Institut für Skandinavistik
Frisistik und Allgemeine
Sprachwissenschaft (ISFAS)

Philosophische Fakultät
der Christian-Albrechts-
Universität zu Kiel

Oliver Niebuhr, Sarah Buchberger (Hrsg.)

KALIPHO

Kieler Arbeiten zur Linguistik und Phonetik

2-3

2014/2015

Institut für Skandinavistik
Frisistik und Allgemeine
Sprachwissenschaft (ISFAS)

Philosophische Fakultät
der Christian-Albrechts-
Universität zu Kiel

Oliver Niebuhr, Sarah Buchberger (Hrsg.)

KALIPHO

Kieler Arbeiten zur Linguistik und Phonetik

2-3

2014/2015

Die KALIPHO-Reihe wird herausgegeben von:

Prof. Dr. Oliver Niebuhr
Juniorprofessor für Analyse gesprochener Sprache
am Institut für Skandinavistik, Frisistik und
Allgemeine Sprachwissenschaft (ISFAS)
Abt. Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
Leibnizstraße 10
D-24098 Kiel

Email: niebuhr@isfas.uni-kiel.de
Tel: +49 431 880 3314

Redaktion und Mitherausgeberin für dieses Heft:

Sarah Buchberger, M.A.
Universität Flensburg
Institut für Germanistik
Auf dem Campus 1
D-24943 Flensburg

ISSN: 2364-2459

Kieler Arbeiten zur Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

Inhaltsverzeichnis

Band 2 (2014)

.

Inhaltsverzeichnis

Band 3 (2015) Special Issue: "Theoretical and empirical foundations of experimental phonetics"

Kieler Arbeiten zur Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

1 You're shtanding on the shtreet— /s/-palatalization in the context of neighboring /r/

Stephanie Berger, B.A.
Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
Steffi_berger.bordesholm@freenet.de

In different varieties of English, the voiceless alveolar fricative /s/ is palatalized to a quality approximating [ʃ]. The process was initially said to appear predominantly in word-initial /str/ clusters. Five Canadian and five Australian speakers were recorded while reading two texts with target words in different /str/ and /sr/ contexts. Center of gravity, standard deviation, skewness and kurtosis were measured in these target words and served as dependent variables in the statistical analyses. The independent variables were the different contexts of the target words: Australian or Canadian English, /t/ as part of the cluster or not, with or without contact between /s/ and /r/, with or without word boundary between /s/ and /r/, and /r/ preceding or following /s/. In one statistical analysis, the independent variables were compared to each other, in another analysis they were compared to reference sibilants. The results suggest that the sibilants were palatalized more when there was contact between /s/ and /r/ and the /s/ was preceding the /r/. The presence or absence of neither /t/ nor a word boundary between /s/ and /r/ played no role in the degree of palatalization. The sibilants were palatalized further in the recordings of the Canadian speakers. Because there was no significant difference in palatalization found in words with or without /t/, the most probable trigger for /s/-palatalization is most likely retroflex /r/, suggesting that the process is assimilation at a distance.

1. Introduction

One assimilation process in the English language has been subject of studies all over the world in recent years: the palatalization of the voiceless alveolar fricative /s/ in consonant clusters like, for example, /str/, resulting in a more /ʃ/-like sound. The contexts

of this process involve retroflex /r/, even though there are similar changes in words with /stj/ clusters that will not be part of this paper. Originally believed to appear only in word-initial /str/ clusters (as in "street", "strong" or "strike"), the phenomenon can now also be found syllable initially (e.g. "catastrophic") and syllable finally (e.g. "first", "worst"), or even across word and syllable boundaries (e.g. "you're standing", "understand"). The /r/ can either precede or follow /s/. The process appears to spread to other contexts as well, for example to words with /sr/ clusters that have no /t/ (for example, cf. Labov 1984; Lawrence 2000; Rutter 2011).

Palatalization was first mentioned by Labov (1984). He defined the process by stating that *"the cluster (str) represents the variation between a hissing and a hushing sibilant before /tr/, though it also extends to /st/ clusters without a following /r/ and across word boundaries"* (Labov 1984: 50). He assumed that /s/-palatalization occurred only with speakers of the lower and working classes (Durian 2007: 66). Labov's data originated in the 1970s, but he already saw a spread to other contexts than word-initial /str/ clusters. Nowadays palatalization seems to be a quite prominent feature of varieties of English all around the world. Singers and actors frequently use a palatalized variant of /s/. The variant is therefore present on television, radio and other spoken-word media.

A lot has been written about palatalization. However, the exact acoustic nature of the palatalized sound does not seem to be clear yet. Is the /s/ realized as a prototypical /ʃ/ or as a sound falling between alveolar /s/ and post-alveolar /ʃ/? What seems to trigger /s/-palatalization?

Those are questions that will be addressed in the course of this paper, and hopefully in the end some tentative answers can be given. For the analyses, speech data of speakers from Canada and Australia were used. In section 2.1 the most important parts of the acoustic background of the involved sounds and sound clusters, as well as general information on palatalization, assimilation and coarticulation will be provided; section 2.2 will offer an overview of the existing research on /s/-palatalization that can be found in the literature. Section 3 will be concerned with the methodology of this investigation, including overviews of the sample (§3.1), the contexts (§3.2) and the methods used to analyze the data (§3.3). The results of the analyses will be presented in section 4 and then critically discussed in section 5, before a conclusion will be reached in section 6.

2. Background information

As already mentioned above, the phenomenon of /s/-palatalization has been part of linguistic research for quite a while. Still no conclusion has been reached as to what might be the cause for the process, seen here as an assimilatory process. Different authors have different suggestions, both for possible triggers and the contexts in which palatalization occurs.

First, some information will be provided on the acoustical background of the sounds involved in the change, as well as the processes of palatalization, assimilation and coarticulation. Then, findings already present in the literature will be looked at. Both will provide an overview of the topic, before the analysis of the speech data can be discussed.

2.1 The sounds and processes involved

2.1.1 /s/ and /ʃ/

The alveolar fricative /s/ and the post-alveolar fricative /ʃ/ are the voiceless members of the group of sibilant fricatives, characterized by their hissing-like sound in high frequencies (Crystal 2003: 417). Fricatives are produced with a constriction in the oral cavity, creating a narrow passage that the airstream has to pass through. Both upon entry and exit of the constriction, the airstream becomes turbulent. The noise gets louder when the channel is narrower or the air hits another obstacle aside from the constriction—for example the teeth (Johnson 2012: 154). Both /s/ and /ʃ/ are louder than other fricatives because *"the upper and lower teeth […] function as turbulence producing obstacles"* (Johnson 2012: 155).

In American English, the [s] is usually unrounded in words like *sea*, but rounded in words like the name *Sue* because of co-articulation (Johnson 2012: 159). The [ʃ], on the other hand, is usually produced with lip rounding, which lowers its frequencies (Johnson 2012: 159). This lowering of the frequencies is further enhanced by the place of articulation being further back in the oral cavity for [ʃ] than for [s]. The rule of thumb is that *"the shorter the front cavity, the higher the frequency of the lowest spectral peak"* (Johnson 2012: 158). Since [s] is produced further towards the front of the mouth it has a shorter front cavity and therefore a higher spectral peak. This spectral peak is at about 2.5 to 3 kHz for [ʃ] and at around 4 to 5 kHz for [s] (Jongman et al. 2000: 1253), though Johnson (2000: 162) even places the spectral peak for [s] near 8 kHz. However, the *"location of the spectral peaks in the frication noise is to some extent speaker […] and vowel dependent"* (Jongman et al. 2000: 1253). Both sounds seem to be characterized by a *"lack of energy below 1200 Hz"* (Olive et al. 1993: 92). Prototypical /s/ is displayed in the spectrogram in Figure 1(a) and prototypical /ʃ/ in Figure 1(b) below. Both sibilants were chosen from the recordings of speaker 01.

To classify sibilants as either /s/ or /ʃ/, Jongman et al. measured four spectral moments. Using a Fast Fourier Transform (FFT), they computed mean, variance, skewness and kurtosis of sibilants. These parameters were also investigated in the present study. What Jongman et al. called *mean* will be referred to as 'center of gravity', reflecting the weighted average energy concentration of the sibilant. They classify the parameter variance as the range of the energy concentration (Jongman et al. 2000: 1253). Together with a later statement (cf. Jongman et al. 2000: 1254), this suggests that they examined the standard deviation of the center of gravity. The four parameters Jongman et al. used are the same that will function as dependent variables in the present study.

Skewness is an indicator of the degree of asymmetry a distribution has. *"Positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies"* (Jongman et al. 2000: 1253). Negative skewness, on the other hand, is *"associated with a positive tilt and a predominance of energy in the higher frequencies"* (Jongman et al. 2000: 1253). Kurtosis indicates the peakedness of a distribution: positive values suggest a spectrum that is clearly defined, while negative values indicate a spectrum that is relatively flat (Jongman et al. 2000: 1253). According to Jongman et al. (2000: 1254), /s/ has *"a higher mean, lower standard deviation, and greater kurtosis,"* while /ʃ/ *"was characterized by a lower spectral mean, positive skewness, and smaller*

kurtosis". That means that when deciding whether a sound is /s/ or /ʃ/, the sound is more /ʃ/-like when it has lower center of gravity, positive (or at least higher) skewness, smaller kurtosis and a greater standard deviation.

At a phonotactic level, as a cluster together with the sounds /t/ and /r/—both individual sounds are discussed below—English /s/ can only occur at the beginning of a word or syllable, and not word-finally (Olive et al. 1993:227). In combination with /r/, /s/ cannot appear word-initially in English. The only word where it is possible—"Sri Lanka"—is only sometimes produced with a [s], but more often with a [ʃ].

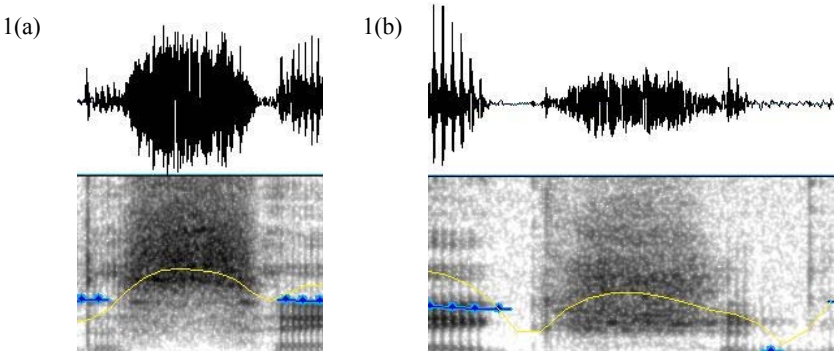


Figure 1. Waveforms and spectrograms of prototypical /s/ on the left (a, taken from the word 'same' in the phrase "… out of the same cycle …") and prototypical /ʃ/ on the right (b, taken from the word 'should' in the phrase "… Crazy Mike should not …"). Both examples were chosen from the text "Mister Stevens" and were produced by speaker 01. The difference in frequency concentration can be seen by the dark patches of energy.

2.1.2 /t/, /r/, and different clusters

The voiceless alveolar plosive /t/ is usually aspirated in English as long as it does not occur in a consonant cluster. As all plosives, the /t/ is produced with a complete oral closure, a following burst and an aspiration period. When compared to the other voiceless plosives /p/ and /k/, *"the energy of the burst and aspiration regions […] is the highest"* for /t/ (Olive et al. 1993: 88). The aspiration period is also called voice onset time (VOT). The +VOT of /t/ is longer than that of /p/, but shorter than the +VOT of /k/ (Olive et al. 1993: 88).

When /t/ follows /s/ in a cluster, as in "still", the /t/ is usually unaspirated—unless the cluster is found across word or syllable boundaries, as in "festive"—and often there is no discernible closure period. Usually, the vibrations of the vocal chords decay at the end of the fricative /s/ and a slight rise of energy can be observed in the beginning of the plosive (Olive et al. 1993: 260).

The affricate /tʃ/—one of the possible triggers for palatalization that will be mentioned below—is a combination of a plosive and a fricative. The sounds are usually homorganic. There is a *"distinct region of short voiceless closure"* that is then ended with a short burst before the energy rapidly decays (Olive et al. 1993: 242). Then, the

energy increases again at the onset of the fricative and decays gradually at its end (Olive et al. 1993: 242). Compared to an isolated fricative, the rise time of the /ʃ/ in an affricate is shorter (Johnson 2012: 179f.). That means that the full level of energy is reached faster in an affricate. Phonologically, the affricate is sometimes considered to be either *"underspecified for place of articulation features"* or *"unordered with respect to each other"* (Lawrence 2000: 84). This means that it is not clear which of the sounds of the affricate comes first or where exactly the affricate is produced.

In English, /r/ is usually realized as the retroflex approximant [ɻ]. At least American speakers seem to have two different variants of producing /r/: retroflexion *"involves curling the tongue tip back to the palate,"* while tongue bunching *"involves raising the tongue body to the velum"* (Rutter 2011: 30). /r/ is also often accompanied by lip rounding (Olive et al. 1993: 27). As a phoneme, the approximant will be transcribed as /r/ here. When the phone is being referred to, the symbol [ɻ] will be used. In some quotes, the phone is transcribed as [ɹ] or [r].

The use of sound characteristics of /r/ is considered to be one of the most important features when different varieties of English are supposed to be distinguished. There are varieties that have non-prevocalic /r/ (also called rhotic varieties). In these varieties /r/ is present both before and after vowels. Rhotic varieties of English are, for example, Standard American English, Scottish English, Irish English and Canadian English. Other varieties only have prevocalic /r/. These varieties are called non-rhotic varieties, and include Australian English, New Zealand English and Received Pronunciation (RP) as well as most British English varieties (Piercy 2012: 77). Even non-rhotic varieties can have the production of a word-final /r/ when the following word begins with a vowel (Piercy 2012: 78). This is a feature called linking /r/. Rhoticity also influences the quality of the vowel before or after the retroflex; and even in non-rhotic varieties the vowel quality seems to be different from words where the /r/ would not be realized.

The difference between the two major groups of varieties can be traced back to the colonial history. Countries in the southern hemisphere—like Australia or New Zealand—were settled only in the nineteenth century. By then, the /r/ was lost in Standard Southern British English, the major British English variety. Northern hemisphere countries like the United States or Canada, on the other hand, were settled in the seventeenth and eighteenth centuries before non-prevocalic /r/ was lost in the prestigious British varieties, and therefore this variant could spread in most parts of North America (Boberg 2010: 102).

The /t/ in the consonant cluster /tr/ is often affricated (Lawrence 2000: 83). Even though this was not acoustically examined, the auditory impressions of clusters found in the speech recordings used for this study seem to be in line with that statement. Olive et al. (1993: 280) say that

> "the center of the /r/ occurs during the aspiration region [of the plosive] and before the onset of voicing. The production of /r/ during the aspiration is possible because after the closure is released, the tongue is free to retroflex."

This statement does not appear to contradict the point of view that /t/ in /tr/ clusters is affricated. If the retroflex is already being produced while the plosive is being aspirated, it seems reasonable to expect that the quality of the aspiration is changed and could very well turn into something [ʃ]-like. The tongue is moved further back in the oral

cavity, while the airstream is still flowing. A constriction is created and the sound can change from a glottal fricative to a sound similar to or even an actual post-alveolar fricative. According to Olive et al., the /t/ is no longer aspirated when the /tr/ cluster is preceded by /s/ (1993: 281). The auditory impression of the evaluated speech recordings do not always concur with this statement.

2.1.3 Palatalization

Palatalization refers *"to any articulation involving a movement of the tongue towards the hard palate"* (Crystal 2003: 333) and it is very common as a secondary articulation. A sound with a place of articulation that is different from a palatal sound can be produced further back towards the palate when a palatal sound is neighboring it.

The /s/-palatalization in English researched in this paper involves the sibilant /s/ that becomes a sound approximating the quality of /ʃ/. That means that the former alveolar fricative is then produced further back in the oral cavity. It has become either indistinguishable from an actual post-alveolar fricative or it is produced somewhere between the alveolar and post-alveolar regions. The trigger for this process is not clear. The sibilants investigated in this study all appear in different contexts containing /s/ together with either both /t/ and /r/ or only /r/. Since the /r/ in English is produced quite far back in the oral cavity, especially when it is produced as a retroflex sound, it seems plausible that this sound pulls the fricative further toward the palate. Studies that already exist do not seem to agree when it comes to the question of what is the trigger for /s/-palatalization in English, as will become clear in the following sections.

2.1.4 Assimilation and coarticulation

Assimilation is a process that is defined as the *"influence exercised by one sound segment upon the articulation of another, so that the sounds become more alike or identical"* (Crystal 2003: 38). In non-linear approaches to the topic, assimilation is seen as a feature spreading from one segment, also called the trigger, to another segment, the target (Crystal 2003: 39). Contact assimilation occurs when the trigger and the target are adjacent to each other and distance assimilation occurs when the trigger is further away than the target (Crystal 2003: 38). The assimilation is classified as a total or complete assimilation when *"the target segment acquires all the features of the trigger"* (Crystal 2003: 39); and it is classified as partial or incomplete assimilation when the target takes on only part of the trigger's features.

Traditionally, assimilation is seen as a phonological process that categorically changes a sound. However, modern studies see assimilations as *"gradual rather than categorical processes that generate a broad array of intermediate forms in terms of both speech timing and spectrum"* (Niebuhr and Meunier 2011: 135). Remnants of the assimilated sound can still be found in preceding vowels or, in the case of sibilant assimilation, can still be found *"in a initial transition in the phonetic quality of the friction"* (Niebuhr and Meunier 2011: 135).

These new findings blur the formerly sharp differentiation between assimilation as a phonological process and coarticulation as a phonetic process. In this paper, the author will refrain from partaking in the complex terminological and theoretical discus-

sion on the topic and will instead use the term assimilation without its traditional, theoretical implications. The term assimilation is also used in other studies dealing with /s/-palatalization and can today be seen as the more conventional term for sound interaction in general.

However, there seems to be proof for classification as assimilation in the traditional sense in the data of the present study. As will become clear in the course of the paper, /s/-palatalization also occurs across word boundaries and with /t/ between /s/ and /r/. That suggests that the temporal distance does not play a role concerning degree of palatalization which argues against coarticulation.

2.2 /s/-palatalization: literature and hypotheses

In 1984, Labov commented on data elicited in the 1970s, as was mentioned above. He found that /s/-palatalization was a change in progress in Philadelphia mainly concerning the cluster /str/ but also occurring in other contexts like *"/st/ clusters without a following /r/ and across word boundaries"* (Labov 1984: 50). Furthermore he described the phenomenon as a feature of the speech of lower and working classes (Durian 2007: 66). While this distribution of palatalized /s/ was probably correct in the 70s, by now the feature seems to have spread—both socially and globally. Prominently occurring in the media and among wide parts of the population, *"/s/ palatalization in /stɹ/ could be considered to be the dominant form,"* at least *"in some parts of the USA"* (Rutter 2011:27).

Apart from its occurrence in American English, the existence of /s/-palatalization is also attested for British English (Altendorf 2003; Bass 2009) and New Zealand English (Lawrence 2000). However, the most extensive research on /s/-palatalization has been conducted with American English: Labov (1984) found the variable in his Philadelphia study; Shapiro (1995) observed the phenomenon in the speech of American public figures; Durian (2007) conducted studies of the speech in Columbus, Ohio; and Rutter (2011) looked at Southern Louisiana American English. Boberg (2010) briefly mentions the process in Canadian English, but only on an anecdotal basis without experimental evidence.

Section 2.2.1 deals with the acoustical studies included in Shapiro (1995), Lawrence (2000), Durian (2007), Boberg (2010) and Rutter (2011). In section 2.2.2, the sociolinguistic embedding of /s/-palatalization as researched by Altendorf (2003), Durian (2007) and Bass (2009) will be presented. Section 2.2.3 will offer a brief summary of what is known about /s/-palatalization so far as well as hypotheses of the possible triggers of this sound change process.

2.2.1 Acoustical studies on palatalization in English

Shapiro published the first extensive description of /s/-palatalization. He looked at the American English of people in the public eye, for example radio hosts or sports commentators. He found that the /s/-to-/ʃ/ change was *"neither dialectal nor regional"* (Shapiro 1995: 101). Shapiro observed the *"change of /s/ to /ʃ/ before /tr/ […], which involves a palatalization of the initial sound in the cluster /str/, typically in initial position but not exclusively"* (Shapiro 1995: 101). The resulting sound is not always a typical American [ʃ], but *"acoustic evidence suggests that the initial fricative could be –*

phonetically – a retroflex [ʂ];" however, he *"consistently heard varieties of [ʃ] and not retroflex [ʂ]"* (Shapiro 1995: 102).

Concerning the type of assimilation involved in /s/-palatalization, he first states that the alveolar plosive /t/ undergoes no phoneme change. Because of his interpretation, Shapiro concludes that it *"can, therefore, not figure in the assimilation of /s/ to /r/, if [it is an] assimilation […] it is not an example of contact assimilation but of assimilation at a distance"* (Shapiro 1995: 103). He mentions that /tr/ clusters can be affricated in American English, the fricative replacing the aspiration. The resulting sound might be either [tʃɻ] or [tʂɻ] and the /s/ *"might be thought to be assimilating to the immediately contiguous retroflex [t]"* (Shapiro 1995: 103). However, he concludes that the /t/ is not changed in /str/ clusters because *"Olive et al. never speak of affrication, only of unaspiration in /str/ clusters"* (Shapiro 1995: 103). Shapiro therefore draws the final conclusion that *"/s/ changes to /ʃ/ owing to the presence of /r/"* (Shapiro 1995: 103), mainly because of the similarity between /r/ and /ʃ/. He argues that while assimilation takes place in an /str/ context, no palatalization could be found *"in /st/ clusters lacking /r/"* (Shapiro 1995: 103), contradictory to what Labov (1984) observed.

In his comment on Shapiro's paper, Lawrence says that /s/-palatalization is *"not assimilation to /r/ and […] not […] assimilation at a distance"* (Lawrence 2000: 82). He also states that the process occurs also across word boundaries, not only word-internally (Lawrence 2000: 82). Lawrence quotes Shapiro by saying that there is a *"similarity between [ʃ] and [r], in that they are both palatal"*. That *"accounts for [s] becoming [ʃ] when it precedes /r/ in the same syllable, as in, for example, Sri Lanka. In the speech of some, this process [also applies] over syllable boundaries",* for example in "classroom" (Lawrence 2000: 82).

Lawrence argues against Shapiro's explanation of /s/-palatalization as assimilation at a distance by giving the example of one speaker in his study who used [ʃtɻ], but not [ʃɻ] (Lawrence 2000: 82). One *"would not expect to have assimilation at a distance (i.e. across /t/) without also assimilation of the same target when adjacent to the trigger, this shows that /str/ → /ʃtr/ cannot be conflated with /sr/ → /ʃr/"* (Lawrence 2000: 82f.).

Shapiro denies affrication in /str/ because it does not appear in Olive et al., as was mentioned above. Their analyzed sequence, however, does not involve palatalization (Lawrence 2000: 83). Lawrence observed that *"/t/ is always affricated in cases where /ʃtr/ is used"* (Lawrence 2000: 83). He offers two views that suggest that /str/ → /ʃtr/ could be seen as *"assimilation of adjacent features"*. The first view is that coronal segments like /t/ are often underspecified for place of articulation, but are rather "open windows" that allow the *"place of articulation features of the following fricative /ʃ/ to be adjacent to those of any segment immediately preceding the /t/"* (Lawrence 2000: 84). The second view is that phonologically the two components of an affricate are unordered: They basically *"lie one above the other," making /s/ "be simultaneously adjacent to both"* affricate components (Lawrence 2000: 84). According to Lawrence (2000: 84), *"all phonological processes are local"*. He therefore concludes that /str/ → /ʃtr/ involves local assimilation of adjacent features due to affrication.

Like Lawrence (2000), Rutter (2011: 27) calls *"the realization of /t/ as [tʃ] [a] possibly related factor"* to the process of /s/-palatalization (Rutter 2011: 27). In his study he sets out to examine whether or not the palatalized variant of the sibilant is identical to typical English [ʃ] and if there are intermediate forms (Rutter 2011: 28). According

to him, three parameters are important regarding the sound change: the tongue placement, the tongue shape and the lip shape. /r/ and /ʃ/ are more similar to each other in these parameters than /r/ and /s/. Possible intermediate forms might arise because *"the parameters associated with [s] are shifting towards [ʃ] at different rates"* (Rutter 2011: 31). Rutter states that the *"change from /s/ to /ʃ/ […] in the context of /ɹ/ could be seen as a harmonizing process, possibly incorporating rounding and retraction of the /t/"* (Rutter 2011: 31). Therefore, it seems as if Rutter considers a combination of both /r/ and affricated /t/ as the trigger of /s/-palatalization.

In his study, Rutter investigated the speech of ten native speakers of Southwest Louisiana American English. He examined target words in a carrier phrase the participants had to produce (Rutter 2011: 32). The target words had four different vowels preceded by the four consonant onsets /s/, /ʃ/, /ʃɹ/ and /stɹ/ (Rutter 2011: 32). In an auditory analysis, Rutter found *"a variety of /ʃ/ like fricatives in the onset of /stɹ/ as well as some more /s/-like productions"* (Rutter 2011: 34). These /s/-like productions later proved to be acoustically more /ʃ/-like.

> "The onset /stɹ/ can be seen to pattern with both /ʃ/ and /ʃɹ/, exhibiting a major spectral peak between 2000 Hz and 4000 Hz and a gradual loss of energy in the higher frequencies [while the] alveolar fricative […] exhibits its major peak between 6000 Hz and 8000 Hz" (Rutter 2011: 34).

The acoustical analyses showed that /str/ had a greater range of spectral peaks than /ʃ/ or /ʃɹ/. This range, however, was lower than that of productions of /s/ (Rutter 2011: 37).

There also seemed to be *"two different types of /ʃ/ in /stɹ/: one with a very low spectral peak [(usually followed by /u/)], and one with a very high spectral peak"* (Rutter 2011: 37). Since sibilants without potential palatalization were also examined, the sibilant in /str/ could be compared to "normal" productions of /s/ and /ʃ/. Most fricatives of /str/ were in their production like canonical /ʃ/, while those that were not still did not fall into the range of /s/ (Rutter 2011: 38). Rutter therefore concludes that there is one intermediate form, acoustically falling between /s/ and /ʃ/. This intermediate form is caused by palatalization but a lack of lip rounding (Rutter 2011: 38).

While Boberg mentions /s/-palatalization in Canada only very briefly and with purely anecdotal evidence, he—like Rutter—seems to lean towards a combined trigger for /s/-palatalization. He agrees with researchers like Shapiro by stating that the *"apparent cause of the retraction [of /s/] is the tongue constriction required for the /r/"* (Boberg 2010: 231). He states further that the /t/ in initial /tr/ clusters is palatalized and that this palatalization is now spreading to the /s/ in /str/ clusters (Boberg 2010: 231). To him it *"seems most likely that this change is diffusing into Canada from the United States via popular culture, such as television and popular music"* (Boberg 2010: 231).

Aside from /r/ and /tʃ/ as possible triggers for /s/-palatalization, Durian (2007) offers a third possibility, together with an account of the sociolinguistic background of the process (addressed further in section 2.2.2) as well as an acoustic analysis of the sibilants' features in /str/ clusters in Columbus, Ohio. According to Durian, *"the alveopalatal [ʃtr] is treated as the prototypical vernacular variant and [the] alveolar [str] as the standard"* in Columbus (Durian 2007: 65). While [ʃtɻ] is more strongly associated with urban speakers there, the use of this variant is also increasing in the suburbs (Durian

2007: 66). Durian mainly investigated the distribution of [ʃtɻ] in neighbourhoods across Columbus and nearby suburbs (Durian 2007: 65), but also looked at the acoustic properties of palatalization in detail.

Three different realizations of /str/ were assumed for Columbus: The standard variant [stɻ]; an intermediate variant [ʂtɻ], as a retroflex but unrounded sound; and the rounded variant [ʃtɻ], which is the vernacular variant (the form used by the community) (Durian 2007: 66). The results of analyses of interviews conducted with middle class speakers showed the acoustical characteristics of each form. [ʃtɻ] appeared to have a high concentration of spectral energy at or below 2500 Hz. The spectral energy of the intermediate form seemed to be roughly between 3000 and 3500 Hz; and the energy for [stɻ] seemed to be at or above 4000 Hz (Durian 2007: 70). The vernacular form was produced more often in word-medial environments (Durian 2007: 71). Therefore he sees the trigger of /s/-palatalization as a vowel preceding or following the sibilant.

2.2.2 The sociolinguistic embedding of /s/-palatalization

According to Durian (2007: 66), the strongest social factors regarding /s/-palatalization are class, age and the location in which the speaker was born and raised. Younger speakers have more [ʃtɻ] or intermediate forms than older speakers (Durian 2007: 73). [ʃtɻ] seems to be quite prominent in Columbus' middle class, the social group Durian's study focused on. His informants were either from Columbus or the city's suburbs. The palatalized variant [ʃtɻ] might be used as a *"marker of 'urban affiliation'"* (Durian 2007: 66), supported by the observation that *"speakers showing a higher level of 'urban affiliation' showed a higher use of (str) than their counterparts, who show lower levels of affiliation in the interviews, regardless of the location in which they presently live"* (Durian 2007: 77). The variant seems to have spread from the city to the suburbs, supported by the movement of older speakers from Columbus—where they acquired the palatalized variant—to the suburbs.

Aside from North American varieties and New Zealand English, /s/-palatalization is also present in British English. However, it seems as if mostly social aspects of the process were investigated, instead of acoustic characteristics. Altendorf (2003) briefly mentions /s/-palatalization in her study dealing with Received Pronunciation (RP) and Estuary English. Estuary English is a *"variety of British English supposedly originating in the counties adjacent to the estuary of the River Thames, and thus displaying the influence of London regional speech, especially in pronunciation"* (Crystal 2003: 166). In this variety of English, the process seems to be restricted to three contexts, two of which are relevant for the present paper: /st/ clusters, though this is considered to be *"extremely rare"* (Altendorf 2003: 70), and /str/-clusters. It seems to be present in working-class and middle-class speech in three cities that were investigated: London, Colchester and Canterbury (Altendorf 2003: 101). The use of /s/-palatalization appears to *"[reflect and convey] an attitude of informality and nonchalance. Those who use these forms affiliate themselves with the 'young' and the 'cool' and distance themselves from the 'formal' and the 'stuffy'"* (Altendorf 2003: 154).

Bass studied the social stratification of /s/-palatalization in the same area as Altendorf, more specifically in Colchester English. In his study, Bass compares younger and older speakers as well as men and women. His results show a clear difference between

the speech of young speakers (between 16 and 21) and old speakers (over 65) (Bass 2009: 12). Younger speakers used [ʃtɻ] in 60 % of /str/ occurrences, while older speakers used this variant only in 25 % of its occurrences. Palatalization therefore presents itself as a change in progress (Bass 2009: 14).

The variant [ʃtɻ] also seems to be influenced by gender. Men use more [ʃtɻ] than women. That is unexpected because usually *"women are said to lead linguistic change"* (Bass 2009: 14). Young men use [ʃtɻ] more often than older men and young women more than older women (Bass 2009: 16). However, while younger women use the standard variant [stɻ] more often than younger men, they use [ʃtɻ] more than older men, suggesting that age is a more important factor in the sound change than gender (Bass 2009: 16). The social embedding of /s/-palatalization will not be investigated in the present study.

2.2.3 Brief summary and possible triggers

To summarize what has already been stated about /s/-palatalization in the literature, it can be said that this process seems to be a change in progress. Younger speakers seem to use the palatalized variant more frequently than older speakers, with men apparently leading the change. Its distribution in society appears to have spread out from being a working class phenomenon in the 1970s to turning into the vernacular way of pronouncing words containing [s], [t] and [ɻ] clusters in different contexts and in different countries and varieties of English.

The acoustical properties appear to be uncertain at this point. Some researchers say there is only one intermediate variant and some say there are more. For some speakers, palatalization also occurs in /sr/ clusters. In different papers, three different possible explanations for what might trigger /s/-palatalization have been offered, all of which make sense. The three ideas are listed below:

1. Assimilation at a distance, triggered by retroflex [ɻ], rounding and retracting the [s] (Shapiro 1995).
2. Assimilation of adjacent features due to affrication of [t] in /tr/ clusters (Lawrence 2000).
3. The trigger is neither [t] nor [ɻ], but a vowel that is preceding or following /s/, /t/ and /r/ (Durian 2007).

The third possible explanation will not be addressed further because it is independent from the retroflex [ɻ] and therefore not connected to the topic of this paper. The first possible trigger, however, has an immediate connection to the topic. The second possible trigger is also indirectly connected to the retroflex: Affrication seems to occur in this context because of the retroflex as part of the cluster. In the present study, some open questions may be tentatively answered. The set-up of the study, explained further throughout section 3, offers the opportunity to investigate possible differences in /s/-palatalization between Australian and Canadian English as well as differences between different word contexts. To reach a conclusion of the degree of palatalization, five independent variables were examined. The elicited data were separated into two levels for each variable to compare opposite conditions that are listed in Table 1 below.

Context group	Context	Explanation
AUS_CAN	*AUS*	Speakers of Australian English
	CAN	Speakers of Canadian English
WithT_noT	*WithT*	Words with /t/ as part of the consonant cluster (e.g. /str/, /r.st/, /st# #r/, etc.)
	NoT	Words without /t/ as part of the cluster (e.g. /sr/, /r.s/, /s# #r/, etc.)
WithWB_noWB	*WithWB*	With a word boundary between /s/ and /r/
	NoWB	Without a word boundary between /s/ and /r/
WCont_nCont	*WCont*	With contact between /s/ and /r/; a /t/ or a word boundary between /s/and /r/ was still counted as *WCont*
	NCont	Without contact between /s/ and /r/, e.g. a vowel or a consonant other than /t/ between /s/ and /r/
Rs_sr	*Rs*	The /r/ is preceding the /s/ in the target word
	Sr	The /r/ is following the /s/ in the target word

Table 1. The context groups used for the analyses and their definitions.

The variables will be referred to with their respective abbreviations in the course of this paper. With this procedure, the discussion will revert to the following questions:

- Which contexts lead to more palatalized realizations of /s/?
- Is /s/-palatalization a gradual process with intermediate forms?
- Is /s/-palatalization stronger in the sample of Canadian or Australian speakers?
- Which of the possible triggers of palatalization is more relevant?

3. Methodology

In the following sections the methodology of the conducted study will be presented. First, the sample of speakers will be introduced. Second, the different contexts and the target words used for the analyses will be shown. Then the analyses will be explained.

3.1 The sample

The sample of speakers used for this study consists of five female speakers from Canada as well as two female and three male speakers from Australia. The Canadian speakers were all students at McGill University in Montréal, Québec. The origins of these speakers cannot be specified further, though. The age can be assumed to be between 17 and 30. These five speakers were recorded by Meghan Clayards at McGill University in 2011. In total, eleven speakers were recorded at the time, each of them reading two different texts containing target words from different contexts—one text (“Mister Stevens”) being the source for target words in different /str/ contexts, the other (“Julianna”) consisted of target words with /sr/ contexts. The five speakers that are part of this study were chosen because their speech seemed most promising in terms of audible /s/-palatalization.

The Australian speakers were recorded by Hywel Stoakes in June 2014 at the University of Melbourne. Each participant recorded the same texts as the Canadian participants—even though the texts were slightly revised. Each text then was read in three different speeds: slow, normal and fast. For this thesis, only the recordings with the fast production of the texts were considered as this speed matched the speed of the Canadian speakers, therefore offering the opportunity to compare both sets of data, since speech rate has a major influence on assimilation (cf. Dilley and Pitt 2007). Both the Canadian speakers and the Australian speakers in their "fast" recording produced on average approximately six syllables per second. In the recording deemed having "normal" speed, the Australian speakers produced only approximately five syllables per second. The average speech rate is illustrated in Figure 2 below. Table 2 offers an overview of the speakers' origins as well as the speakers' number.

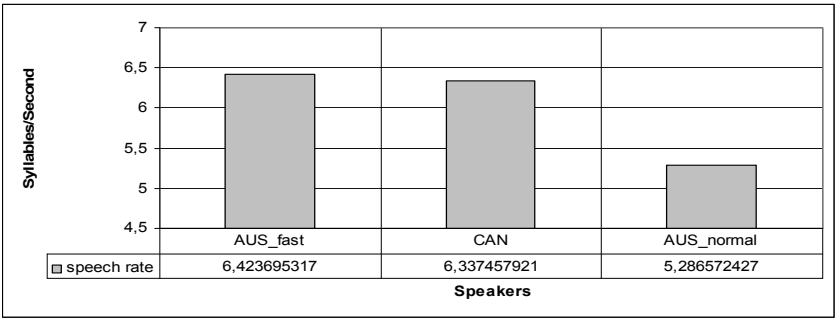


Figure 2. Average speech rates of the speakers in the sample. The averages were created by counting the syllables in a stretch of approximately 1 second. For each speaker, 10 seconds were analyzed and then averaged. This average was then used to calculate the approximate syllables per second for all speakers in a group, depicted above.

Number	Origin	Gender	Age	Profession
01	Canada	female	ca. 17-30	Student
02	Canada	female	ca. 17-30	Student
03	Canada	female	ca. 17-30	Student
04	Canada	female	ca. 17-30	Student
05	Canada	female	ca. 17-30	Student
06	Melbourne, Australia	female	26	PhD student
07	Melbourne, Australia	female	27	PhD student
08	Melbourne, Australia	male	40	Audio engineer
09	Australia, England	male	39	Phonetician
10	Sydney, Australia	male	36	Student

Table 2. Speaker numbers, their origins, age and professions.

3.2 The contexts

The targets chosen allow a consideration of two of the three possible triggers for /s/-palatalization mentioned above. Should the process be a case of assimilation at a dis-

tance, as proposed by Shapiro (1995), there should be a difference in /s/ retraction between words with /t/ or without /t/. If the process was assimilation at a distance, the /s/ should also be palatalized when it is adjacent to the trigger, and possibly even more severely.

Assimilation to affricated /t/, considered an assimilation of adjacent features (cf. Lawrence 2000), can not be the case if palatalization can be attested in /sr/ contexts as well. Should /s/-palatalization occur in these contexts, it could indicate that the process might in fact be triggered by the retroflex and in that case it should also occur with /sr/-clusters. Should there be no palatalization in contexts without /t/, it could be an indicator that the [ɻ] might not be the trigger, but possibly the affricate. If there is an effect, but slighter than with the /t/, a combination of both possible triggers could be reasonable. Since the Australian speakers do not have non-prevolcalic /r/ as the Canadian speakers do, not all of the contexts investigated will be left with the retroflex /r/ involved. Even though remnants of the retroflex can usually be observed as a difference in vowel quality, the retroflex itself is usually no longer present after vowels (cf. Heid and Hawkins 2000). Less palatalization in the Australian data could also indicate that there might be an influence of /r/. All target words were part of five independent variables that were already mentioned in section 2.2.3: *AUS_CAN*, *WithT_noT*, *WithWB_noWB*, *Rs_sr* and *WCont_nCont*. Each target word was assigned to one of the two levels of each variable. The target *strong*, for example, was part of the variables *WithT*, *WCont*, *NoWB* and *Sr* as well as either *AUS* or *CAN*, depending on the speaker.

The target words were taken from two texts that were used for the speech recordings. These texts have already been used in an earlier survey with Canadian speakers, were designed by Karo Kress at Kiel University in 2010 and slightly revised by the author. The data of five of these speakers is also included in this paper. The text entitled "Mister Stevens" includes 50 target words from 19 contexts of clusters with /s/, /t/ and /r/. In some contexts, the /r/ precedes the /st/, in others it follows /st/. In 18 target words, the sounds occur across word boundaries. One target—"Mister Stevens"—has four occurrences, another target—"Costa Rica"—occurs three times. There are 33 target words without a word boundary between /s/ and /r/. /s/, /t/ and /r/ are either in the same syllable or occur across syllable boundaries. The target "Mister" occurs four times.

The target "restless" was added to the text later and is only present in the recordings of the Australian participants, but not in the recordings of the Canadian speakers. The target "first" is followed by the word "rays" in the data of the Canadian speakers. In this context, it is not clear which /r/ might have influenced the alveolar fricative. For the Australian speakers, "rays" was substituted by "light", therefore eliminating a possible effect of retroflex sounds both before and after the fricative. For the statistical analysis the second /r/ in the data of the Canadian speakers will be left aside. Table 3 below offers the targets included in the text, as well as the contexts the words are part of.

The second text has the title "Julianna" and includes 34 target words belonging to different /sr/-contexts. 16 targets have a word boundary between /s/ and /r/, while 18 targets do not. These targets have /s/ and /r/ in the same or different syllables. The different target words can be found in 16 contexts, with /r/ either following or preceding the alveolar fricative. Table 4 below shows the targets and contexts of the /sr/-category.

Contexts				Examples
Rs	*WithWB*	*WCont*	-r# st-	Miste**r** **St**evens, othe**r** **st**ates, an-othe**r** **st**ory
		NCont	-r# __st-	he**r** a**st**onishment, Pete**r** U**st**inov
			-r__# st-	eve**r**y **st**ep, ve**r**y **st**able
	NoWB	*WCont*	-rst#	fi**rst**, wo**rst**
			-r.st-	doo**rst**ep, unde**rst**and
		NCont	-r.__st-	unde**rest**imate
			-r__.st-	p**rist**ine, tu**rnst**ile
			-r__s.t	a**rrest**ed, characte**rist**ics
			[-]r__st-	**rest**less
Sr	*WithWB*	*WCont*	-st# r-	mu**st** **r**emember, te**st** **r**esults
			-s# tr-	Steven**s** **dr**ank
		NCont	-st__# r-	Co**st**a **R**ica, du**st**y **r**oad, pe**st**o **r**ice
			-st# __r-	be**st** **f**riend, ca**st** **i**ron, la**st** **F**riday
			-st# __#r-	ju**st** a **r**outine
	NoWB	*WCont*	[-]str-	cata**str**ophic, **str**ange, **str**eam, **str**eet, **str**ike, **str**iking, **str**ong
			-s.tr-	de**str**oyed, di**str**acting, di**str**ict, fru**str**ation, regi**str**ation
			-st.r-	po**st** **r**egistration, po**st** **r**oom
		NCont	[-]st__r-	di**st**u**r**bing, Mi**st**e**r**, si**st**e**r**, **st**a**r**ted, Ul**st**e**r**
			st.__r-	po**st**-**gr**aduate
			-st__.r-	ma**st**e**r**ing, re**st**o**r**ed, **st**e**r**eo

Table 3. /str/ contexts and target words in the text "Mister Stevens".

Contexts				Examples
Rs	*WithWB*	*WCont*	-r# s-	eve**r** **s**ince, fo**r** **s**everal, o**r** **s**ome
		NCont	-r# __s-	fo**r** a**s**ylum, he**r** e**ss**ay, neve**r** a**s**ked
			-r__# s-	eve**r**y **S**unday, ve**r**y **s**imilar
	NoWB	*WCont*	-rs#	nu**rs**e, shoulde**rs**, yea**rs**
			-r.s-	he**rs**elf, pe**rs**on, unive**rs**ity
		NoCont	-r.__s-	unde**r**-a**ss**essed
			-r__.s-	c**ris**is, p**res**ence
			-r__s	Se**ries**
Sr	*WithWB*	*WCont*	-s# r-	hi**s** **r**ole, wa**s** **r**eally, wa**s** **r**equired
		NCont	-s__# r-	a**ls**o **r**esponsible, ea**sy** **r**each
			-s#__r-	hi**s** a**r**m, thi**s** a**r**ea
	NoWB	*WCont*	sr-	**Sr**i [Lanka]
			-s.r-	cla**ss** **r**oom, di**sr**egarded, I**sr**ael
		NCont	[-]s__r.-	**s**e**r**ies, **s**e**r**vice, **s**u**r**face, nece**ss**a**ry**
Rs/Sr	*WithWB*	*WCont*	-r# #sr-	o**r** **Sr**i

Table 4. /sr/ contexts and target words in the text "Julianna".

3.3 The analyses

For the analyses, four acoustic properties have been measured using Praat (Boersma and Weenik 2014). The properties include the center of gravity and the standard deviation (both in Hz) as well as skewness and kurtosis of the sibilant. All properties have been measured in the sibilants of the target words listed in tables 3 and 4 above as well as in reference sibilants. These reference sibilants include eleven alveolar fricatives as well as eleven post-alveolar fricatives. The reference sibilants can be found in Table 5 below. They were chosen from both texts and taken from contexts where neither [t] nor [ɹ] were closer than two syllables to the sibilant. This set-up minimizes the chance of palatalization in these contexts and allows a comparison of acoustical properties to the potentially palatalized targets. A wide range of contexts was chosen for the reference sibilants (i.e. word-initially, word-medially, word-finally, between different vowels) to also achieve a wide range of the sibilant including co-articulatory influences. This allows a classification of the target sibilants as part of the range or not.

/s/			/ʃ/
Texa**s**	**s**omething	**s**ame	**sh**e (7 times)
cycle	him**s**elf	**s**ome	**sh**ould (2 times)
sun	son**s**	hou**s**e	**sh**own
poli**c**e	**s**o		a**sh**amed

Table 5. Target words for reference sibilants.

The sound files were annotated using Praat. Each TextGrid had four tiers: one for interval, word, sibilant and reference, respectively. On the interval tier, the phrases in which a target word appeared were segmented and numbered. The target words were segmented and annotated orthographically on the word tier. The sibilants were segmented from the rest of each target word on the sibilant tier, as were the sibilants used for reference sibilants on the reference tier. The boundaries for the sibilants were determined by looking at the spectrogram and then listening to each sound to make sure only the sibilant was chosen. This was especially crucial in cases where /t/ followed /s/ but did not have a real closure period, and instead seemed to be the same as /s/ when looking at the spectrogram. The boundaries were set so that the increase and decrease of noise energy in and out of the fricative was not included, but the friction spanned to the top of the spectrogram. The spectrogram settings in Praat were set to a frequency range from 0 to 15,000 Hz.

The sibilants were extracted, saved, and then measured. Then the sounds were analyzed as spectra using a Fast Fourier Transform (FFT). From these spectra, center of gravity, standard deviation, skewness and kurtosis were computed. To establish whether or not these values are in fact similar to prototypical sibilants, the four values were also measured for reference sibilants without influence of [t] or [ɹ]. This way, the context groups could be compared to the reference sibilants.

Statistics have been carried out using SPSS (IBM Corp. 2012). A five-way multivariate analysis of variance (MANOVA) has been computed with the independent variables *AUS_CAN*, *NoWB_withWB*, *Rs_sr*, *WithT_noT* as well as *WCont_nCont*. The most significant interactions between different factors were also looked at. Only effects with a significance of $p \leq 0.001$ will be addressed in this paper. A lot of variables were

part of the statistical analyses; the sample of speakers was quite small; the small sample size led to a relatively small amount of data; and the sample was not evenly distributed within the different contexts. These problems can influence the statistical analyses and could potentially lead to misleading significance values. Choosing the highly significant effects only, the chance of accidental classification as significant can be minimized. The dependent variables in the analyses were the four measured variables: the center of gravity and its standard deviation, skewness and kurtosis. Interactions between the independent variables were approached in the same way: only those that were statistically highly significant were included in this analysis. The problem with interaction is that the simple effects cannot be separated from the interactions. It is not always clear if the dependent variables are really significant when only one independent variable is examined, or if this significance is the result of an interaction of an independent variable with other independent variables.

After this analysis, two further MANOVAs have been carried out. One MANOVA focused on the sibilants produced by the Australian speakers; the second contained only the sibilants produced by the Canadian speakers. Both analyses compared the independent variables *NoWB_withWB*, *Rs_sr*, *WithT_noT* and *WCont_nCont* to the reference sibilants of each group, always in terms of the dependent variables. This way, the distinction of /s/ and /ʃ/ can be made in reference to the major groups of word contexts and nationality. It is generally expected that the sibilants in the contexts *WithT*, *Sr*, *NoWB* and *WCont* are the contexts with larger degrees of palatalization, because these contexts are the ones that were first and most prominently recognized to be involved in /s/-palatalization.

4. Results

Considering the effect of the entire model where all independent variables were taken together, all of the four dependent variables were highly significant. This was the first MANOVA that was computed. There were significant main effects on skewness ($F[1,865] = 24.039$, $p < 0.001$), standard deviation ($F[1,865] = 9.796$, $p < 0.001$), and the center of gravity ($F[1,865] = 9.711$, $p < 0.001$). Kurtosis was the dependent variable with the lowest significance ($F[1,865] = 3.065$, $p < 0.001$).

In section 4.1, the significant main effects of the independent variables will be presented. In section 4.2, the highly significant interactions will be discussed. Section 4.3 then discusses the results of the second MANOVA and will compare the major context groups that were mentioned above to the reference sibilants /s/ and /ʃ/. The p-values are always $p \leq 0.001$ and will not be mentioned unless the p-values differ. All values mentioned are differences between the average values of the variables. In the following sections, it will not be mentioned that a value is an average.

4.1 Main effects of the independent variables

For the independent variable *AUS_CAN*, all dependent variables (center of gravity, standard deviation, skewness and kurtosis) had highly significant main effects. Both the center of gravity ($F[1,865] = 92.033$) and the standard deviation ($F[1,865] = 131.459$) were significantly lower in the Canadian data. Skewness ($F[1,865] = 514.666$) and kur-

tosis (F[1,865] = 29.125) had significantly higher values in the target words produced by Canadian speakers than in the Australian speakers' productions. The values of all four dependent variables can be seen in Figures 3 (a)-(d) below.

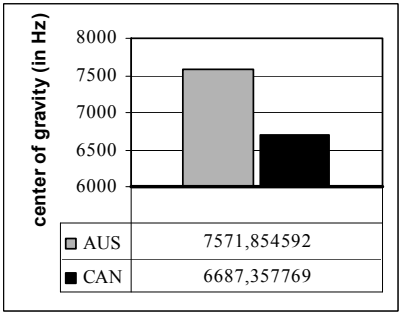


Figure 3(a). Average center of gravity of the sibilants in the Australian and Canadian data in Hz.

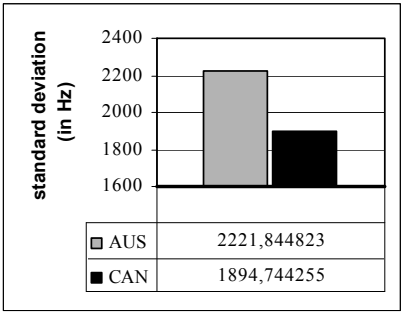


Figure 3(b). Average standard deviation of the sibilants in the Australian and Canadian data in Hz.

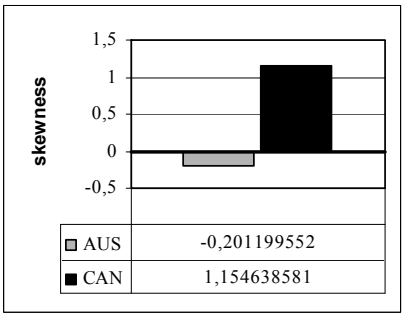


Figure 3(c). Average skewness of the sibilants in the Australian and Canadian data.

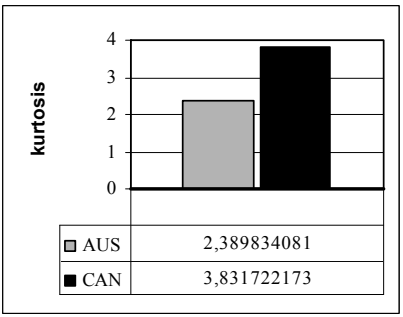


Figure 3(d). Average kurtosis of the sibilants in the Australian and Canadian data.

WithWB_noWB (F[1,865] = 12.082) and *WithT_noT* (F[1,865] = 15.982) both had highly significant effects for the standard deviation of the center of gravity. The standard deviation was significantly higher in words in the contexts *NoWB* and *NoT*. The *NoWB* context is one that was first included in the process of /s/-palatalization, *NoT* was not. The mean values of the standard deviation are depicted in Figures 4 and 5.

Rs_sr and *WCont_noCont* both had the same patterns and highly significant effects for center of gravity, standard deviation and kurtosis. The center of gravity was lower in *Sr* (F[1,865] = 14.939) as well as *WCont* (F[1,865] = 77.748). The standard deviation was higher in *Sr* (F[1,865] = 33.899) and *WCont* (F[1,865] = 29.209); and the kurtosis was again lower in *Sr* (F[1,865] = 11.684) and *WCont* (F[1,865] = 12.713). *Sr* and *WCont* are two of the contexts that are expected to be palatalized more severely. The values can be seen in Figures 6 (a)-(c) and 7 (a)-(c) below.

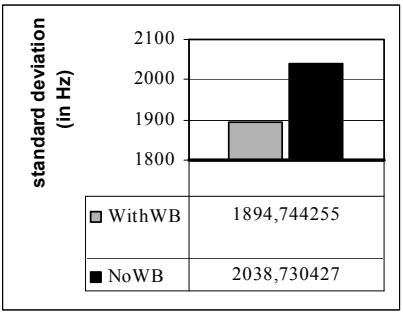


Figure 4. Average standard deviation of the sibilants in words in the context WithWB_noWB in Hz.

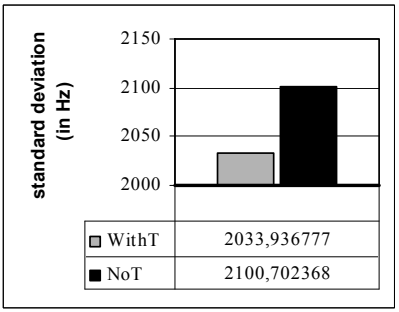


Figure 5. Average standard deviation of the sibilants in words in the context WithT_noT in Hz.

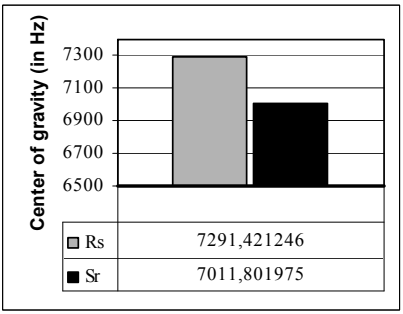


Figure 6(a). Average center of gravity of the sibilants in words in the context Rs_sr in Hz.

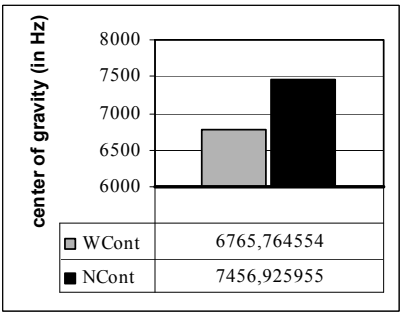


Figure 7(a). Average center of gravity of the sibilants in words in the context WCont_nCont in Hz.

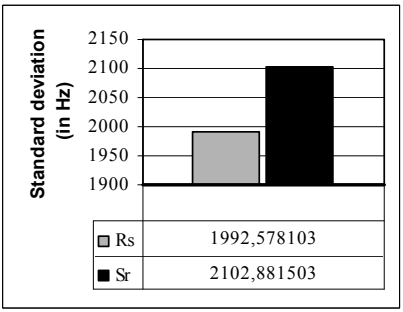


Figure 6(b). Average standard deviation of the sibilants in words in the context Rs_sr in Hz.

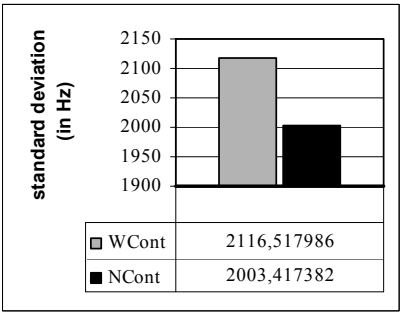


Figure 7(b). Average standard deviation context of the sibilants in words in the WCont_nCont in Hz.

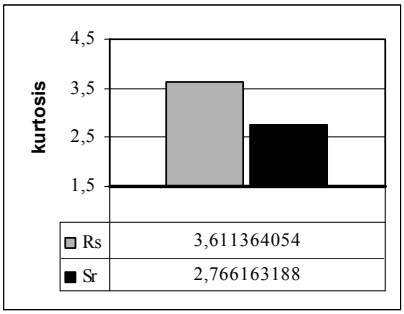


Figure 6(c). Average kurtosis of the sibilants in words in the context Rs_sr.

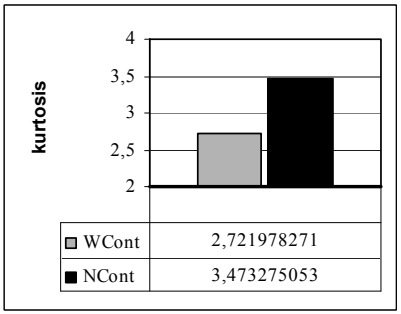


Figure 7(c). Average kurtosis of the sibilants in words in the context WCont_nCont.

4.2 The independent variables' interactions

Six interactions of the independent variables were highly significant with $p \leq 0.001$. Of the interactions, five were two-way interactions, and four were three-way interactions.

In the interaction *NoWB_withWB*Rs_sr*, the dependent variable skewness ($F[1,865] = 14.397$) was highly significant. However, skewness was not even close to being significant for either of the independent variables alone. The interaction, shown in Figure 8, had higher skewness values in the context *NoWB* and lower values in the *WithWB* context when the /s/ was preceding the /r/. The opposite was the case when the /s/ was following /r/. The difference between the skewness values in the context *WithWB_NoWB* in interaction with the context *Sr* is extremely striking.

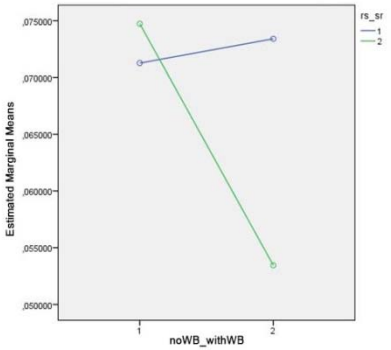


*Figure 8. Estimated marginal means of the skewness of NoWB_withWB*Rs_sr.*

The interaction *Rs_sr*WithT_noT*, depicted in Figure 9 below, was highly significant for the standard deviation ($F[1,865] = 11.647$). The standard deviation was also highly significant in the interaction R*s_sr*WCont_nCont* ($F[1,865] = 11.548$), shown in Figure

10. The difference in standard deviation was much greater in the sequence /sr/ than /rs/ for both interactions. This difference was also a lot greater in words without /t/ in the cluster than in words without /t/. The difference between *Rs_sr* was greater in words with contact between /s/ and /r/ than in words without. The standard deviation was also found to be highly significant for all three independent variables involved in these two interactions.

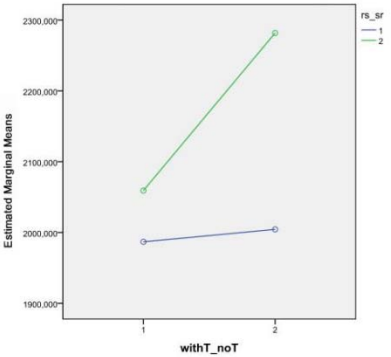


*Figure 9. Estimated marginal means of the standard deviation in the interaction Rs_sr*WithT_noT.*

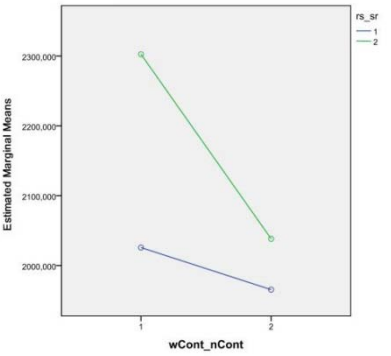


*Figure 10. Estimated marginal means of the standard deviation in the interaction Rs_sr*WCont_nCont.*

*WithT_noT*WCont_nCont* was the last two-way interaction that was highly significant. Both the center of gravity (F[1,865] = 10.948) and its standard deviation (F[1,865] = 18.363) were highly significant. The center of gravity, whose interaction is shown in Figure 11, was lower in the context *WithT* when there was no contact between /s/ and /r/ and it was higher with contact. The opposite was the case for words in the context *NoT*. There was no difference in standard deviation between *WithT* and *NoT* in the interaction with *NCont*, but the standard deviation was a lot higher in words with contact between /s/ and /r/ and without /t/ in the cluster. This interaction can be seen in Figure 12. Both independent variables had highly significant main effects for the standard deviation when they were not interacting; *WithT_noT* also had a highly significant effect for the center of gravity. The independent variable *WCont_nCont*, on the other hand, was far from being significant concerning center of gravity (F[1,865] = 0.208, p = 0.648).

The last two interactions were three-way interactions. The interaction of *AUS_CAN*Rs_sr*WCont_nCont* was highly significant for center of gravity (F[1,865] = 20.389). In the simple analyses without interactions, center of gravity also proved to be highly significant for all three independent variables. The interaction *NoWB_withWB*Rs_sr*WithT_noT* was significant for the standard deviation (F[1,865] = 12.185) which was also highly significant for all three independent variables without interaction.

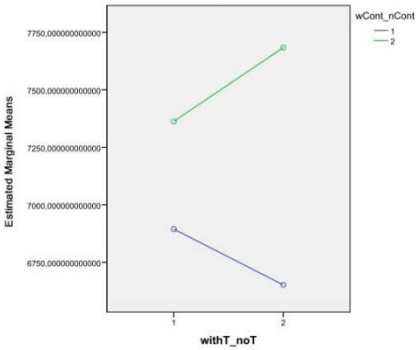


*Figure 11. Estimated marginal means of the center of gravity in the interaction WithT_noT*WCont_nCont.*

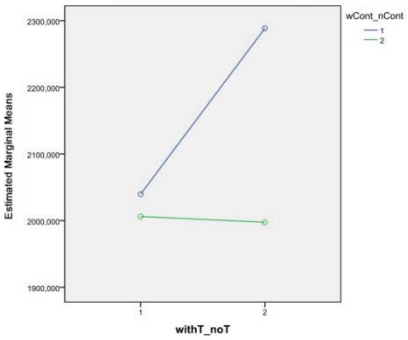


*Figure 12. Estimated marginal means of the standard deviation in the interaction WithT_noT*WCont_nCont.*

4.3 Comparison to reference sibilants

For this part of the analysis, two MANOVAs comparing the major context groups with the reference sibilants have been computed. One featured only data from the Canadian speakers and the other from the Australian speakers. The major context groups are the same independent variables as above: *NoWB_withWB*, *Rs_sr*, *WCont_nCont* and *WithT_noT*. Results of skewness will be looked at briefly in section 4.3.1, but the focus will be on the results regarding center of gravity, presented in section 4.3.2. These two variables were especially sensitive towards differences in degree of palatalization between the sibilants in the context groups and the reference sibilants.

Center of gravity had significant effects ($p \leq 0.001$) for all contexts and for both varieties of English. Skewness had significant effects ($p \leq 0.001$) only in the Australian data, but not in the data of the Canadian speakers. Nonetheless, in Australia, skewness proved to be a significant indicator of differences between all context groups and reference sibilants.

4.3.1 Skewness

In section 2.1.1 it was mentioned that positive or greater skewness was associated with a more /ʃ/-like sound. It can be assumed that the higher the skewness values are compared to reference /s/, the more palatalized the sounds are in a particular context. The results presented here will be discussed in section 5.

In the data of the Australian speakers, the sibilants of all contexts followed the same patterns in comparison to the reference sibilants, with two exceptions. In the context *NCont*, the skewness of the target sibilants was 0.014 lower than reference /s/ and even lower compared to reference /ʃ/. The skewness of the sibilants in the context *Rs* was 0.054 lower when compared to reference /s/. Both differences are fairly small and not

significant, suggesting a /s/-like sound rather than a palatalized variant in both contexts. In all other contexts, the skewness values were between 0.047 and 0.250 higher than those of reference /s/ and between 1.244 and 1.507 lower than those of reference /ʃ/. Except for the non-significant difference in the context *Rs* that also falls into that range, all other differences in skewness were highly significant.

4.3.2 Center of gravity

The dependent variable center of gravity was highly significant both in the Australian and the Canadian data. There were differences in frequencies between all variables looked at for each analysis, offering the chance to hypothesize on possible degrees of palatalization in the different context groups. The results will be interpreted in section 5.1. In general, the post-alveolar fricative /ʃ/ has a lower center of gravity than the alveolar fricative /s/. Should the investigated sibilants have lower centers of gravity than reference /s/, this could be interpreted as an indicator for palatalization.

In the Australian and the Canadian data, the comparison of the sibilants of the investigated contexts and the reference sibilants showed consistent patterns throughout the sample. The center of gravity was always significantly lower than reference /s/ and significantly higher than reference /ʃ/. This pattern had two exceptions: In the comparison of Australian sibilants, the center of gravity was with 180 Hz lower in the context *NCont* than reference /s/ and lower by 209 Hz in the context *Rs*. Both differences were non-significant which suggests that there is no palatalization in this context.

When the centers of gravity of the different contexts were compared to each other, the contexts that were first recognized as being part of the /s/-palatalization process—*Sr*, *NoWB*, *WithT* and *WCont*—always had lower centers of gravity than the inverse context. However, this difference was only highly significant in the comparison between *Rs* and *Sr* in the Australian data and the comparison between *WCont* and *NCont* in both Canadian and Australian English. *NoWB_withWB* and *WithT_noT* do not differ significantly when compared to each other. It is important to mention the non-significant differences in this case. A non-significant difference in center of gravity between two opposite contexts suggests that both contexts are palatalized to practically the same degree, something that will be addressed further in the discussion in section 5.

Table 6 below shows the differences in center of gravity in the different contexts, both compared to themselves and the reference sibilants, for the Australian data. Non-significant results are shown in parentheses. Table 7 shows the same for the data of the Canadian speakers. By looking at the comparison of the contexts with reference /s/, the values suggest that the sibilants in the Canadian sample are palatalized further. The centers of gravity have a larger difference to reference /s/ in Canadian English than in Australian English. However, the center of gravity is always closer to that of reference /s/, suggesting that there is never complete palatalization to prototypical /ʃ/.

Context	lower than /s/	higher than /ʃ/	lower than inverse context	higher than inverse context
NoWB	1172 Hz	1930 Hz	(38 Hz)	—
WithWB	1133 Hz	1969 Hz	—	(38 Hz)
Rs	1064 Hz	2039 Hz	(156 Hz)	—
Sr	1219 Hz	1883 Hz	—	(156 Hz)
WithT	1239 Hz	1863 Hz	(232 Hz)	—
NoT	1008 Hz	2094 Hz	—	(232 Hz)
Wcont	1585 Hz	1518 Hz	826 Hz	—
Ncont	758 Hz	2344 Hz	—	826 Hz

Table 6. Differences in center of gravity in the different contexts, compared to themselves and to the reference sibilants /s/ and /ʃ/ in the data of the Australian speakers. Parentheses indicate non-significant differences.

Context	lower than /s/	higher than /ʃ/	lower than inverse context	higher than inverse context
NoWB	1172 Hz	1930 Hz	(38 Hz)	—
WithWB	1133 Hz	1969 Hz	—	(38 Hz)
Rs	1064 Hz	2039 Hz	(156 Hz)	—
Sr	1219 Hz	1883 Hz	—	(156 Hz)
WithT	1239 Hz	1863 Hz	(232 Hz)	—
NoT	1008 Hz	2094 Hz	—	(232 Hz)
Wcont	1585 Hz	1518 Hz	826 Hz	—
Ncont	758 Hz	2344 Hz	—	826 Hz

Table 7. Differences in center of gravity in the different contexts, compared to themselves and to the reference sibilants /s/ and /ʃ/ in the data of the Canadian speakers. Parentheses indicate non-significant differences.

5. Discussion

In the following section 5.1, the results of section 4 will be discussed and interpreted in terms of what the results might indicate for the process of /s/-palatalization in this sample. Problems faced during the analyses and interpretations as well as an idea of what would be necessary in terms of further research will be discussed in section 5.2. The discussion will revert back to the four questions posed at the end of section 2:

- Which contexts lead to more palatalized realizations of /s/?
- Is /s/-palatalization a gradual process with intermediate forms?
- Is /s/-palatalization stronger in the sample of Canadian or Australian speakers?
- Which of the possible triggers of palatalization is more relevant?

Some of the contexts included in this study can be found in some of the words that are most prominently acknowledged to participate in the assimilation process of /s/-palatalization. These contexts include *WithT*, *Sr*, *NoWB* and *WCont* as in words like "street" or "strong". Because these contexts are the ones that seem to be involved in the process for a longer time, these contexts are assumed to have sibilants that are palatalized more severely. To be classified as having "more severe palatalization," the sibilants in these contexts need to have smaller kurtosis, greater standard deviation, higher skewness and lower center of gravity than the opposite contexts.

5.1 Interpretation of the results

Standard deviation and kurtosis seem to be inconclusive in some cases of the first MANOVA that was conducted. Expected was that the contexts *CAN*, *WithT*, *NoWB*, *Sr* and *WCont* were palatalized more and therefore were expected to have greater standard deviation and smaller kurtosis. That was the case for *NoWB*, *Sr* and *WCont*, suggesting that these contexts might have sibilants that were palatalized more than their opposite contexts. For the contexts *CAN* and *WithT*, standard deviation was lower than in their inverse contexts, suggesting they might not be palatalized as much.

The center of gravity and skewness, on the other hand, connect the acoustical framework and the expectations from above nicely, even though center of gravity was only significant for *AUS_CAN*, *Rs_sr* and *WCont_nCont*. The sibilants in the context *CAN* had higher skewness and lower center of gravity, suggesting more severe palatalization in Canadian English. In the contexts *Sr* and *WCont*, the sibilants had lower center of gravity, also suggesting more palatalization compared to their inverse contexts and at the same time hinting at more than one realization of the sibilant in terms of palatalization. The fact that in this analysis the difference in center of gravity in the contexts *WithWB_noWB* and *WithT_noT* was non-significant might suggest that in each context group, the contexts are palatalized to the same degree. This already tentatively answers the first question posed above.

Regarding the interactions between the independent variables, it seems unclear what impact the results have on the interpretation of the results without interaction. Kurtosis never seems to show significant effects. Most of the highly significant interactions involved independent variables where the examined dependent variables were also highly significant without interactions. All but one of the highly significant interactions involved the independent variable *Rs_sr*, suggesting that significances of this variable are mostly connected to other significant effects. The main effects of this independent variable on its own might not be significant without the other variables. It also appears as if some of the combinations of the contexts were more important for palatalization. For example, the center of gravity was lower when the two contexts *WithT* and *NCont* were interacting, suggesting more severe palatalization in words of these two contexts.

The most answers can be given by looking at the results from the second MANOVA, the comparison to the reference sibilants. The first question, which of the contexts had sibilants that were palatalized more severely, can be answered by looking at the results both from the comparison of skewness and center of gravity. The contexts mentioned above that were expected to be palatalized more severely—*WithT*, *NoWB*, *Sr* and *WCont*— had larger differences to reference /s/ than the opposite contexts. They

had lower center of gravity and in the Australian data the skewness was higher than /s/. Two of the contexts that are not expected to be palatalized as severely—*Rs* and *NCont*—even seem to be not palatalized at all in the Australian data. Their centers of gravity were not significantly lower than /s/ and the skewness was lower or almost the same as /s/. There was no difference in center of gravity between sibilants in words with and without /t/ or in words with and without word boundaries between /s/ and /r/, suggesting that the presence of a word boundary or the absence of /t/ do not figure in as influences on the process.

The second question was whether or not intermediate forms between /s/ and /ʃ/ could be found. The comparison between the different contexts gives some clues that there seem to be intermediate forms in all context groups. There was no significant difference in center of gravity between the contexts in the group *WithT_noT* as well as *NoWB_withWB*. That suggests that the sibilants in both contexts in each group are palatalized to the same degree. At the same time the values of the center of gravity always stay closer to reference /s/ than reference /ʃ/ which means that, in this sample, there is never complete palatalization. That means that palatalization is not a phonemic change like Shapiro (1995: 101) concluded. The contexts *Rs_sr* and *WCont_nCont* show different patterns in Australian and Canadian English. In the Canadian data, the contexts *Rs* and *Sr* do not differ significantly from each other, suggesting that the sibilants in both contexts are palatalized to the same degree. In the Australian data, however, there is a significant difference of almost 400 Hz between both contexts. One of the contexts, *Rs*, does not have a significant difference to the center of gravity of reference /s/. That means that in the Australian English in this sample, the sibilants in words of the context *Sr* are palatalized while the ones in the context *Rs* are not. This is the same for the contexts *WCont* and *NCont* in the Australian data. *NCont* showed no difference in center of gravity to reference /s/ which means that there is most likely no palatalization in this context. *WCont*, on the other hand, is palatalized. The indication of no palatalization in the contexts without contact between /s/ and /r/ as well as the contexts where /r/ preceded /s/ was also found in the analysis of the skewness values as well as the interpretation of the results of the first MANOVA above. In the data of the Canadian speakers in this sample, both contexts differ significantly from each other but also from reference /s/. That suggests that the sibilants in both contexts are palatalized, but those in the context with contact between /s/ and /r/ are palatalized more severely.

The third question was if the sibilants were more severely palatalized in the data of the Australian or the Canadian speakers. In general, the Canadian speakers in this small sample seem to palatalize the sibilants more severely than the Australian speakers, and also in more contexts. The Australian speakers' sibilants usually had a center of gravity that was only between 300 and 720 Hz below that of reference /s/, which is a significant difference, but not that much in terms of center of gravity. This indicates only slight palatalization. The Canadian speakers' sibilants, on the other hand, were between 750 and 1600 Hz lower than reference /s/ in terms of center of gravity, which is more severe. The reason for this could be that a lot of the target words had non-prevocalic /r/, a feature not present in Australian English, and all values were taken together for the analysis. Especially in the context *Rs*, one of the contexts found to not involve palatalization in the Australian data, a lot of the target words include non-prevocalic /r/. Therefore, the opportunity for palatalization was reduced. An overview of the differences in

center of gravity between the contexts, reference sibilants as well as Australian and Canadian speakers is provided in Figure 13 below.

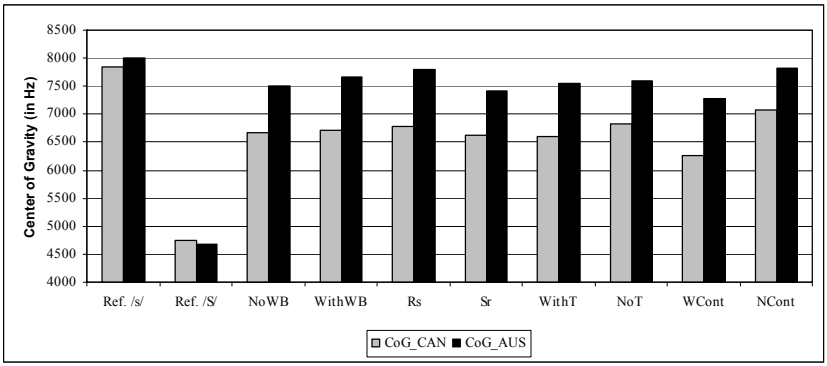


Figure 13. The average centers of gravity of the reference sibilants /s/ and /ʃ/ and the different contexts both in the Australian and the Canadian data. It is obvious that there is never complete palatalization and that the palatalization is more severe in the data of the Canadian speaker.

The difference between Australian and Canadian English in terms of /s/-palatalization also might be one answer for the fourth question, asking for the most probable trigger for this assimilation process. Taking especially the center of gravity analysis into account, it seems reasonable that the trigger for /s/-palatalization is most likely the retroflex /r/, as Shapiro (1995) suggested. The interpretation of the differences between Australian and Canadian English, offered above, was that the reason might be the lack of non-prevocalic /r/ in Australian English, leading to less severe /s/-palatalization in comparison to Canadian English. Another proof might be that there was no difference in center of gravity in the context *WithT_noT*, there seems to be the same amount of palatalization in both contexts. If the sibilants in words without /t/ are palatalized the same way as in words with /t/, this would suggest that the /t/—affricated or not—has no influence on the assimilation process. Shapiro called palatalization a *"case of distant assimilation"* (Shapiro 1995: 101) because the /r/ is not adjacent to the /s/. The results of this study seem to prove this. However, the assimiliation is not categorical. It is not a phonological process. The sound never becomes a prototypical [ʃ] but always stays closer to [s]. That is in line with the more modern view on assimilation, defining the resulting sounds as being intermediate forms (cf. Niebuhr and Meunier 2011). The sounds involved in assimilation at a distance would be expected to behave in the same way when they are adjacent to each other, which seems to be the case in the small sample used for this study: There is no significant difference in center of gravity between sibilants in words with or without /t/. The /t/ seems to play no important role in degree of palatalization, meaning that on average there is as much palatalization in words like "post room" as in words like "classroom".

5.2 Problems and future research

There were some results seemingly contradicting each other. However, the majority of the results from the analyses seem to support both the acoustical framework presented in section 2.1 and the expectations one would have regarding the contexts that would most likely have more severely palatalized sibilants. But it is hard to determine why the interactions between the independent variables happen and to what extent that influences the significances of the dependent variables when there is no interaction between the independent variables.

The target words from the texts used to elicit the data were not evenly distributed which might have interfered with the statistical analyses. There were a lot more target words with /str/ contexts than with /sr/ contexts. The entire sample of speakers was also too small to make definite statements on the exact nature of /s/-palatalization in a particular variety of English. The age distribution was too variable to draw conclusions on social factors of the process. In further studies, the sample size should be enlarged, and every context should be represented by the same number of target words. The reference sibilants ranged a lot of different contexts—word-initial, word-final, different vowels preceding and following, but also different prosodic contexts. That way a wider range of values could be assembled, covering differences in each sound. There were two further contradictory problems that are both important to consider in future research. The many parameters used for the analyses—five independent variables, four dependent variables—made the amount of data very large and confusing, yet all parameters seemed to be crucial for the study. However, one more independent variable should be included: whether a word had non-prevocalic /r/ or not. That parameter might be another indicator for /r/ being the trigger of /s/-palatalization or not. Another observation that should be noted in the data is whether or not the /t/ is affricated.

The last but maybe most important thing to mention here is that for this study, the sibilants and reference sibilants of all speakers have been put together before the analyses and comparisons. In future research, the sibilants of one speaker should be compared directly to that same speaker's reference sibilants. This was unfortunately not possible for the present study. However, since /s/-palatalization is to a large degree speaker-specific, this comparison would be crucial. In this study, only the general nature of palatalization could be investigated. Taking the inter-speaker variability into account, a much clearer picture could be painted.

6. Conclusion

At the end of section 2.2.3, four questions were introduced that were also addressed in further detail in the discussion in section 5. The answers found there will be summarized in this section. Each question can be answered, even though these answers might only reflect this particular small sample. In the future, a larger scale study with more speakers and balanced contexts should be performed.

The first question was which segmental contexts lead to realizations of /s/ that are more palatalized. After reviewing the results of the analyses, it seems as if the contexts expected to have more palatalized sibilants—in words with /s/ preceding /r/ and in

words with contact between /s/ and /r/—were in fact more strongly palatalized. There seemed to be no difference in degree of palatalization in the contexts *WithT_noT* and *NoWB_withWB*. However, in the Canadian data there were more contexts palatalized than in the Australian data: In Canadian English, there was no distinction between the sequence /sr/ and /rs/ like there was in Australian English, and sibilants in both *WCont* and *NCont* contexts were palatalized, but to differing degrees.

In the second question it was asked if there were intermediate forms of the sibilant, produced in between /s/ and /ʃ/. In general, there was on average one intermediate form in each context group. This intermediate form was shared by both contexts in some cases—as in words with and without word boundaries between /s/ and /r/, in words with and without /t/ as part of the cluster, and, in the data of the Canadian speakers, in words with /s/ either following or preceding /r/. In these contexts, there was no statistical difference between the sibilants in any of the contexts. In other cases only one context was palatalized and the other was not: In the Australian data, only sibilants in words with contact between /s/ and /r/ and sibilants in words with /s/ preceding /r/ were palatalized. Only the context group *WCont_nCont* seemed to have two intermediate forms in the data of the Canadian speakers. There has also never been complete palatalization—the sibilants always remained closer to /s/ than to /ʃ/, but far enough away from /s/ to be interpreted as palatalization.

Another question was concerned with the possible triggers for /s/-palatalization and which seem to be probable. The influence of preceding or following vowels (Durian 2007) could not be investigated. The suggestion of Lawrence (2000), who sees the trigger as affricated /t/, does not seem to be in line with the results. Since there was also palatalization in words without /t/—and without a significant difference between both contexts as well—the /t/ does not appear to be an active part of the process, at least not in the present data. Therefore, the most probable trigger is the retroflex /r/, as Shapiro (1995) suggested. Evidence for this hypothesis is the same as the evidence against affricated /t/ as the trigger: Palatalization also takes place in /sr/ clusters. Another indicator for /r/ as the trigger seems to be the answer to the last question—whether there is a significant difference between Australian and Canadian English. The answer is: yes. The Canadian speakers palatalized the sibilants further than the Australian speakers. A reason for this could be that a lot of the target words included had non-prevocalic /r/ which is not a feature of Australian English. Therefore, fewer possibilities for palatalization could be found in the data of the Australian speakers, probably accounting for the difference between both varieties. This could also be seen as evidence that /r/ is the most likely trigger for /s/-palatalization.

There is still a lot of room for further research—be it with a larger sample, more varieties of English, maybe even studies on how speakers of English as a second language realize /s/ in the contexts discussed here. It is still not clear if the /r/ really is the trigger—even though it seems to be the only probable trigger in this study. A much more detailed investigation could be carried out on the nature of the intermediate forms. All in all, this assimilation process appears to be present around the globe in different varieties of English. The linguistic contexts in which /s/-palatalization occurs also seem to widen. Palatalization is very prominent and there is a lot of research still to be done.

7. Acknowledgements

First of all, I would like to thank Oliver Niebuhr (Kiel University) for his support with the statistical analyses, his comments and the incredible mentoring during the working process on this—now revised—Bachelor thesis. Thank you to Meghan Clayards (McGill University, Montréal, Canada) and Hywel Stoakes (University of Melbourne, Australia) for the recording of the data used for the study and Karo Kress for the design of the elicitation texts. Thank you to John Peterson (Kiel University), Lena Marr and Karin Bubelach for corrections and suggestions given during the different stages of the paper.

8. References

ALTENDORF, U. 2003. Estuary English – Levelling at the Interface of RP and South-Eastern British English. Tübingen: Gunter Narr Verlag.

BASS, M. 2009. "Street or Shtreet? Investigating (str-) Palatalisation in Colchester English." Estro: Essex Student Research Online 1, 10-21. http://www.essex.ac.uk/ journals/estro/documents/issue1/FullIssue.pdf, accessed July 21, 2014.

BOBERG, C. 2010. The English language in Canada – Status, history and comparative analysis. Cambridge: Cambridge University Press.

BOERSMA, P. / D. WEENINK. 2014. Praat: doing phonetics by computer. Version 5.3.66, retrieved 18 March 2014 from http://www.praat.org/

CRYSTAL, D. 2003. A Dictionary of Linguistics and Phonetics. Oxford: Blackwell.

DILLEY, L.C. / M.A. PITT. 2007. A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. Journal of the Acoustical Society of America 122, 2340-2353.

DURIAN, D. 2007. Getting [ʃ]tronger Every Day? More on Urbanization and the Sociogeographic Diffusion of (str) in Columbus, OH. University of Pennsylvania Working Papers in Linguistics 13, 65-79.

HEID, S. / S. HAWKINS. 2000. An Acoustical Study of Long Domain /r/ and /l/ Coarticulation. Proc. 5th International Conference on Speech Signal Processing, Kloster Seeon, Germany, 77-80.

IBM CORP. 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

JOHNSON, K. 2012. Acoustic and Auditory Phonetic*s*. Chichester: Wiley-Blackwell.

JONGMAN, A. / R. WAYLAND / S. WONG. 2000. Acoustic Characteristics of English Fricatives. Journal of the Acoustical Society of America 108, 1252-1263.

LABOV, W. 1984. Field Methods of the Project on Language Change and Variation. In: John Baugh / Joel Schezer (Eds), Language in Use (pp. 28-53). Englewood Cliffs: Prentice Hall.

LAWRENCE, W. 2000. /str/ → /ʃtr/: Assimilation at a distance? American Speech 75, 82-87.

NIEBUHR, O. / C. MEUNIER. 2011. The phonetic manifestation of French /s#ʃ/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. Phonetica 68, 133-160.

OLIVE, J.P. / A. GREENWOOD / J. COLEMAN. 1993. Acoustics of American English Speech: A Dynamic Approach. New York: Springer.

PIERCY, C. 2012. A Transatlantic Cross-Dialectal Comparison of Non-Prevocalic /r/. University of Pennsylvania Working Papers in Linguistics 18, 77-86.

RUTTER, B. 2011. Acoustic Analysis of a Sound Change in Progress: The Consonant Cluster /stɹ/ in English. Journal of the International Phonetic Association 41, 27-40.

SHAPIRO, M. 1995. A Case of Distant Assimilation: /str/ → /ʃtr/. American Speech 70, 101-107.

2 Prosodische Parameter im Poetry Slam – Eine Pilotstudie anhand von Beispielen

Maike Thießen, M. Ed.
maike_thiessen@yahoo.de

In der vorliegenden Master Arbeit werden erste Studien zum Thema „Prosodie im Poetry Slam“ durchgeführt. Das erst seit den 90er Jahren populäre Veranstaltungsformat ermöglicht den Zugang zu einem jungen Publikum und generiert neue Formen mündlicher Dichtung. Daher ist es für die Literatur- und Sprachwissenschaft von großer Bedeutung.

Der erste Teil der Arbeit stützt sich auf eine theoriebasierte Exploration und stellt den aktuellen Forschungsstand zu Prosodie und ihren Parametern dar. Außerdem werden Aussagen, die auf die auditiven Eigenschaften von Slam Poetry hindeuten, aus wissenschaftlichen Veröffentlichungen extrahiert.

Der zweite Teil dokumentiert eine Umfrage, mit welcher getestet wurde, ob Slam Poetry aufgrund seiner besonderen prosodischen Eigenschaften ausschließlich anhand von technisch-delexikalisierten Vorträgen identifizierbar ist. Es folgt eine prosodische Analyse der delexikalisierten Hörbeispiele, durch die bestehende Hypothesen geprüft und neue generiert werden. Eine abschließende Betrachtung fasst die Hypothesen zusammen und deutet neue Forschungsaspekte an.

1. Einleitung

> „Auch auf der Bühne gehen die Slammer unkonventionell vor. Sie schreien, flüstern, stöhnen und fluchen- die Dichtung wird Teil einer Performance.“ (Westermayr 2010:153)

Stefanie Westermayr (2010) beschreibt in dem Buch „Poetry Slam in Deutschland. Kulturelle Praxis einer multimedialen Kunstform“ eine literarische Bewegung, die sich seit einigen Jahren mehr und mehr im deutschen Kulturbetrieb etabliert. Das Veranstaltungsformat „Poetry Slam“ spricht vor allem ein junges Publikum an und eröffnet Poeten jeden Alters die Möglichkeit, ihre Texte öffentlich zu präsentieren. Wie in dem Zitat Westermayrs angedeutet, zeichnet die Texte, die für dieses Event geschrieben

werden, nicht nur das Alter der Protagonisten aus. Neben den Darstellern selbst war bereits in einigen wissenschaftlichen Arbeiten von dem besonderen Klang der Vorträge die Rede und es wurden Vermutungen angestellt, wie dieser zustande kommt. Eine konkrete Analyse der nicht lexikalischen sprachlichen Gestaltungsmittel wurde bislang nicht vorgenommen.

2. Theoriebasierte Exploration

2.1 Prosodie

2.1.1 Eingrenzung der Parameter

Baldur Neuber (2002) beginnt seine Arbeit zu prosodischen Formen und ihrer Funktion mit einem Kapitel zu dem aktuellen Forschungsstand zur Prosodie. Er macht gleich zu Anfang darauf aufmerksam, dass eine einfache Verwendung der in diesem Kontext zu nennenden Begrifflichkeiten zu Missverständnissen führen kann, da in der Literatur nur bedingt Einigkeit darüber besteht, wie prosodische Phänomene wie zum Beispiel die Intonation begrifflich einzuordnen sind.

> „Die Termini Prosodie, Suprasegmentalia und Intonation sind in der Fachliteratur divers besetzt. Unterschiede bestehen sowohl zwischen verschiedenen fachspezifischen (insbesondere phonetischen, linguistischen und psychologischen) Ansätzen, als auch innerhalb der einzelnen Disziplinen.“ (Neuber 2002:15)

Aus dieser Problematik schließt Neuber, dass eine einheitliche Definition der Begriffe *Intonation*, *Suprasegmentalia* und *Prosodie* zu dem aktuellen Zeitpunkt nicht möglich ist. Er stellt verschiedene Definitionsversuche aus der Fachliteratur vor, bei denen grundsätzlich zwischen einer eng und einer weit gefassten Interpretation unterschieden wird. Neuber betont, dass eine terminologische Festlegung nur in Bezug auf den konkreten Forschungsgegenstand möglich ist, und informiert darüber, dass die Begriffe in seiner Arbeit synonym verwendet werden und somit der enge Intonationsbegriff ausgeschlossen wird (Neuber 2002:20). Robert Ladd löst das Problem, indem er seine eigene Definition von Intonation an den Anfang seines Buches setzt.

> „One of the many difficulties of writing on the subject of intonation is that the term means different things to different people. It is therefore appropriate, right at the beginning of the book, to offer my own definition of intonation, or at least to try to delimit the area I propose to cover.“ (Ladd 1996:6)

Christiane Miosga (2006) interpretiert Prosodie in dem Buch „Habitus der Prosodie. Die Bedeutung der Rekonstruktion von personalen Sprechstilen in pädagogischen Handlungskontexten“ aus einer funktionalen Perspektive. Als Ergebnis des Habitus, der

gesamtkörperlichen Selbstpräsentation, gehen die Funktionen der Prosodie über die phonetische Realisierung hinaus und sind bedingt durch innere Einstellungen des Sprechers und die interpersonelle Einbettung in eine Kommunikation. Damit begründet Miosga auch die schwer einzuordnenden Merkmale der Prosodie. In dem Kapitel zur Definition des Forschungsgegenstandes legt Miosga (2006:50) drei Fragen zugrunde: *„Welche Sprechgestaltungsmittel werden unter dem Begriff der Prosodie subsumiert? Welche interpersonellen Funktionen können ihnen zugeordnet werden? Welche Klassifikationskriterien ergeben sich auf deskriptiver, präskriptiver und interpretierender Ebene?“*

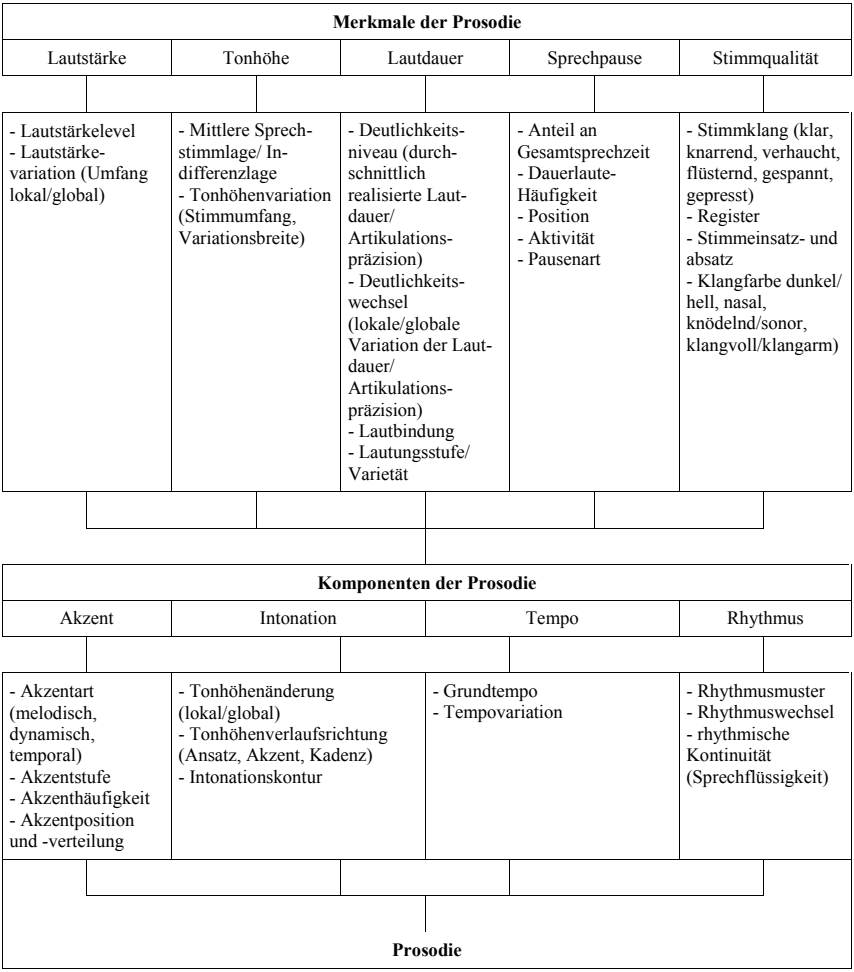


Abbildung 1. Parameter der Prosodie: Einteilung in Merkmale und Komponenten (modifiziert nach Hargrove und McGarr 1994:15); (Miosga 2006:58).

Um diese Fragen zu beantworten, zieht Miosga Erkenntnisse aus verschiedenen Disziplinen, wie der Linguistik, Kommunikationstheorie und Sprechwissenschaft, heran. Sie lässt dabei die Schwierigkeiten, die Neuber beschreibt, nicht außer Acht und merkt an, dass aus funktionaler Perspektive viele Parameter der Prosodie aus verschiedenen Wissenschaften zu berücksichtigen sind. Abbildung 1 zeigt eine Übersicht aller von Miosga aufgezählten Parameter.

Im Grammatik-Duden (2009) ist für den Begriff *Intonation* folgende Definition zu finden: *„Mit Intonation bezeichnet man die melodische Gestalt einer Äußerung. Sie ergibt sich aus der Wahrnehmung von Tonhöheneigenschaften durch das Gehör“* (S.95). Ähnlich wie bei Miosga wird die Intonationskontur genannt, das heißt eine Abfolge von abstrakten Tönen, welche in lexikalische und intonatorische aufgeteilt sind. Erstere beschreiben phonologische Laute innerhalb von Wortformen, welche bedeutungskonstituierend der Unterscheidung verschiedener Wortformen dienen. Letztere sind von Äußerungsbedeutungen gelöst und *„nicht an Einheiten der lexikalischen Ebene gebunden“* (Duden 2009:95).

2.1.2 Auflistung aller Parameter mach Miosga

Um für die vorliegende Arbeit eine Auswahl an Parametern zu ermöglichen, ist es notwendig, zunächst die potenziellen aufzulisten. Dazu wird die Arbeit von Christiane Miosga verwendet, welche diese ausführlich behandelt. Dabei ist zu erwähnen, dass Miosga zur Feststellung des persönlichen Sprachstils die lautlichen Signale nicht als akustisch-physikalische Erscheinungen oder in ihrer physiologischen Produktion, sondern als auditive Empfindungsgrößen verwendet. Wie in Abbildung (1) abzulesen ist, unterteilt Miosga die prosodischen Bestandteile in *Parameter* und *Merkmale*.

> „Die Komponenten Akzent, Intonation, Tempo und Rhythmus bestimmen durch ihr Zusammenspiel in verschiedenen Kombinationen die prosodische Gestaltung. Sie bilden Ganzheitskategorien, die wiederum durch das Zusammenspiel von messbaren Einzelmerkmalen realisiert werden. Die Parameter Lautstärke, Tonhöhe, Lautdauer, Sprechpause und Stimmqualität sind also Merkmale der Komponenten Akzent, Intonation, Tempo und Rhythmus, die wiederum durch ihr Zusammenspiel die prosodische Gestaltung konstituieren.“ (Miosga 2006:65f.)

Mit dieser Differenzierung verdeutlicht Miosga die Tatsache, dass die prosodischen Parameter nicht unabhängig voneinander zu betrachten sind, sondern in ständiger wechselseitiger Beziehung zueinander stehen.

2.1.2.1 Lautstärke

Das Merkmal der *Lautstärke* entspricht dem akustischen Parameter der *Intensität*, da dieser durch die *„Auslenkung der Schwingung bei einem Schwingungsvorgang“* beschrieben wird (Miosga 2006:59). Die durchschnittliche Lautstärke bei Konversationen beträgt 60 bis 65dB. Physiologisch wird die Variation durch den Atemdruck und die Spannung der Stimmlippen realisiert. Dadurch kommt es bei einem größeren

Lautstärkelevel meist zu einer erhöhten Stimmlage, da die Vibrationsfrequenz der Stimmlippen dies hervorruft. Im Deutschen wird in diesem Fall aufgrund der erhöhten Atemanstrengung eine Silbendehnung festgestellt. Das Merkmal der Lautstärke beeinflusst daher Tonhöhe und Lautlänge. Als auditive Empfindungsgröße ist die Lautstärke allerdings auch kontextabhängig. In einer Bibliothek kann zum Beispiel eine Konversation als lauter empfunden werden als dieselbe in einem gut besuchten Restaurant. Bei einer Untersuchung kann zum einen herausgefunden werden, auf welchem Lautstärkelevel jemand spricht, aber auch, inwiefern ein Lautstärkewechsel vorliegt. Interessant ist in diesem Zusammenhang auch die Spannweite der Variation. Setzt der Sprecher zum Beispiel in einem Poetry Slam-Vortrag unterschiedliche Lautstärken ein, um bestimmte Passagen zu betonen? Unter diesem Aspekt ist der Einbezug der Größe Lautstärke für die Untersuchung relevant.

2.1.2.2 Tonhöhe

Das Merkmal der *Tonhöhe* entspricht dem akustischen Parameter der *Frequenz*. Dieser beschreibt die Quantität der Schwingungen pro Zeiteinheit. Sie wird Grundfrequenz genannt und in Hertz (Hz) gemessen. Je höher die Grundfrequenz, das heißt, je öfter sich die Stimmlippen öffnen und schließen, desto höher die Stimmlage. Frauen haben meist eine höhere Stimmlage als Männer. Als mittlere Sprechstimmlage wird die durchschnittlich realisierte Tonhöhe bezeichnet, welche *„sich im unteren Drittel des individuellen Gesamtumfangs [befindet] und nicht angestrengt tief oder hoch, sondern entspannt“* klingt (Miosga 2006:62). Die Stimmlage kann in bestimmten Kontexten gezielt eingesetzt werden, daher ist auch sie für die Analyse von Poetry Slam-Texten interessant. Besonders bei Betrachtung der Tonhöhenvariation.

> „In konkreten Kontexten beschreibt der Stimmumfang die Bandbreite der Tonhöhenvariation und kann ermittelt werden als Variationsbreite (F_0- Range) der vorkommenden Grundfrequenzen (F_0 Minima/Maxima). Die Standardabweichung von der mittleren Stimmlage („standard deviation“) liegt zwischen zwei und vier Halbtönen.“ (Miosga 2006:62f.)

Nach Hargrove und McGarr wirken Sprecher monoton, wenn sie weniger als zwei Halbtöne variieren. Dies muss allerdings nicht für alle Sprachgemeinschaften gelten.

2.1.2.3 Lautdauer

Das Merkmal der *Lautdauer* entspricht dem akustischen Parameter der *Zeit*. Bei einer spontanen Kommunikation beträgt die übliche Silbendauer nach verschiedenen Untersuchungen 0,20 bis 0,27 Sekunden. Diese kann stark variieren, da die Dehnung einzelner Laute als rhetorisches Mittel dienen kann. Die Lautdauer steht in direktem Zusammenhang zur Sprechgeschwindigkeit, da die Verlängerung einzelner Silben oder Laute dazu führt, dass ein Sprecher eindringlicher wirkt und sich der akustische Beitrag zum Gespräch dehnt. Damit geht auch eine Erhöhung des Deutlichkeitsniveaus einher. Dies ist vor allem bei öffentlichen Auftritten („Bühnenlautung“ (Miosga 2006:65)) von Bedeutung: *„Bei einer überdeutlichen Lautdehnung handelt es sich um eine*

silbenstechende Überbetontheit, was außer in der Bühnenlautung vom Zuhörer oft als theatralische Unnatürlichkeit empfunden wird" (Miosga 2006:66). Es gilt also herauszufinden, wie der Sprecher sich zum durchschnittlich realisierten Deutlichkeitsniveau verhält.

2.1.2.4 Sprechpausen

Pausen entsprechen innerhalb eines Redebeitrages dem akustischen Parameter der *Stille* (vgl. Miosga 2006:67). Es wird zwischen verschiedenen Pausenarten unterschieden. Sie haben nicht nur die Funktion, einzelne Worte voneinander zu unterscheiden, sondern können gezielt eingesetzt werden, um zum Beispiel bei einem Vortrag den Zuschauern die Möglichkeit zu geben, über das Gesagte nachzudenken. Folgende Möglichkeiten zählt Miosga auf, um Pausen in Analyseabschnitten zu untersuchen: Anteil an der Gesamtsprechzeit (viel/wenig), Dauer (lang/kurz; mindestens 0,13 Sekunden bis zu 0,27 Sekunden), Häufigkeit (häufig/selten), Position (innerhalb des Redebeitrages/ zwischen den Redebeiträgen von verschiedenen Sprechern), sprechbezogene Aktivität (gefüllt, spannungsvoll, spannungsvoll leer). Pausen, die kürzer als 0,13 Sekunden sind, sind auf *„artikulationsmotorische Umschaltvorgänge zurückzuführen"* (Miosga 2006:67). Außerdem differenziert sie zwischen unterschiedlichen Pausenarten, die von der Position in einer Äußerung oder in einem Gespräch abhängen.

> „Innerhalb des Redebeitrages: Gliederungspausen (lexikalische/semantische/ syntaktische Gliederungshilfe), Abgrenzungspausen (phrasale Gliederung), Betonungspausen (Hervorhebung wichtiger Redeeinheiten oder Begriffe), Denkpausen (repräsentieren Denk- und Verarbeitungsprozesse des Sprechenden). Zwischen den Redebeiträgen von verschiedenen Sprechern: Abgrenzungspausen (Signal für den Sprecher-Hörer-Wechsel) bezüglich der sprechbezogenen Aktivität werden die Atempausen (unwillkürliches Anspannen/ willkürliches Luftholen) in ihrer biologischen Funktion der Luftergänzung von den Pausen ohne Atmung unterschieden." (Miosga 2006:70)

2.1.2.5 Stimmqualität

Das Merkmal der *Stimmqualität* ist mehreren akustischen Parametern zuzuordnen und daher relativ komplex. Diese sind nach Miosga (2006:71) die Grundfrequenz, Intensität, Obertoncharakteristika und der Störungsgrad. In diesem Zusammenhang ist zu sagen, dass durch Variation des Stimmklangs, der Klangfarbe, von Vokaltrakteinstellung und Stimmlippenschwingung verschiedene Stimmqualitäten erzeugt werden können. Mit Stimmklang ist die Umwandlung der Atemluft in Klang gemeint. Wie viel dieser Atemluft zur Klangproduktion genutzt wird, kann durch den Sprecher beeinflusst werden.

Miosga (ebd.) nennt zur Beschreibung des Stimmklangs Begriffe wie brüchig, rau, belegt, zittrig, kratzend oder knarrend. Der Hörereindruck ist kontextabhängig. Der Stimmklang wird nicht nur durch die mehr oder weniger ökonomische Nutzung der Atemluft beeinflusst, sondern auch durch das *„Material, [die] Form und [die] Bauart des Resonanzkörpers"* (ebd.). Die Vokalbildung konstituiert die Klangfarbe, die mit

Hilfe von Gegensatzpaaren wie zum Beispiel hell-dunkel, zart-spitz oder eng-weit beschrieben wird. Um das Variationsspektrum der Stimmqualität im Einzelnen wissenschaftlich benennen zu können, wurden einige *„Settings"* festgelegt, die physiologische Hintergründe und das entstehende akustische Signal aufzeigen (Miosga 2006:72). Diese werden bei Miosga ausgeführt und hier dann herangezogen, sobald sie für die vorliegende Analyse relevant erscheinen.

2.1.2.6 Akzent

Die Komponente *Akzent* bezeichnet die Betonung einzelner Silben in einem Wort gegenüber der unbetonten Variante. Durch den Vorgang der Akzentuierung kann auf einen bestimmten Laut aufmerksam gemacht werden. Er dient auch dazu, einen Satz zu strukturieren oder eine Melodisierung vorzunehmen. Nach Miosga zeichnet sich ein Akzent durch folgende lautliche Merkmale aus:

> „Wechsel der Tonhöhenrichtung auf der akzentuierten Silbe, Tonhöhensprünge vor der Akzentsilbe, erhöhte Lautstärke auf der Akzentsilbe, erhöhte Lautdauer des Vokals der Akzentsilbe, Pausen vor der Akzentsilbe, Wechsel der Klangfarbe des Vokals der Akzentsilbe." (Miosga 2006:82f.)

Zur Markierung eines Akzents können nur eine, aber auch mehrere dieser Möglichkeiten vom Sprecher genutzt werden. Je nachdem werden folgende Akzentarten unterschieden:

- dynamischer Akzent (Lautstärke in der betonten Silbe nimmt zu)
- temporaler Akzent (Dehnung des Vokals oder des stimmhaften Konsonanten, Pausen vor wesentlichen Begriffen)
- melodischer Akzent (Variation der Tonhöhenrichtung: steigend, fallend, gleichbleibend) (vgl. Miosga 2006:83)

Eberhard Stock veröffentlichte 1996 das Buch „Deutsche Intonation", in dem er unter anderem die Besonderheiten der deutschen Akzentuierung beschreibt. Er macht darauf aufmerksam, dass die Wortakzentuierung in vielen Sprachen klar festgelegt ist. Als Beispiel nennt er das Französische, in dem der Akzent häufig auf der letzten Silbe liegt. Falsch gesetzte Akzente erschweren das Verständnis eines Wortes oder eines Satzes.

Die Akzentuierung im Deutschen ist laut Stock einfacher als im Polnischen, jedoch schwieriger als im Französischen. Dies liegt zum Beispiel daran, dass zwischen ursprünglich deutschen und fremden Worten unterschieden werden muss. Bei fremden Ausdrücken gilt die Akzentuierung der Herkunftssprache. Im Deutschen wird bis auf einige Ausnahmen der Wortstamm akzentuiert. Dieser ist als bedeutungstragender Teil eines Wortes meist einsilbig und in Präfix, Suffix und Endungen eingebettet (z.B. Ge-*zog*-en). In einem Satz oder in einer Sinneinheit sind meistens mehrere Akzente vorhanden. Daher wird zwischen mehreren Akzentgraden unterschieden. Diese ermöglichen eine Abstufung in Hauptakzent, Nebenakzent und besondere Akzentformen (Miosga 2006:85f.).

2.1.2.7 Intonation

Der Begriff der *Intonation* ist, wie bereits Neuber anführte, mehrfach besetzt. Miosga ordnet die Intonation als prosodische Komponente ein und bezeichnet damit *„den wahrgenommenen Tonhöhen- oder Melodieverlauf von Äußerungen (oft „Intonation im engeren Sinne" genannt, vgl. [...]) und Äußerungsteilen"* (Miosga 2006:87). Folgende sind die wesentlichen Vokabeln zur Beschreibung der Intonation nach Miosga:

- „Tonhöhenänderung
  - ➔ lokale Tonhöhenänderung: Tonsprünge nach oben oder nach unten
  - ➔ globale Tonhöhenänderung: Stimmlagenwechsel nach oben oder unten / ROL [„relative onset level"] (mittel, hoch, tief)
- lokale Tonhöhenrichtung
  - ➔ des Tonhöhenansatzes [Tonhöhe der ersten betonten Silbe einer Äußerung]
  - ➔ des melodischen Akzents [prominenteste Silbe einer Äußerung]
  - ➔ der Kadenz (fallend, steigend, gleichbleibend) [Tonhöhenbewegung vor einer Pause]
- globale Intonationsstruktur [Anfang bis Ende eines Redebeitrages]: monoton, variationsreich, gleichförmig (isoton, „Singsang")" (Miosga 2006:90)

Die Produktion verschiedener Tonhöhenverläufe ist durch Variation der Stimmlippenspannung und des subglottalen Drucks möglich. Wird zum Beispiel ein fallender Tonhöhenverlauf festgestellt, so ist dies in einem Absinken der Grundfrequenz messbar. Dies ist in der Regel der Fall, wenn nicht gezielt entgegengesteuert wird („downstep" oder „Abfall pro Akzent"). Der Hörer bemerkt diesen Vorgang nicht unbedingt durch den Abfall der Grundfrequenz, sondern durch andere prosodische Parameter wie der Intensität oder einer Dehnung der finalen Laute. Aufgrund dieser Tatsache betrachten viele Sprachwissenschaftler Intonation als komplexes Phänomen, in dem die verschiedenen Einflusskriterien beinhaltet sind, was Miosgas Meinung nach häufig zu Irritationen führt, weswegen der engere Begriff zweckdienlicher ist.

2.1.2.8 Sprechtempo

Das Sprechtempo wird nach Miosga in Silben oder Wörtern pro Minute gemessen. Mit Bezug zu einer vorangegangenen Studie ist die Variationsbreite im Deutschen zwischen 100 und 400 Silben pro Minute angegeben. Folgende Abstufungen sind tradiert: sehr langsam (100 Silben/min.), langsam (150 Silben/min.), untermittel (200 Silben/min.), mittelrasch (250 Silben/min.), übermittel (300 Silben/min.), rasch (350 Silben/min.) und sehr rasch (>350 Silben/min.) (Miosga 2006:90).

2.1.2.9 Rhythmus

Verschiedenen Sprachen liegen unterschiedliche Rhythmusmuster zugrunde. Diese äußern sich nicht nur durch die Lautdauer, sondern auch durch die verschiedene

Akzentuierung der Worte. Der Rhythmus ist also, ähnlich wie die Intonation, durch mehrere Parameter bedingt, und kann als komplexes Phänomen angesehen werden.

Besonderes Augenmerk sollte auf Unterbrechungen des Sprechrhythmus gelegt werden. Die rhythmisch-einheitliche Wirkung eines Beitrages hängt unter anderem von der Sprechflüssigkeit („fluency“) ab. Beispiele für Zögerungsphänomene sind nach Miosga (2006:93): *„leere oder gefüllte Pausen, Füllwörter, Füllphrasen, Wortwiederholungen, Wortkorrekturen und Satzrevisionen“*.

Eberhard Stock interpretiert die Rhythmik als Gliederung der sprachlichen Äußerung. Als Grund hierfür nennt er wie Miosga die physiologische Notwendigkeit, die Atemluft einzuteilen. Um die korrekten Stellen dieser Zäsur beschreibbar zu machen, bezeichnet Stock (1996:69) sie im spontanen Sprechen als rhythmische Gruppen. Diese variieren je nach Aussageabsicht oder Rhythmisierungsgewohnheiten einer Sprachgemeinschaft.

Die Besonderheit des Deutschen ist die im Vergleich zu silben- oder akzentzählenden Sprachen ungewöhnliche Behandlung von akzentlosen Silben. Bei einer akzentuierten Silbe wird im Deutschen die Sprechspannung und die Lautstärke beträchtlich gesteigert. Akzentlose Silben werden dagegen leiser gesprochen oder reduziert. Dadurch entsteht ein staccato-ähnlicher Rhythmus. *„Der Akzent hat hier gleichsam eine zentralisierende Wirkung: Er zieht fast die gesamte Artikulationsenergie auf eine Silbe“* (ebd.). Im Gegensatz zum Französischen, das einen *legato*-Rhythmus aufweist und eher weich klingt.

Trotzdem ist die Einteilung der rhythmischen Gruppen nach Stock nicht rein zufällig. In einem Abschnitt über reproduziertes Sprechen tätigt er einige Aussagen zu Umständen, von denen die Bildung rhythmischer Gruppen beim Vorlesen, Sprechen gelernter Texte oder Reden Halten mit Manuskript abhängt:

- *„von der syntaktischen und inhaltlichen Struktur der einzelnen Sätze und – sofern vorhanden – von der Textgebung,*
- *von den Wortgruppenakzenten,*
- *von der Gestaltungsabsicht und dem sich daraus ergebenden Sprechtempo“* (Stock 1996:70)

Stock merkt an, dass die Akzentuierung von häufig vorlesenden Personen auch von den äußeren Rahmenbedingungen abhängig gemacht wird: *„von den Hörenden, von der Art des Textes, von den Raumbedingungen usw. Er wird ein Märchen langsam, mit vielen Pausen und kleinen rhythmischen Gruppen vortragen. Ein Bericht dagegen kann viel schneller und mit wenig Pausen vorgelesen werden“* (Stock 1996:70). Diese Erkenntnisse sind besonders interessant bei der Betrachtung von Poetry Slam-Vorträgen, da diese zum Teil vorgelesen, zum Teil aber auch auswendig-gelernt aufgesagt werden.

2.2 Poetry Slam

2.2.1 Was ist Poetry Slam?

> „Der Poetry Slam ist ein der Subkultur zugehöriges Format, das sich be sonders an junge Autoren wendet und ihnen eine Möglichkeit bietet, abseits vom klassischen Literaturbetrieb ihre Texte zu veröffentlichen. Der Poetry Slam in Deutschland ist ein Hybrid aus der amerikanischen Spoken-Word-Kultur und deutscher Poesie. Tatsächlich ist der Poetry Slam die radikalste Infragestellung des derzeitigen Literaturbetriebs und seiner zelebrierten Hochkultur, er ist das Modell einer demokratisierten Geschmacksbildung.“ (Westermayr 2010:153)

Die aus Amerika stammende Literaturveranstaltung namens *Poetry Slam* wird seit dem Jahre 1993 mit großem Erfolg auch in Deutschland durchgeführt. Zunächst nur vereinzelt, finden inzwischen in fast allen großen Städten der Bundesrepublik regelmäßig Poetry Slams statt. Dieser Wettstreit der Dichter, welche den mittelalterlichen Bänkelsängern nahe steht, ist in seiner aktuellen Umsetzung einmalig in der Literaturgeschichte und vermehrt Gegenstand von Diskussionen. Das Veranstaltungsformat genießt vor allem bei einem jungen Publikum große Popularität und die häufig jungen Poeten, die mit selbst-verfassten Texten um die Gunst des Publikums slammen, bieten Anlass zur Auseinandersetzung mit den Texten und deren Rezeption in der Literaturkritik, aber auch unter den Jugendlichen, denen ein Zugang zur Poesie ermöglicht wird, welcher nicht nur für die Pädagogik fruchtbar sein kann.

Petra Anders analysiert in dem Buch „Poetry Slam im Deutschunterricht: Aus einer für Jugendliche bedeutsamen kulturellen Praxis Inszenierungsmuster gewinnen, um das Schreiben, Sprechen und Zuhören zu fördern“, das im Jahr 2010 veröffentlicht wurde, Texte des Poetry Slams und deren Einsatzmöglichkeiten im Deutschunterricht. Dabei beleuchtet sie nicht nur die Ursprünge des Veranstaltungsformats, sondern auch dessen Leitlinien: *„Es sollen also selbstverfasste Texte (Regel 1) innerhalb eines Zeitlimits (Regel 2) ohne Kostüme und Hilfsmittel (Regel 3) mit anschließender Publikumsbewertung (Regel 4) aufgeführt werden.“* (Anders 2010:23)

Poetry Slam gilt sowohl in Deutschland als auch in Amerika als ein gemeinsames Format. So haben sie die gleichen Leitlinien, welche von dem Begründer der Szene, Marc Kelly Smith, festgelegt wurden. Anders macht allerdings deutlich, dass sich der Poetry Slam in Deutschland in mehreren Aspekten unterschiedlich von dem amerikanischen Vorbild entwickelt hat. Sie merkt zum Beispiel an, dass bei Texten deutscher Poeten zurzeit die Tendenz zur Komik erkennbar ist. Diese Entwicklung hat es in der amerikanischen Slam Szene in ähnlicher Form gegeben und sie mündete in eine Differenzierung verschiedener Profile für Poetry Slam Veranstaltungen. Anders zitiert an dieser Stelle Edward Garcia, der einen Unterschied zwischen den eher „serious poets“ und den „performance poets“ feststellt. Im Folgenden wird ausschließlich die deutsche Form des Poetry Slams untersucht. Wie also läuft ein typischer Poetry Slam in Deutschland ab?

„Üblicher Ablauf eines Poetry Slam in Deutschland

- *Die Slam-Poeten, die auftreten möchten, melden sich beim Betreten der Veranstaltung (oder während der Veranstaltung) an der Kasse oder bei dem Moderator an (und erhalten dafür freien Eintritt).*
- *Der Moderator begrüßt das Publikum.*
- *Der Moderator erläutert die Regeln und ermittelt (meist per Los) die Reihenfolge der auftretenden Slam-Poeten, deren Namen (meist auf eine Tafel) geschrieben werden.*
- *Der Moderator vergibt die Stimmtafeln (oder Rosen, Dichtungsringe etc. je nach Abstimmungsmodalitäten) an die von ihm willkürlich aus dem Publikum gewählte Jury (aus drei bis zehn oder mehr Juroren) und erläutert die Abstimmungsregeln (z.B. Beurteilung nach jedem Poeten oder nach den ersten drei Poeten; gleichzeitiges Hochhalten der Stimmtafeln, Löschung der höchsten und der niedrigsten Punktzahl)*
- *Der Moderator bittet einen (meist unbekannten) „Featured" Poeten auf die Bühne, der als „Opferlamm" außerhalb des Wettbewerbs auftritt, für Stimmung sorgt und an dessen Auftritt die Jury ihre erste Abstimmung testen kann […].*
- *Der Poetry-Slam-Wettbewerb beginnt, indem der Moderator unter Applaus des Publikums die einzelnen Slam-Poeten oder -Teams (von maximal sechs Personen) nacheinander auf die Bühne bittet, diese die vorher bekannt gegebene Zeit (von meist fünf Minuten) für die Aufführung nutzen, und nach dem Auftritt von der Jury beurteilt werden. Die Punkte werden (auf der Tafel) notiert.*
- *Meist treten ca. 12 Slam-Poeten pro Veranstaltung auf; nach den ersten sechs Auftritten wird eine Pause eingelegt, in der ein DJ Musik macht und das Publikum entspannen kann.*
- *Bei Gleichstand erfolgt ein Stechen, die Poeten sollten also mehr als einen Text in petto haben, können nach Absprache aber auch mit demselben Text erneut auftreten.*
- *Bei der Siegerehrung werden symbolische Preise verteilt.*
- *Der Moderator kündigt den Termin für den nächsten Poetry Slam an"*
  (Anders 2010:24).

Der Ablauf eines Poetry Slams variiert je nach Veranstalter, ist aber in seinen Grundzügen laut Anders (2010:25) als eine Art kulturelles Ritual mit verbindlichen Regeln zu bezeichnen. Die Tatsache, dass eine Jury zufällig aus dem Publikum gewählt wird, also das Werturteil nicht von einem renommierten Literaturkritiker, sondern möglicherweise von einem Laien vorgenommen wird, verdeutlicht den revolutionären Geist der Veranstaltung und das demokratische Verständnis von Bildung und Kultur. Gerade deswegen steht das Veranstaltungsformat allerdings vermehrt in der öffentlichen Kritik. Ein literarisch-anspruchsvoller Text, den ein erfahrener Literaturkritiker vermutlich als solches erkannt hätte, wird möglicherweise schlechter vom Publikum bewertet als das populär-formulierte Pendant.

2.2.2 Forschungsstand

Als relativ neue Strömung wurde das Format des Poetry Slams in der wissenschaftlichen Literatur bislang als eher nebensächlich wahrgenommen. Anders (2010:7ff.) fasst die Schriften zu diesem Thema in einem Kapitel ihrer Arbeit zusammen. Bei der Lektüre fällt auf, dass sich mehrere Literaturwissenschaftler mit einer Einordnung in die Gegenwartsliteratur beschäftigt haben. Boris Preckwitz zum Beispiel erstellte in einer Magisterarbeit einen Kriterienkatalog, der die Merkmale des Formats zusammenfasst. Kutsch entwickelte eine Art Pressespiegel. Es wurden auch Arbeiten zu Erfolgskriterien eines Poetry Slam-Vortrags und der Geschichte des Phänomens verfasst. Laut Anders stammen die meisten Autoren aus der Szene selbst.

2.2.3 Was ist Slam Poetry?

Der Begriff *Slam Poetry* bezeichnet nicht, wie zuerst annehmbar, ein eigenes Genre. Nach Anders gibt es in der Szene ein ungeschriebenes Gesetz, dass ein Poetry Slam als eine offene Bühne anzusehen ist, *„auf der alles vorgetragen werden kann, sofern es selbstverfasst ist“* (Anders 2010:45). Somit ist *Slam Poetry* der Oberbegriff für alle Textsorten, die für und auf einem Poetry Slam verfasst und präsentiert werden.

Trotzdem können einige der Slam-Poesie eigenen Merkmale festgestellt werden, welche einen charakteristischen Stil dieser Textsorten konstituieren. Diese fünf werden von Anders (vgl. 2010:46 f.) in Aktualität, Klanglichkeit, Interaktion, Intertextualität und Kürze eingeteilt.

Die Texte haben meistens einen direkten Lebensweltbezug und behandeln Themen, die für die Autoren aktuell sind. Dies kann zum einen der persönliche Alltag sein, aber auch das momentane Geschehen in Politik und Kultur. Durch die Nähe der Inhalte zu einem potenziell jedem Zuschauer zugänglichen mentalen Modell wird es diesem leicht gemacht, einen Zusammenhang zwischen seinem eigenen Leben und dem des Vortragenden zu finden. Dadurch kann der Rezipient Wirkungszusammenhänge besser nachvollziehen.

Anders (ebd.) bezeichnet die klangliche Wirkung eines Poetry Slam-Vortrags dem amerikanischen Vorbild ähnlich als *liedartig* und nennt damit ein mögliches Erkennungsmerkmal von Slam Poetry. Als weiteres Merkmal wird hier die hohe Sprechgeschwindigkeit genannt, einhergehend mit großen rhythmischen Einheiten.

Kontrastiv zu üblichen Autorenlesungen provozieren bei einem Poetry Slam vorgetragene Texte häufig Reaktionen beim Publikum. Dies liegt nicht nur an dem Format der Veranstaltung, das eine Bewertung der Texte durch das Auditorium vorsieht, sondern auch an den Texten selbst. Zum Teil fordert der Poet, zum Beispiel durch einen mitzusprechenden Refrain, direkt zur Interaktion auf.

Der vierte von Anders (vgl. 2010:47) benannte Aspekt ist der der Intertextualität. Gerade durch die offene Form des Poetry Slams wird mit mündlichen wie schriftlichen Genres verfremdend gespielt. Es werden häufig der Alltagssprache relativ entfernte Begrifflichkeiten verwendet, die in mehreren Texten, auch Poeten übergreifend, wiederkehren.

Zuletzt ist die durch das Regelwerk bestimmte Kürze der Vorträge zu nennen. Anders beschreibt die Neigung deutschsprachiger Slam-Poeten, lyrische Elemente

sowie Reimstrukturen und Merkmale des Storytelling zu verwenden. Häufig werden auch komische Elemente integriert, die den Kontakt zwischen Vortragenden und Publikum in der kurzen zur Verfügung stehenden Zeit vertiefen soll.

Auch Stefanie Westermayr hat sich mit den Merkmalen von Texten im Poetry Slam auseinandergesetzt. In ihrem Werk „Poetry Slam in Deutschland: Theorie und Praxis einer multimedialen Kunstform" nimmt sie eine Kategorisierung in Themen, Präsentationsformen und Stilmittel vor (Westermayr 2010:91 ff.).

Poetry Slam-Texte zeichnen sich durch eine große Vielfalt aus. Es wird nicht nur Lyrik, sondern auch Prosa vorgetragen, die allein oder in einem Team präsentiert werden. Die Übergänge zwischen den Bereichen sind fließend und in einem Poetry Slam-Text können mehrere eine Rolle spielen. Die folgende Übersicht, welche eine Zusammenfassung von Westermayrs Ausführungen darstellt, erhebt keinen Anspruch auf Vollständigkeit, ist aber möglicherweise für spätere Erläuterungen hilfreich.

Themen	**Präsentationsformen**	**Stilmittel**
➢ Beschreibung eines bestimmten „Typus" ➢ Selbstreflexion, ➢ Aktuelle politische oder gesellschaftliche Kommentare, ➢ Milieuschilderung, ➢ Kritischer Standpunkt, ➢ Alltägliches (Mode, Marken, Sex, Partys, Großstadtleben)	➢ Gedicht (am wenigsten Interaktion mit Publikum), ➢ Prosaminiatur, ➢ Monolog (viel Interaktion mit Publikum)	➢ Stabreime ➢ Lautmalerei ➢ Interaktion durch Fragen

Abbildung 2. Einteilung nach Westermayr (2010:91ff.).

Aber auch sprachwissenschaftlich ist die Vortragsform der im Poetry Slam präsentierten Texte interessant. Vor allem durch die Rahmung der Vorträge ist eine eigene Gattung entstanden, die sowohl im Text als auch in der Textpräsentation ungewöhnliche und einmalige Ausprägungen angenommen hat.

Der deutsche Slam-Poet *Sebastian23* wurde von Sebastian Rabsahl von der Zeitschrift „Sprachnachrichten", die monatlich von dem Verein Deutsche Sprache veröffentlicht wird, unter anderem zu diesen Besonderheiten befragt. Auf einen Slam-Wettbewerb in England angesprochen, befürwortete er seine deutschsprachige Performance damit, dass beim Poetry Slam der Klang der Sprache im Mittelpunkt stehe und nur so als akustisches Erlebnis erhalten bliebe. Die Frage, ob bei Slam Poetry das Schreiben der Texte oder das Darstellen auf der Bühne wichtiger ist, beantwortet er mit einem klaren Statement: *„Die schreibende Seite. Es ist wichtiger, dass man Ideen mitbringt, um Sprache kreativ zu gestalten, als es dann auf der Bühne umzusetzen, obwohl man für beides Talent haben muss – so wie ich [...]"* (Rabsahl 2011:3).

Bislang liegt noch keine wissenschaftliche Analyse der Prosodie von Slam Poetry vor. Anhand der Literaturrecherche werden im Folgenden Anhaltspunkte aufgelistet, die Hinweise zur Ausprägung von prosodischen Parametern im Poetry Slam zulassen.

2.2.4 Prosodie im Poetry Slam

2.2.4.1 Begriffsherkunft

Die Bezeichnung *slam* kommt häufig in Zusammenhang mit sportlichen Wettkämpfen vor, wie zum Beispiel der *Grand Slam* beim Basketball. Das englische Verb „to slam" bedeutet übersetzt *„jemanden schlagen, (eine Tür) zuknallen, schlagen, jemanden heruntermachen etc."* (Anders 2010:18). In einem Interview mit dem amerikanischen Slam-Poeten K-Swift oder Marc Kelly Smith, den Anders befragte, umschreibt dieser Slam Poetry als Texte, *„which hit the bullseye"* (Anders 2010:18). Dieser erneut dem Sport entspringende Begriff beschreibt einen Treffer im inneren Kreis einer Dartscheibe, welcher die höchste Punktzahl bedeutet.

Diese Benennung erlaubt eine Aussage über das Wesen des ursprünglichen Veranstaltungsformats: Es ist nicht nur eine Lesung, sondern ein Wettkampf der Dichter, in dem es wie in einem sportlichen Ereignis einen Gewinner gibt. Dem Bedeutungshof von *slam* wohnt ein aggressiver Grundton inne, der sich auf die Texte im Poetry Slam auswirken kann. *„Das „Slammen" kann also bedeuten, dass ein Poet etwas schnell und treffsicher auf den Punkt bringt, eine Aussage dem Publikum präzise und durchsetzungsstark darbietet bzw. den Zuhörer mit der eigenen Meinung konfrontiert" (*Anders 2010:19).

Christine Miosga hat zur Beschreibung des Habitus der Prosodie lautliche Eigenschaften aufgeführt, die bestimmten emotionalen Zuständen oder kommunikativen Gattungen zuzuordnen sind. Aggression zeichnet sich durch einen lauten Level aus (vgl. Miosga 2006:136f.), sowie steigende Tonhöhe (vgl. Miosga 2006:146) bei zunehmender Wut (vgl. Miosga 2006:189). Ein lauter Lautstärkelevel kann laut Miosga dazu führen, dass der Sprecher unteren Gesellschaftsklassen zugeordnet wird. Ein erhöhter Aktivierungsgrad zeichnet sich außerdem trotz der hohen Sprechgeschwindigkeit (vgl. Miosga 2006:196) durch eine verlängerte Lautdauer und erhöhte Artikulationspräzision aus (vgl. Miosga 2006:152). Sprechpausen kommen vermindert vor und sind mit Spannung gefüllt (vgl. Miosga 2006:159). Der Stimmklang ist rau bzw. hart und die Klangfarbe wird als „eng" oder „knödelnd" bezeichnet. Die Stimme setzt hart ein und es wird das Brustregister zum Sprechen genutzt (vgl. Miosga 2006:171). Ein dynamischer Akzent charakterisiert Zorn und Ärger (vgl. Miosga 2006: 179f.).

2.2.4.2 Kulturhistorische Prägung

Im deutschen Poetry Slam spielt die kulturelle Vorgeschichte eine nicht zu unterschätzende Rolle. Im Gegensatz zum englischen Begriff „poet" ist das deutsche Pendant *Poet* eher in der literarischen Tradition verhaftet und bezeichnet einen Dichter, der sich klassischen Gattungen wie dem Drama oder der Lyrik verschrieben hat, wie zum Beispiel Goethe oder Lessing. Daher ist es für Slam-Poeten wichtig, sich von der geschichtlichen Prägung abzusetzen. Dies kann bedeuten, dass zum Beispiel klassische Reimformen neu interpretiert werden oder moderne Elemente wie Rap oder Beatbox in die Vorträge eingebaut werden. Für die prosodische Analyse heißt das, dass nicht davon ausgegangen werden kann, dass wie zum Beispiel bei einem Gedicht ein Reimschema

die gesamte Zeit über durchgehalten wird, sondern es ist ein virtuoser Umgang mit Klang zu erwarten (vgl. Anders 2010:19).

2.2.4.3 Rahmung

Drittens sind die Regeln eines Poetry Slams Einflusskriterien für die Prosodie der vorgetragenen Texte. Es müssen eigene Texte vorgetragen werden, sodass der Künstler einen persönlichen Bezug zu dem Präsentierten hat. Das kann unter Umständen dazu führen, dass er emotionalisiert spricht. Je nach Ausprägung der Empfindung gibt Miosga Anhaltspunkte, wie sich diese auf die Prosodie auswirkt. Bei Gefühlen mit hohen Aktivierungsgrad, zum Beispiel Freude oder Leidenschaft, kommt es zu einer erhöhten Stimmlage, während es bei Gefühlen mit vermindertem Aktivierungsgrad, zum Beispiel Trauer oder Gleichgültigkeit, eine erniedrigte Stimmlage ist (vgl. Miosga 2006:146). Die Texte müssen außerdem der begrenzten Zeit angepasst werden. Prosa wird meistens sehr schnell vorgetragen.

2.2.4.4 Komik

Im deutschsprachigen Poetry Slam sind witzige Texte sehr erfolgreich. Das kann unter Umständen auch an der klanglichen Qualität liegen.

> „Komik entsteht auch durch lustige oder skurrile Sprach- und Klangspiele. So gibt es Slam-Texte, die nur durch Lautmalerei, Rhythmus, Reim, Wiederholung eine sinnliche Lust an der lautlichen Substanz von Sprache zeigen, ohne einen Anspruch auf inhaltliche Nachvollziehbarkeit zu erheben.“ (Anders 2010:48)

Ein Beispiel dafür ist die Schüttelprosa von Lasse Samström, bei der er die Anfangsbuchstaben von Worten vertauscht und es so schafft, auf komische Art und Weise eine an sich triviale Geschichte zu vermitteln.

2.2.4.5 Vortragsstil

Anders (2010:46) beschreibt die liedartige Wirkung sowohl amerikanischer als auch deutscher Slam Poetry. Dies begründet sie mit dem fließenden Lese- und Vortragsstil der Poeten und dessen Nähe zum Rap. *„Geschichten im „Lesebühnenstil“ sind aggregativ (mehrgliedrig) und eher additiv (auflistend), sie bieten durch ihre inhaltliche Komposition und den schnellen Vortrag nahezu rhythmische Einheiten“* (Anders 2010:46). Es ist folglich eine hohe Sprechgeschwindigkeit zu erwarten, sowie ein klarer Intonationsverlauf.

2.2.4.6 Mündlichkeit und Schriftlichkeit

> „Slam-Texte werden von den Verfassern sorgfältig als Schrifttexte vor bereitet und erst nach Fertigstellung in die mündliche Version, also den Vortrag, übertragen.“ (Anders 2010:56)

Peter Koch und Wulf Österreicher legten in dem Aufsatz „Schriftlichkeit in Sprache“ eine Theorie zum Verhältnis von Nähe und Distanz zu konzeptioneller und medialer Schriftlichkeit und Mündlichkeit dar. Die klassische Einteilung von Mündlichkeit und Schriftlichkeit beruht nach Koch und Österreicher auf dem Medium. So ist alles mündlich, das phonisch umgesetzt wird und alles schriftlich, was graphisch zum Beispiel durch Schrift auf Papier dargestellt ist. Dieses Konzept genügt nicht, um Phänomene wie die verschiedene Wortwahl bei einem Geschäftsbrief und einem Privatbrief zu erklären.

Daher muss eine neue Kategorisierung vorgenommen werden. Koch und Österreicher nutzen die Konzeption, das heißt die Merkmale schriftlicher und mündlicher Sprache, um Unterschiede zu beschreiben. Die konzeptionelle Mündlichkeit und Schriftlichkeit ist abhängig *„von Parametern wie 'raum- zeitliche Nahe oder Distanz der Kommunikationspartner', 'Öffentlichkeit', 'Vertrautheit der Kommunikationspartner', 'Emotionalität', 'Situations- und Handlungseinbindung', (...), 'Dialog/Monolog', 'Spontaneität', 'Themenfixierung' usw.“* (Koch und Österreicher 1994:587 ff.).

Das heißt, je vertrauter ich mit meinem Gesprächspartner bin, umso *mündlicher* bzw. *umgangssprachlicher* spreche ich mit ihm, unabhängig davon, ob ich medial-mündlich oder schriftlich mit ihm kommuniziere. Koch und Österreicher haben dieses Verhältnis in einem Diagramm dargestellt.

Die Unterscheidung ist insofern wichtig für die Prosodie im Poetry Slam, da konzeptionelle Mündlichkeit anders realisiert wird als Schriftlichkeit. Poetry Slam-Texte sind nach Anders (2010:57) aufgrund der Vortragsform medial-mündlich, von der Umsetzung her aber konzeptionell-schriftlich, *„so ähnelt eine Slam Poetry-Darbietung am ehesten einer Predigt oder einem wissenschaftlichen Vortrag. Denn diese sind als phonisch realisierte Texte genauso wie der geschriebene Slam-Text weitgehend monologisch, vorgeplant und themafixiert[...]“*. Sie grenzen sich allerdings insofern von einem wissenschaftlichen Vortrag oder Ähnlichem ab, als dass die Sprecher eine Nähe zum Publikum inszenieren. Dessen Umfang variiert, je nach gemeinsamen Hintergrundwissen des Publikums mit dem Poeten.

> „Versteht es ein Poet, seinen Text an regionale, lokale bzw. zielgruppenspezifische Bedingungen anzupassen, dann entsteht mehr (simulierte) Nähe zum Publikum, was oft den Erfolg des Auftritts begünstigt, als wenn Texte nicht an die wechselnden Auftrittsbedingungen angepasst werden.“ (Anders 2010:57)

Es ist dem Poeten demnach nicht nur möglich, Nähe zum Auditorium aufzubauen, sondern es ist auch für den Erfolg des Textes notwendig. Walter Ong formuliert Charakteristika oraler Kulturen, die Anhaltspunkte geben, wie die Prosodie bei Poetry Slam-Texten umgesetzt sein kann. Unter oralen Kulturen sind solche zu verstehen,

welche Wissen nicht schriftlich fixierten, sondern von Druiden oder Ähnlichem mündlich tradieren ließen. Diese inszenierten die Wissensweitergabe durch sprachliche Mittel wie Reimschemata oder Rhythmus.

„Nach Ong ist Oralität
1. stark rhythmisch
2. eher additiv als subordinierend,
d. h. bezüglich der Syntax neigen orale Menschen zu additiven Verbindungen („und"), statt zu solchen wie „damit", „dann", „deshalb", die hierarchische Strukturen entstehen lassen,
3. eher aggregativ als analytisch,
d. h. Wörter werden zu mehr oder weniger festen Formen zusammengefügt (z.B. die schöne Prinzessin, der tapfere Soldat),
4. redundant und nachahmend,
d. h. Wiederholungen und Ausschweifungen bieten dem Sprecher die Zeit zur gedanklichen Weiterentwicklung der Rede und vermindern akkustische (sic!) Verständnisprobleme des Zuhörers,
5. konservativ oder traditionalistisch, […]
6. nahe am menschlichen Leben,
d. h. Wissen wird in Bezug zur menschlichen Lebenswelt gewonnen und verbalisiert, indem die fremde, objektive Welt in das unmittelbare, bekannte Miteinander überführt wird,
7. kämpferisch im Ton,
d. h. in oralen Kulturen wird Sprache oft für verbale, intellektuelle Kämpfe eingesetzt,
8. eher einfühlend und teilnehmend als objektiv-distanziert,
d. h. Lernen und Wissen bedeutet für eine orale Kultur eine nahe, einfühlende gemeinsame Identifikation mit dem Wissensstoff,
9. homöostatisch, […]
10. *eher situativ als abstrakt, […]"* (Ong 1987:42 ff.; referiert von Anders 2010:62f.)

2.2.4.7 Redundanz

Für eine gelungene Wissensübermittlung ist zusätzlich Redundanz wichtig. Werden Informationen ausschließlich mündlich vermittelt, ist es notwendig, Störfaktoren des Vortrags einzuplanen und den Zuhörenden trotz dieser das Verständnis zu ermöglichen. Angesichts der gelösten Stimmung in Clubs, in denen Poetry Slams vornehmlich durchgeführt werden, enthalten Poetry Slam-Texte meistens mehr als eine Anschlussmöglichkeit, von der aus der Text verfolgt und verstanden werden kann. Daher werden bei einigen Refrains oder Anaphern zu Satzbeginn eingebunden. Diese Praxis ist durch Interviews mit Poeten belegt.

2.2.4.8 Textbezogene Inszenierung/ Performance

Bei deutschen Slam-Poeten sind laut Anders häufig prosodische und paraverbale Mittel zu beobachten, durch die bestimmte Textabschnitte zum Beispiel verstärkt werden. Neben einigen Beispielen beschreibt sie deren Funktion, den Text anschaulicher zu

machen (vgl. Anders 2010:72). In Amerika kommt es zurzeit zu einer Bewegung weg von nonverbalen Gestaltungsmitteln hin *„zum bloßen Sprechen des Textes, zur textbezogenen Performance. Die Poeten reduzieren ihre Rolle als Performer und nutzen lediglich Schnelligkeit und Lautstärke, um den Vortrag zu akzentuieren“* (ebd.). Bei der Analyse deutschsprachiger Slam Poetry ist demnach zu vermuten, dass sprachliche Mittel ungewöhnlich verwendet werden, wie zum Beispiel die Darstellung unterschiedlicher Protagonisten durch eine veränderte Tonhöhe oder die Einleitung eines Textes mit einer Art „Schlachtruf“.

Zusammenfassend ist zu sagen, dass Slam Poetry durch eine erhöhte Sprechgeschwindigkeit und den unkonventionellen Umgang mit sprachlichen Mitteln wie zum Beispiel der Tonhöhe charakterisiert ist. Teilweise wird auf klassische Stilmittel der Lyrik zurückgegriffen. Insgesamt besteht eine Ähnlichkeit zur Übermittlungspraxis oraler Kulturen aufgrund von Merkmalen mündlicher Informationsvermittlung und der Zuordnung zur konzeptionellen Schriftlichkeit.

3 Umfrage

3.1 Methodisches Vorgehen

In dem vorangegangenen theoretischen Teil wurde der Forschungsgegenstand der 'Prosodie im Poetry Slam' dargestellt, sowie das Begriffsspektrum in diesem Bereich eingegrenzt. Es wurde deutlich, dass Slam Poetry von Vertretern der Sprachwissenschaft und von den Poeten selbst eine besondere Sprechmelodie unterstellt wird. Angenommen, dies ist der Fall, dann sollte es möglich sein, diese anhand von auf technischem Wege delexikalisierten Vorträgen zu identifizieren.

Ausgehend von dieser Überlegung und auf der Grundlage der wissenschaftlichen Literatur soll in Hinblick auf das Ziel dieser Arbeit, die Spezifika der prosodischen Parameter im Poetry Slam herauszustellen, mit Hilfe einer empirischen Untersuchung getestet werden, ob die Unterscheidung auch in der Praxis funktioniert.

Um zu gewährleisten, dass die Teilnehmer der Studie nicht durch den lexikalischen Gehalt eines Vortrags in ihrer Wahl beeinflusst werden, wurde zunächst die Delexikalisierung der verschiedenen Textvorträge vorgenommen. Bei den bewusst prägnant gewählten Beispielen, die für den Vergleich verwendet wurden, handelte es sich jeweils um ein vorgelesenes Gedicht und ein vorgelesenes Essay. Den Probanden wurden diese nacheinander vorgespielt und sie konnten mit Hilfe eines Onlineportals eine Schätzung abgeben, wie die gehörten Beispiele zuzuordnen sind.

3.1.1 Korpus[1]

Für die Studie wurden zwölf verschiedene Textpräsentationen ausgewählt. Für die Vortragsformen wurden Aufnahmen von jeweils zwei weiblichen und zwei männlichen Sprecherinnen und Sprechern verwendet, um mögliche geschlechtsspezifische Unterschiede einzubeziehen.

Die Audiodateien der Poetry Slams wurden dem Online-Video-Portal Youtube entnommen, über das viele Slam-Poeten oder Fans derselben Aufzeichnungen veröffentlichen. Die Aufzeichnung von Slam Poetry wird nur von wenigen Veranstaltern vorgenommen, zum Teil entstammen sie, wie bei dem WDR Poetry Slam, einer Fernsehsendung. Neuere Formen, wie zum Beispiel der Poetry Clip, werden grundsätzlich aufgezeichnet, was Bestandteil ihrer Gattung ist. Weil viele Besucher von Poetry Slams diese privat dokumentieren, sind die Videos häufig in einer schlechten Qualität und für diese Studie nicht brauchbar. Die gewählten Beispiele von Andy Strauß, Wolfgang Lüchtrath und Pauline Füg wurden im WDR ausgestrahlt. Der Vortrag von Anke Fuchs wurde von dem Veranstalter des Poetry Slams Ulm veröffentlicht.

Der Text „Stefanie Hitzer" wurde aufgrund seiner großen Popularität ausgewählt. Andy Strauß verzerrt seine Stimme dabei sehr stark und inszeniert sich dadurch unkonventionell. Wolfgang Lüchtraths Vortrag verdeutlicht die emotionale Seite von Slam Poetry. Das stark emotionalisierte Sprechen ist möglicherweise charakteristisch für Slam Poetry. In dem Text „Was wisst ihr denn schon davon" von Anke Fuchs geht es um Vorurteile und verschiedene Lebensgeschichten. Es ist im Vergleich zu den anderen Beispielen für Slam Poetry ein eher ruhiger Text und gibt eine weitere Facette der Präsentationsformen innerhalb der Szene wieder. Dieser Vortrag sollte relativ schwer von den anderen abzugrenzen sein, da er nicht an die Sprechgeschwindigkeit der anderen heranreicht. Bei „Nen Orientierungslos" von Pauline Füg fällt der interessante und gleichförmige Intonationsverlauf auf.

Die Dateien der Gedichtbeispiele wurden über das Online-Portal Amazon käuflich erworben und befinden sich auf Tonträgern, auf denen mehrere Gedichte medial-mündlich wiedergegeben werden. Bei der Auswahl waren das Geschlecht des Sprechers und die Länge des Gedichts ausschlaggebend, da viele Gedichte kürzer als eine Minute und daher für diese Studie nicht geeignet sind.

Als erstes Beispiel wurde die Ballade von Friedrich Schiller „Die Bürgschaft" gewählt, die von Peter Reimers vorgetragen wird. Jürgen von der Lippe spricht das Gedicht „Das Ideal" von Kurt Tucholsky. In dem Gedicht „Das Riesenspielzeug" geht es um die Rolle des Bauern in der Gesellschaft. Der Name der Sprecherin konnte nicht herausgefunden werden. Das Gedicht „Havelland" von Theodor Fontane wird von Ute Beckert gesprochen.

---

[1] Die Begriffe, die im Folgenden verwendet werden, sowie die Definitionen zu den unterschiedlichen Forschungsdesigns, wurden dem Handbuch von Jürgen Bortz und Nicola Döring „Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler" aus dem Jahr 2006 entnommen. Die Autoren stellen in diesem Lehrbuch alle quantitativen und qualitativen Methoden der empirischen Sozialforschung und die dazugehörigen Musterlösungen vor. Außerdem wurde das Werk von Andreas Diekmann, „Empirische Sozialforschung" zur Ausarbeitung dieses Kapitels verwendet.

Für die Hörbeispiele der Gattung des Essays wurden vorgetragene Essays aus der Radiosendung „Gedanken zur Zeit“ von NDR Kultur ausgewählt. In dieser kommen bekannte Persönlichkeiten wie zum Beispiel Wilhelm Schmid oder Rainer Burchard zu Wort, die sich zu aktuellen Themen äußern. Die Dateien sind frei auf der Website des Radiosenders erhältlich und haben eine sehr gute Klangqualität. Sie wurden von Harald Eggebrecht, Cora Stephan, Reinhard Kahl und Christiane Grefe formuliert und vorgetragen. Die Auflösung im Test ist:

- Vergleichsgruppe 1 (m)
  1. Hörbeispiel 1 → Gedicht Peter Reimers ("Die Bürgschaft" von Schiller)
  2. Hörbeispiel 2 → Essay Harald Eggebrecht
  3. Hörbeispiel 3 → Slam Andy Strauß „Stefanie Hitzer“
- Vergleichsgruppe 2 (m)
  1. Hörbeispiel 1 → Slam Wolfgang Lüchtrath „Der Leuchtwart“
  2. Hörbeispiel 2 → Gedicht Jürgen von der Lippe ("Das Ideal" von Tucholsky)
  3. Hörbeispiel 3 → Essay Reinhard Kahl
- Vergleichsgruppe 3 (w)
  1. Hörbeispiel 1 → Essay Cora Stephan
  2. Hörbeispiel 2 → Gedicht weibl. Spr. ("Das Riesenspielzeug", von Chamisso)
  3. Hörbeispiel 3 → Slam Pauline Füg „Nen Orientierungslos“
- Vergleichsgruppe 4 (w)
  1. Hörbeispiel 1 → Gedicht Ute Beckert ("Havelland" von Fontane)
  2. Hörbeispiel 2 → Slam Anke Fuchs „Was wisst ihr denn schon davon“
  3. Hörbeispiel 3 → Essay Christiane Grefe

3.1.2 Delexikalisierung

Zunächst wurden die gewählten Beispiele mit dem Programm Audacity jeweils auf eine Länge von etwa zwei Minuten gekürzt, um vergleichbare Voraussetzungen zu schaffen. Die Delexikalisierung wurde mit dem Freeware Programm PRAAT durchgeführt. Das manipulierte Sprachsignal enthält nach der Bearbeitung mit PRAAT nur noch prosodische Merkmale wie Intonation, Rhythmus und Intensität. Indem die lexikalische Nachricht aus dem Hörbeispiel entfernt wurde, können die Hörer nicht mehr durch den Inhalt des jeweiligen Textes in ihrer Auswahl beeinflusst werden und die Studie richtet sich konkret auf die Prosodie.

3.1.3 Untersuchungsdesign

Bei der vorliegenden Studie handelt es sich um eine Hypothesen prüfende Untersuchung, weil explizite Vorkenntnisse im Bereich der Prosodie und des Poetry Slams bestehen und eine präzise Hypothesenformulierung ermöglichen. Es sollen außerdem neue Ideen und Hypothesen zur Beschreibung prosodischer Phänomene in der Slam-Vortragsweise generiert werden, daher kann hier von einem explorativen Wesen der Studie gesprochen werden (vgl. Bortz, Döring 2006:490). In einem analytischen Teil werden mit Hilfe der Ergebnisse der Umfrage weitere Hypothesen zu Prosodie im Poetry Slam generiert.

Die Entscheidung, eine quantitative Analyse durchzuführen, beruht unter Anderem auf der Tatsache, dass in Bezug auf Poetry Slam und dessen Prosodie bei den Probanden vermutlich noch kein Fachwissen besteht und sie ebenso wie die Verfasserin dieser Arbeit Neuland erkunden. Es handelt sich zudem möglicherweise nicht um trainierte Hörer, die spezifische Auskünfte über prosodische Merkmale geben können. Es sind hauptsächlich Laien, die laut Elena Travkina (2010), die sich im Rahmen ihrer „Sprechwissenschaftliche[n] Untersuchungen zur Wirkung vorgelesener Prosa (Hörbuch)“ mit dieser Problematik beschäftigte, zwar zum analytischen Hören in der Lage sind, das Gehörte aber „mit ihren eigenen Hörmustern und Hörgewohnheiten“ vergleichen und „auch ihre Persönlichkeit, individuelle Erfahrungen und Kenntnisse auf das Gehörte“ übertragen (S.111). Daher wäre ein detailliert-qualitativer Fragebogen nicht angebracht. Um eine Tendenz festzustellen, bietet sich eine quantitative Studie mit mehreren Auswahlmöglichkeiten an. Bortz und Döring positionieren sich bezüglich quantitativ-explorativer Forschung folgendermaßen:

> „Numerische Daten stellen Wirklichkeitsausschnitte in komprimierter, abstrakter Form dar. Überraschende Effekte und prägnante Muster in den Daten lenken die Aufmerksamkeit auf Phänomene, die der Alltagsbeobachtung möglicherweise entgangen wären. Ziel der quantitativen Explorationsmethoden ist es deshalb, Daten so darzustellen und zusammenzufassen, dass derartige Muster problemlos erkennbar werden.“ (Bortz, Döring 2006:269)

3.1.4 Instrumente

Um möglichst viele Probanden zu gewinnen, wurde die Studie über Facebook und StudiVZ veröffentlicht. Bei Facebook wurde zunächst eine Veranstaltung gegründet. Bei StudiVZ wurde eine Gruppe gegründet. Beide trugen den Titel „Umfrage zu meiner Master Arbeit“ und wurden über das persönliche Profil der Verfasserin dieser Arbeit veröffentlicht.

Bei Facebook sagt die zu einer Veranstaltung eingeladene Person entweder zu, ab, oder beantwortet die Anfrage nicht. Durch diesen Vorgang ist es den Probanden freigestellt, ob sie die Teilnahme an der Studie öffentlich nachvollziehbar machen möchten oder nicht. Wer bei StudiVZ an der Studie teilnehmen wollte, konnte in die Gruppe eintreten, sich aber auch ausschließlich von dem zitierten Erklärungstext leiten lassen und seine Teilnahme anonym halten. Anonymität und Datenschutz wird in einem Abschnitt des Lehrbuches von Bortz und Döring (2006:45) explizit erwähnt und deren Einhaltung als Bedingung für die Durchführung einer empirischen Studie gesetzt. Die Links führten zu den delexikalisierten Hörbeispielen, die in Gruppen zusammengefasst bei Youtube veröffentlicht wurden. Dies gewährleistete, dass das Hören und Beurteilen der Hörbeispiele zeitlich und lokal ungebunden durchgeführt werden konnte und diese Punkte keinen Widerstand zur Teilnahme auslösen konnten (vgl. Bortz, Döring 2006:45).

Im Anschluss daran kamen die Teilnehmer über einen Link zu dem Online-Abstimmungs-Tool Doodle, bei dem mehrere Antworten anonym-wählbar waren. Die Teilnehmer konnten ihren Namen eintragen, um so die tatsächliche Teilnahme zu beweisen und Nachfragen zu gewähren.

Dabei war es möglich, die Antworten der vorherigen Teilnehmer einzusehen. Dies ist zum einen eine Eigenschaft der Plattform *Doodle* selbst, zum anderen hat es den Vorteil, dass sich die Teilnehmer, wenn sie weniger als vier Beispiele bearbeiten wollten, sich gleichmäßig auf die Gruppen verteilen konnten. Die Wahl konnte dann noch verändert werden. Wurde dies vorgenommen, kann es im Verlauf nachgelesen und geprüft werden. Folgende Wege konnte ein Teilnehmer, der nur ein Hörbeispiel bearbeitet hat, bei dem erläuterten Untersuchungsdesign gehen:

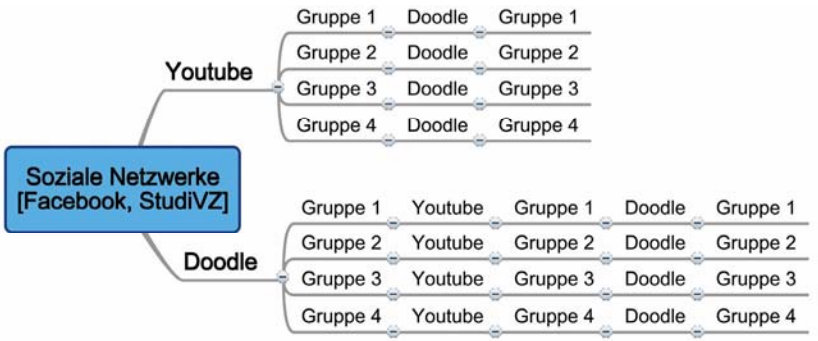


Abbildung 3. Mögliche Wege der Teilnehmer.

Wie deutlich zu erkennen ist, mussten Teilnehmer, die sich an den bisherigen Ergebnissen orientieren wollten, einen Arbeitsschritt mehr durchführen als Teilnehmer, die sich zuerst die Youtube-Videos oder eines davon angesehen haben. Daher ist es unwahrscheinlich, dass dieser Mehraufwand angestrengt wurde.

3.1.5 Gütekriterien einer Messung

Andreas Diekmann (2005:216) führt in dem Grundlagenwerk „Empirische Sozialforschung“ Gütekriterien auf, an denen die Objektivität, Reliabilität und Validität der gewonnenen Daten gemessen werden kann.

Zur Durchführungsobjektivität der vorliegenden Studie ist zu sagen, dass die Teilnehmer möglicherweise unterschiedliche Bedingungen hatten. So konnte ein Teilnehmer, der zu einem späteren Zeitpunkt nach Freischaltung der Umfrage seine Vermutung abgegeben hat, die Ergebnisse der vorherigen Teilnehmer sehen. Dies kann unter Umständen die Ergebnisse beeinflusst haben. Wie nachfolgend aufgeführt, bestanden kaum Anreize für die Teilnehmer, sich an den vermeintlich richtigen Ergebnissen der Vorgänger zu orientieren. Bis auf diese Tatsache ist die Durchführungsobjektivität als hoch zu betrachten, da der Korpus für alle Teilnehmer gleich war und jeder, der einen internetfähigen Rechner hat, auf die Daten zugreifen konnte. Zur Klangqualität der Hörbeispiele kann hier keine Aussage getroffen werden, da es nicht möglich ist, die technischen Voraussetzungen des jeweiligen Computers zu ermitteln.

Um herauszufinden, ob Daten dem Kriterium der Reliabilität gerecht werden, können verschiedene statistische Tests verwendet werden. Der bei dieser Studie angewendete Test wird im nächsten Abschnitt erläutert.

Anhand der im zweiten Teil dieser Arbeit durchgeführten Analyse der Hörbeispiele kann festgestellt werden, welche prosodischen Parameter möglicherweise ausschlaggebend für die Wahl der Untersuchungsteilnehmer waren. Welche oder wie viele dieser Kriterien Einfluss auf den einzelnen Teilnehmer hatten, wird allerdings nicht feststellbar sein, da im Anschluss an die Vermutungsäußerung in Form von Multiple-Choice-Tests keine weitere Befragung durchgeführt wurde. Diese kann eventuell im Rahmen einer weiteren Forschung in Zusammenhang mit den ausgewerteten Daten und den neu entstandenen Ideen und Hypothesen ausgefeilt werden. Die Untersuchung ist insofern intern valide, dass die Ergebnisse eindeutig auf die Unterschiedlichkeit der einzelnen Klangbeispiele zurückzuführen ist. Bei der Fragestellung, ob eine Unterscheidung in der Praxis möglich ist oder nicht, kann eine Tendenz erkennbar werden.

Da lediglich eine Auswahl zwischen den einzelnen Klangbeispielen zu treffen ist und das Untersuchungsergebnis keinen weiteren Einfluss auf die Probanden hat, gibt es keine Gründe für gefährdende „Confounder“ (Einflussfaktoren) wie mangelnde instrumentelle Reliabilität (vgl. Bortz, Döring 2006:503).

Das Untersuchungsinstrument erfasst unter gleichwertigen Auswahlmöglichkeiten die Vermutungen der Teilnehmer. Mangelnde instrumentelle Validität oder „Hawthorne-Effekte“ werden auch dadurch ausgeschlossen, dass eine anonyme Teilnahme an der Studie möglich ist. Wem seine Ergebnisse unangenehm sind, kann diese anonymisieren oder die Teilnahme verweigern.

Das Bewusstsein, einer Freundin bei Facebook oder StudiVZ durch die Teilnahme einen Gefallen zu tun, verpflichtet die Probanden zur gewissenhaften Ausführung der Arbeitsanweisungen, da das Ausmaß der Schädigung im Falle einer Missachtung oder vorsätzlich-falscher Angaben bekannt ist. Personen, die ohne persönliche Verknüpfung an der Studie teilgenommen haben und zum Beispiel eine generelle Aversion gegen Studenten oder Ähnliches haben und denen die Schädigung zum Vorteil gereichen könnte, werden erstens durch die Größe der Studie vereinzelt und zweitens wäre eine Teilnahme mitsamt Hören der Klangbeispiele höchstwahrscheinlich zu aufwendig für einen Trittbrettfahrer. Außerdem gab es keine Aussicht auf eine Belohnung für richtige Ergebnisse, bis auf das menschliche Bedürfnis, mit einer Auswahl richtig liegen zu wollen. Um eventuelle Zweifel an der Reliabilität oder Validität der Daten auszuschließen, wird an dieser Stelle mit Hilfe eines statistischen Tests geprüft, ob eine Beeinflussung durch die Sichtbarkeit der Ergebnisse der vorangegangenen Teilnehmer stattgefunden hat.

Diesem Test liegt die Annahme zugrunde, dass eine Beeinflussung sich durch „Abgucken“ vom Vorgänger oder von den Vorgängern äußert. Dies würde bedeuten, dass die Antworten der Probanden sich im Verlauf der Studie angepasst haben und weniger unterschiedlich getippt wurde. Martin Förster ist an der Universität Flensburg am Zentrum für Methodenlehre (ZML) tätig und half der Autorin dieser Abschlussarbeit bei der Bearbeitung dieser Problematik. Er entwickelte eine Möglichkeit zur Identifizierung von „Response Adaption“. Die Berechnungsschritte und ein Papier zu den einzelnen Werten der Studie wurden dem Anhang beigefügt. Ergebnis des Tests auf eine Verringerung der Streuung der Antworten ist, dass dies nicht der Fall war und somit keine Beeinflussung der Probanden stattgefunden hat. Lediglich bei zwei Beispielen konnte kein Wert ermittelt werden. In der ersten Gruppe konnte bei Beispiel

eins und Beispiel drei keine Streuung festgestellt und dementsprechend konnte auch keine Veränderung derer berechnet werden.

3.1.6 Stichprobenkonstruktion

Durch die Veröffentlichung der Studie bei StudiVZ und Facebook wurden sowohl Studenten als auch andere Teilnehmer für die Untersuchung gewonnen. Es kann davon ausgegangen werden, dass sich unter den Probanden auch geschulte Hörer befanden. Dies hat den Vorteil, dass sowohl der von Travkina beschriebene Effekt des Einsatzes von eigenen Hörmustern und Hörgewohnheiten bei ungeschulten Hörern gemessen wird, als auch differenzierte Wahrnehmungsmöglichkeiten der geschulten Hörer aufgenommen werden. In welchem Umfang dies zum Gesamtergebnis beigetragen hat, kann allerdings nicht ermittelt werden.

Der Stichprobenumfang ist nicht durch die Anlage der Studie reguliert. Es wird von einer Mindestgröße von zehn Personen pro Vergleichsgruppe ausgegangen, um eine Tendenz feststellen zu können.

Es handelt sich um eine Querschnittstudie, da sie einmalig und nicht wiederholt zu unterschiedlichen Zeitpunkten durchgeführt wurde. Die eingeladenen Personen setzen sich aus dem Bekanntenkreis der Verfasserin dieser Arbeit und dem Bekanntenkreis weiterer Personen zusammen, die ebenfalls Einladungen verschickt haben. Bei StudiVZ wurden ausschließlich Personen aus dem Bekanntenkreis der Verfasserin eingeladen. Eine Öffnung der Gruppe ist bei StudiVZ nicht möglich.

3.1.7 Durchführung

Zuerst wurde ein Benutzerkonto bei Youtube erstellt. Unter dem Benutzernamen „LateralerApproximant“ wurden die Videos, die die Klangbeispiele enthalten, hochgeladen. Danach wurden über ein bestehendes Benutzerkonto die Umfragen bei Doodle erstellt. Im Anschluss daran wurde der Erklärungstext für die Veranstaltung bei Facebook und die Gruppe bei StudiVZ formuliert. Dieser wurde gemeinsam mit den dazugehörigen Links mit der Veranstaltung veröffentlicht.

Am Dienstag, den 07.Juni 2011 wurde die Veranstaltung auf öffentlich geschaltet. Kurz nachdem dies vorgenommen wurde, erhöhte sich die Anzahl der Personen mit dem Status online von etwa 30 auf etwa 60 Personen, die anscheinend alle auf die Benachrichtigung reagierten. Um Unklarheiten zu vermeiden, wurde der ungewöhnliche Begriff der Delexikalisierung auf der Pinnwand expliziert. Über die Chat- und die Pinnwand-Funktionen bei Facebook kamen kurz darauf die ersten Nachfragen, was Poetry Slam, Delexikalisierung und Prosodie seien. Die Nachfragen ernst nehmend, wurden Erklärungstexte gesendet. Auf diese Erläuterungen reagierten die User positiv und bekundeten ihre Bereitschaft zur Teilnahme. Die Beteiligung an der Untersuchung belegt die Vorgehensweise ungeübter Hörer, aber auch, dass großes Interesse seitens der Probanden besteht, einen Beitrag zu der vorliegenden Arbeit zu leisten. Um weitere Hypothesen oder Vermutungen abfragen zu können, bietet sich eine qualitative-empirische Untersuchung an, die an dieser Stelle allerdings den Rahmen überschreiten würde. Die Veranstaltung wurde am 22.06.2011 geschlossen und die Ergebnisse der Umfrage unter dem Erklärungstext veröffentlicht.

Die Gruppe bei StudiVZ wurde am 10.06.2011 auf Nachfrage einiger Kommilitonen der Forscherin gegründet, die nicht bei Facebook angemeldet sind, aber trotzdem gerne an der Studie teilnehmen wollten. Insgesamt haben sich bis zum 24.06.2011 insgesamt 33 Personen in der Gruppe angemeldet. Zum Teil überschneiden sich Profile mit Personen, die bereits bei Facebook ihre Teilnahme zugesichert haben. Auch bei diesem Forum kam es zu Begriffsunklarheiten, die durch eine kurze Erklärung aufgelöst werden konnten. Die Doodle Umfrage wurde am 22.06.2011 geschlossen und die Teilnahme war ab diesem Tag nicht mehr möglich.

3.2 Ergebnisse

Bei Facebook haben 55 Personen zugesagt, an der Studie teilzunehmen. Bei 19 Usern ist die Teilnahme unsicher und 47 haben die Teilnahme abgelehnt. 358 Personen haben die Veranstaltungseinladung nicht beantwortet. Von den Personen, die Ihre Teilnahme angegeben haben, sind 25 männlich und 30 weiblich. Teilgenommen haben fast ausschließlich Studenten der Universitäten Flensburg oder Kiel oder junge Berufseinsteiger im Alter von 20 bis 30 Jahren. Unsicher waren sich Personen der gleichen Gruppen, die allerdings zum Teil lange nicht mehr in Kontakt zu der Veranstalterin standen. Zu den der Studie gegenüber ablehnend Positionierten gehören ebenfalls Personen der genannten Gruppen. Dabei ist auffällig, dass unter ihnen viele nicht an der Universität Flensburg studieren oder in ihrer persönlichen Karriere der Sprachwissenschaft relativ fern sind. Ebenfalls abgesagt hat eine ehemalige Dozentin der Veranstalterin. Die Informationen zu der Testgruppe sind den Profilinformationen der Probanden entnommen. Bei StudiVZ sind insgesamt 33 Personen in die Gruppe „Umfrage zu meiner Master Arbeit“ eingetreten. Davon sind 12 weiblichen und 20 männlichen Geschlechts. Die Teilnehmer sind im Alter von 20 bis 30 Jahren und zum Großteil Studenten und junge Berufseinsteiger, wie zum Beispiel Referendare. Tatsächlich teilgenommen haben bei Doodle für die erste Vergleichsgruppe 38 Personen, für die zweite Gruppe 28 Personen, für die dritte Gruppe 29 Personen und für die vierte Gruppe 31 Personen. Die angekreuzten Antworten sind in Tabelle 1 zusammengefasst. Mit Hilfe dieser Daten sollen nun folgende Fragen geklärt werden[2]:

- Ist Slam Poetry erkennbar?
- Wurde das Slam-Beispiel innerhalb der einzelnen Gruppen identifiziert?
- Gibt es einen Unterschied zwischen den Slam-Beispielen?

Zur Erörterung der ersten Frage, ob Slam innerhalb dieser Studie generell als erkennbar gelten kann, werden alle richtigen Antworten bei den Slam-Beispielen aller Gruppen daraufhin verglichen, ob der Anteil der richtig erkannten Werte signifikant über 50% Prozent liegt. Wären die richtigen Antworten nur zufällig zustande gekommen, wäre ihr Anteil nicht deutlich von 0,5 unterschieden. Um bestimmen zu können, ob der Anteil der richtigen Antworten über 50% liegt, muss dieser zunächst

[2] Bei der Auswertung der Daten stand der Autorin erneut Herr Förster zur Seite, der zur Beantwortung der vorangegangenen Fragen Möglichkeiten zur statistischen Berechnung aufzeigte.

ermittelt werden, um dann mit einem Signifikanzniveau von 0,1, das heißt mit 10% Irrtumswahrscheinlichkeit, mit Hilfe von t-Tests geprüft zu werden. Auf Basis einer von Herrn Förster erstellten t-Test-Matrix und der Ergebnisse der Studie wurde festgestellt, dass Slam Poetry in dieser empirischen Erhebung in der Summe signifikant-häufiger erkannt als nicht erkannt wurde ($T=0{,}99$; $df=122$; $p<0{,}001$). Das stützt die These, dass Texte, die in dem Rahmen eines Poetry Slams vorgetragen wurden, über prosodische Eigenschaften verfügen, die im Vergleich mit anderen mündlichen Präsentationsformen zu identifizieren sind.

In den insgesamt vier Beispielgruppen wurde die Frage nach der Vortragsform unterschiedlich oft richtig beantwortet. Daher stellt sich als Zweites die Frage, ob das Slam-Beispiel innerhalb der einzelnen Gruppen signifikant-häufiger richtig erkannt wurde. Um dies herauszufinden, wurde erneut die t-Test-Matrix von Herrn Förster verwendet. In der ersten Beispielgruppe wurde, wie in der nachgestellten Grafik abzulesen ist, bei dem Slam-Beispiel 29-mal richtig getippt. Im Verhältnis zu den sieben Falschantworten ist dies ein signifikanter Anteil ($T=0{,}99$; $df=35$; $p<0{,}001$). Das erste Slam-Beispiel war anscheinend relativ leicht zu erkennen.

Optionen	Häufigkeit der Antworten in jew.Gruppen			
	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
Beispiel 1 ist der Slam-Vortrag.	1	15	4	15
Beispiel 1 ist das vorgelesene Gedicht.	33	5	8	10
Beispiel 1 ist das vorgelesene Essay.	4	8	17	6
Keine Ahnung, was Beispiel 1 ist!	1	0	0	0
Beispiel 2 ist der Slam-Vortrag.	6	5	3	9
Beispiel 2 ist das vorgelesene Gedicht.	6	17	19	2
Beispiel 2 ist das vorgelesene Essay.	25	6	8	20
Keine Ahnung, was Beispiel 2 ist!	1	0	0	0
Beispiel 3 ist der Slam-Vortrag.	29	7	23	7
Beispiel 3 ist das vorgelesene Gedicht.	0	6	2	19
Beispiel 3 ist das vorgelesene Essay.	8	15	4	6
Keine Ahnung, was Beispiel 3 ist!	1	0	0	0
Teilnehmer insgesamt	38	28	29	31

Tabelle 1. Umfrageergebnisse.

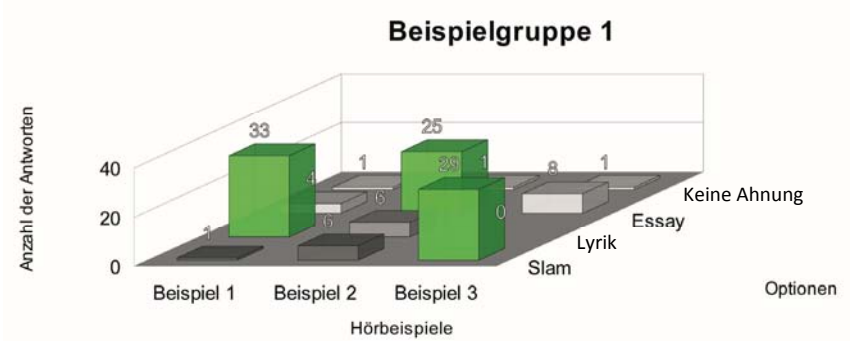


Abbildung 4. Umfrageergebnisse Beispielgruppe 1.

Das Slam-Beispiel in der zweiten Beispielgruppe wurde fünfzehn Mal richtig und damit nicht signifikant-häufiger richtig als falsch erkannt (T=0,78; *df*=25; p=0,56).

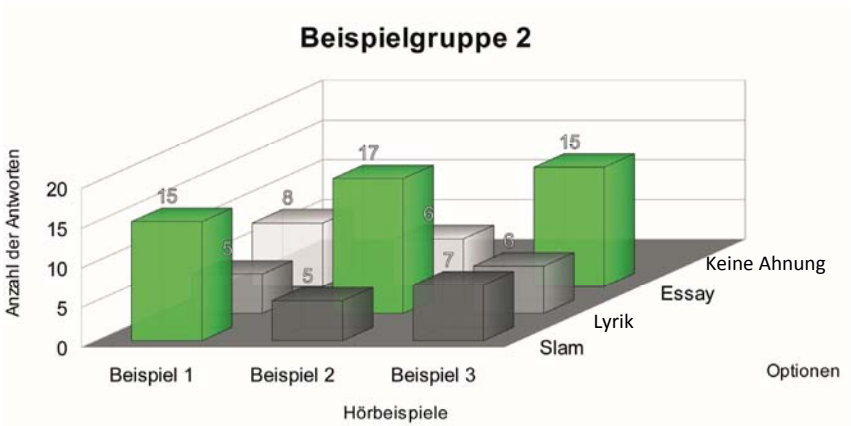


Abbildung 5. Umfrageergebnisse Beispielgruppe 2.

In Beispielgruppe 3 zeigt sich ein anderes Bild. Die 23 richtigen Antworten bei dem Slam-Beispiel bilden einen signifikant-größeren Anteil als die sieben falschen (T=0,99; *df*=28; p<0,001).

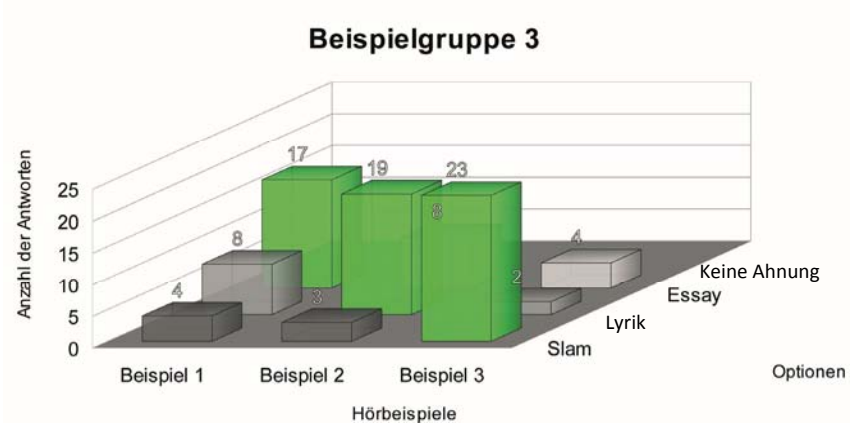


Abbildung 6. Umfrageergebnisse Beispielgruppe 3.

In der letzten Gruppe, in der die verschiedenen Präsentationsformen abgefragt wurden, kam es, wie in der Grafik veranschaulicht, zu in Hinblick auf die vorherigen Ergebnisse unerwarteten Werten. Der Anteil der falschen Antworten, der 22 Messungen umfasst, ist signifikant-größer als der Anteil der neun richtigen Antworten ($T=0{,}29$; $df=29$; $p<0{,}01$).

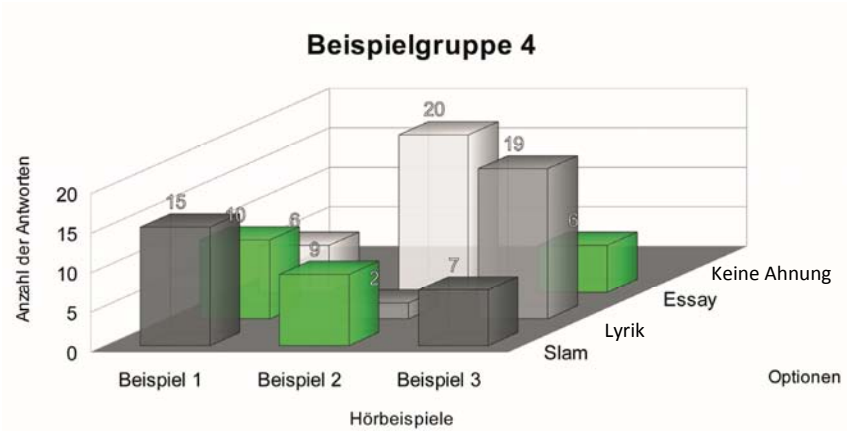


Abbildung 7. Umfrageergebnisse Beispielgruppe 4.

Die Tatsache, dass die Anteilsgrößen in den vier Beispielgruppen so unterschiedlich ausfallen, lässt darauf schließen, dass die Slam-Beispiele verschieden schwer zu identifizieren waren. Folglich ist die Frage angebracht, ob es einen Unterschied zwischen den Slam-Beispielen gibt. Dies wird exemplarisch an den Gruppen eins und zwei unter-

sucht. Dank der t-Test-Matrix, die von Herrn Förster zur Verfügung gestellt wurde, konnte ermittelt werden, dass sich die Anteile der richtigen Antworten bei den Slam-Beispielen der Beispielgruppen eins und zwei signifikant voneinander unterscheiden ($T=2{,}15$; $df= 28{,}5$; $p<0{,}01$).

3.3 Schlussfolgerung

In dieser Studie konnte bewiesen werden, dass Slam Poetry anhand von technisch-delexikalisierten Hörproben erkannt werden kann. Das heißt, dass diese Informationen enthalten, anhand derer Slam Poetry zu identifizieren und damit zu charakterisieren ist. Allerdings besteht ein Unterschied zwischen den einzelnen Slam-Beispielen und dies deutet unter Umständen an, dass die enthaltenen Merkmale verschieden ausgeprägt sind. Um Klarheit darüber zu gewinnen, welche Parameter möglicherweise ausschlaggebend für die Wahl der Probanden waren und welche Schlüsse in Hinblick auf die generellen prosodischen Eigenschaften von Slam Poetry zu ziehen sind, werden die einzelnen Hörproben prosodisch analysiert.

4. Prosodische Analyse der delexikalisierten Hörbeispiele

Die Analyse der technisch-delexikalisierten Hörbeispiele wurde mit dem Programm PRAAT durchgeführt.

4.1 Auswahl der Parameter

4.1.1 Durchschnittliche Grundfrequenz (F_0)

Zuerst wurde die durchschnittliche Grundfrequenz der einzelnen Hörbeispiele ermittelt. Wie bereits aus Christiane Miosgas Ausführungen entnommen, wird mit der Grundfrequenz die jeweils durchschnittlich realisierte Tonhöhe ermittelt. Dabei haben Frauen zumeist eine höhere Grundfrequenz als Männer. Eine erhöhte Tonlage kann auch emotionale Erregung bedeuten. Dieses prosodische Merkmal ist für die Analyse der delexikalisierten Hörproben besonders geeignet, da keine inhaltlichen Informationen durch die Stimmlage vermittelt werden und die Gattung nicht anhand derer zu erkennen ist.

PRAAT hat die Besonderheit, dass Signale zum Teil falsch interpretiert werden und sogenannte Oktavfehler zustande kommen. Gerade bei der Tonhöhe ist dies von Bedeutung. So kann zum Teil nicht festgestellt werden, ob es sich bei einem Ausschlag des Tonhöhenverlaufs um eine fehlerhafte Information handelt. Bei der Größe der Dateien sind die Oktavfehler allerdings zu vernachlässigen, weil sie nur vereinzelt auftreten und der Wert der tatsächlichen durchschnittlichen Grundfrequenz sehr nahe ist.

4.1.2 Regelmäßigkeit der Grundfrequenzmaxima

Welchen Rhythmus ein Signal in etwa hat und wie stark die Worte akzentuiert wurden, kann herausgefunden werden, indem die Schwankungen im Tonhöhenverlauf fest-

gehalten werden. Bei den delexikalisierten Tonbeispielen kann nicht von der lexikalischen Information auf Akzentuierung oder Rhythmus geschlossen werden. Daher ist es vonnöten, die Maxima der Grundfrequenz zur Analyse heranzuziehen. Besonders bei einem melodischen Akzent kann die Grundfrequenz als Indiz gelten. Mit Hilfe des Programms PRAAT, welches die Grundfrequenz linienartig darstellt, kann Zeitpunkt und Höhe des Maximums ermittelt werden. Da die Dateien relativ groß sind und die vollständige Analyse zu umfangreich wäre, wurden nur bestimmte Abschnitte des Signals verwendet. Dazu gehörten zehn Sekunden am Anfang, zehn Sekunden aus der Mitte und zehn Sekunden am Ende des Signals, um einen möglichst umfassenden Eindruck zu bekommen.

Zum Teil stellte es sich als schwierig heraus, Grundfrequenzmaxima klar zu definieren. Das lag daran, dass einige Hörbeispiele nur über wenige Ausprägungen verfügten und nicht klar war, ab wann kein Maximum mehr festzustellen ist. Infolge dessen kam es zustande, dass bei der Analyse von Jürgen von der Lippes Interpretation des Gedichts kaum Grundfrequenzmaxima zu verzeichnen waren. Die exakten maximalen Ausprägungen des Grundfrequenzverlaufes sind dem Spektrum im Anhang zu entnehmen.

Die Grundfrequenzmaxima werden Abschnitt um Abschnitt als Strichdiagramme dargestellt, sodass die Abstände zwischen ihnen deutlich werden. Dabei zeigt die x-Achse einen Zeitstrahl in Millisekunden und die y-Achse die drei Hörbeispiele an.

4.1.3 Variabilität der Intensität

Als Drittes wurde die Variabilität der Intensität der Signale untersucht. Das Merkmal der Lautstärke bedingt ebenfalls Komponenten wie Rhythmus und Akzentuierung und kann daher für die Identifizierung einzelner Präsentationsformen wichtig sein. Die Tonbeispiele, die für diese Studie verwendet wurden, sind in unterschiedlicher Qualität und die Lautstärke wurde nicht manipuliert, sodass sie sich nicht auf einem einheitlichen Level befanden. Daher wäre eine Untersuchung des Intensitätsverlaufes nicht sinnvoll. Allerdings bedingt dies keine Veränderung des verwendeten Lautstärkespektrums. Infolge dessen eignet sich die Variabilität der Intensität an dieser Stelle. Sie stellt die Differenz zwischen dem höchsten Gipfel und dem tiefsten Gipfel und damit den Dynamikbereich des Sprechers dar. Dafür wird per Augenmaß das ungefähre Minimum und Maximum der Intensitätskurve in den extrahierten Sekundenabschnitten erfasst und notiert.

4.1.4 Sprechgeschwindigkeit

Gerade in Hinblick auf die Hypothesen, die auf der Grundlage der Literatur zu Slam Poetry aufgestellt wurden, kann die Sprechgeschwindigkeit des Signals ausschlaggebend für die Wahl eines Hörbeispiels gewesen sein. Und eventuell charakteristisch für Slam Poetry. Um sie in einem delexikalisierten Signal zu ermitteln, werden die Silben pro Minute errechnet. Dafür werden aus den drei je zehn Sekunden andauernden Abschnitten alle stillen Pausen entfernt.

4.1.5 Pausen im Signal

Wie bei Miosga (2006:153 ff.) beschrieben, wird zusätzlich die Anzahl und Länge der jeweiligen Pausen festgehalten, um den Gesamtanteil an der Sprechzeit, sowie ihre Häufigkeit und Dauer zu analysieren. Ihre Position innerhalb eines Sprechbeitrages oder ihr Wesen kann aufgrund der Delexikalisierung nicht festgestellt werden und ist daher auch für die empirische Studie nicht von Bedeutung.

4.2 Einzelanalyse

Bei der Analyse der einzelnen Gruppen wurde entsprechend dem Aufbau der empirischen Studie vorgegangen. Da ein Vergleich aller Slam-Beispiele nicht praktikabel ist, wird damit begonnen, die Unterschiede zwischen den jeweiligen Hörbeispielen herauszustellen. Im Anschluss daran werden die Hörbeispiele zu Slam Poetry gegenübergestellt.

4.2.1 Durchschnittliche Grundfrequenz

Genre: Interpret - *„Titel“*		mean pitch in selection
1	Slam: Andy Strauß – *„Stefanie Hitzer“*	268 Hz
2	Lyrik: Peter Reimers – *„Die Bürgschaft“*	171 Hz
3	Essay: Harald Eggebrecht – *„Von Risiken und Ängsten“*	119 Hz

Tabelle 2. Durchschnittliche Grundfrequenz Beispielgruppe 1.

Die jeweiligen durchschnittlichen Grundfrequenzen der Hörbeispiele aus der ersten Gruppe unterscheiden sich deutlich voneinander. Während Harald Eggebrechts und Peter Reimers' Sprache durchschnittlich unter 200 Hertz realisiert wird, kommt es bei Andy Strauß zu einem für Männer eher extraordinären Wert von über 250 Hertz. Der durchschnittliche Grundfrequenzbereich liegt bei Männern in der Regel zwischen 100 und 200 Hertz. Hört man sich dieses Beispiel nicht manipuliert an, ist festzustellen, dass er seine Stimme stark verzerrt und zum Teil die lautlichen Merkmale einer Frau imitiert. Damit bestätigt sich die Hypothese, dass Slam-Poeten ungewöhnliche sprachliche Mittel einsetzen, welche zum Beispiel anhand der durchschnittlichen Grundfrequenz zu erkennen sind. Möglicherweise hat der große Unterschied zwischen den Werten auch das Urteil der Probanden bedingt.

Genre: Interpret - *„Titel“*		mean pitch in selection
1	Slam: Wolfgang Lüchtrath – *„Der Leuchtwart“*	222 Hz
2	Lyrik: Jürgen von der Lippe – *„Das Ideal“*	123 Hz
3	Essay: Reinhard Kahl – *„Die Wiederentdeckung des Übens“*	146 Hz

Tabelle 3. Durchschnittliche Grundfrequenz Beispielgruppe 2.

In der zweiten Beispielgruppe ist dieses Phänomen ebenfalls zu beobachten. Während Jürgen von der Lippes und Reinhard Kahls Hörbeispiele eine für Männer übliche durchschnittliche Grundfrequenz aufweisen, liegt der Wert von Wolfgang Lüchtrath wesentlich höher. Das kann daran liegen, dass Lüchtrath diesen Text sehr emotionalisiert und auf eine aggressive Art und Weise vorträgt.

Die Übereinstimmung zwischen den ersten beiden Beispielgruppen ist besonders interessant, betrachtet man die Zuordnungen von Christiane Miosga (2006:146). Einer erhöhten Stimmlage schreibt sie einen hohen Aktivierungsgrad durch zum Beispiel Freude, Angst, Wut oder Engagement zu. Gleichzeitig kann eine hohe Stimmlage Submissivität, Aufdringlichkeit oder Quängeligkeit bedeuten. Im Gegensatz dazu stehen die Zuweisungen bei einer erniedrigten Stimmlage, wie sie bei den Essay- und Lyrik-Beispielen zu vermuten sind. Neben selbstexpressiven Funktionen wie dem Eindruck von Dominanz, Weisheit und Souveränität besteht die pragmatische Funktion der Relevanz. In diesem Zusammenhang wird Wichtigkeit, Beruhigung und Nachdruck genannt.

Es können Schlüsse in Hinblick auf die Zielgruppe der jeweiligen sprachlichen Textumsetzung gezogen werden. Während die auf einer Audio-CD befindlichen Gedichte nur in einer entspannten und beruhigenden Umgebung vorstellbar sind und auch die Rezeption des Essays Ruhe bedarf, findet der Slam-Vortrag, der aufwühlen und anstacheln soll, in einem Club mit vielen Nebengeräuschen statt. So verzerrt Andy Strauß seine Stimme auf komische Art und Weise und Wolfgang Lüchtrath schlägt einen kämpferisch-aufgebrachten Ton an. Dies stimmt mit der Hypothese überein, die anhand der Begriffsherkunft aufgestellt wurde: *Zusammenfassend ist zu sagen, dass Slam Poetry durch eine erhöhte Sprechgeschwindigkeit und den unkonventionellen Umgang mit sprachlichen Mitteln wie zum Beispiel der Tonhöhe charakterisiert ist.*
Bei der Untersuchung der durchschnittlichen Grundfrequenz bei den weiblichen Sprecherinnen ist bei den Beispielen aus Beispielgruppe drei Ähnliches zu verzeichnen.

Genre: Interpret - *„Titel“*		mean pitch in selection
1	Slam: Pauline Füg – *„N Orientierungslos“*	280 Hz
2	Lyrik: Weibliche Sprecherin – *„Das Riesenspielzeug“*	177 Hz
3	Essay: Cora Stephan – *„Frauen können alles und machen es niemandem recht“*	191 Hz

Tabelle 4. Durchschnittliche Grundfrequenz Beispielgruppe 3.

Bei Frauen ist ein Wert um die 200 Hertz üblich. Im Vergleich dazu und zu den anderen beiden Beispielen liegt der Wert von Pauline Füg wesentlich höher. Auch sie spricht sehr engagiert. Betrachtet man allerdings die letzte Beispielgruppe, so zeichnet sich ein anderes Bild. Hier ist der Slam-Vortrag das Beispiel mit der niedrigsten durchschnittlichen Grundfrequenz.

	Genre: Interpret - *„Titel“*	mean pitch in selection
1	Slam: Anke Fuchs – *„Was wisst ihr denn schon davon“*	195 Hz
2	Lyrik: Ute Beckert – *„Havelland“*	212 Hz
3	Essay: Christiane Grefe – *„Auf dem Weg ins Solarzeitalter“*	196 Hz

Tabelle 5. Durchschnittliche Grundfrequenz Beispielgruppe 4.

Anke Fuchs spricht in den ausgewählten Abschnitten in einer für Frauen normalen Tonhöhe. An diesem Hörbeispiel ist erkennbar, wie unterschiedlich Slam-Texte sein können. Während Andy Strauß einen komischen Text vorträgt, geht es bei Pauline Füg und Wolfgang Lüchtrath um existenzielle Fragen über Gesellschaft und Persönlichkeit, die in einem aggressiven Tonfall gestellt werden. In „Was wisst ihr denn schon davon“ verwendet Anke Fuchs dagegen eine gewöhnliche Stimme und verleiht ihren Worten auf andere Weise Nachdruck. Ihr Text ist sehr ernst und behandelt die Interaktion von sich gegenseitig weitgehend-unbekannten Personen.

Aufgrund der relativen Nähe der Werte zueinander und zur üblichen Werten der Grundfrequenz war es für die Probanden bei dieser Vergleichsgruppe besonders schwer, das Slam-Beispiel zu identifizieren. Beim Vergleich aller Slam-Beispiele miteinander fällt auf, dass Anke Fuchs auch hier eine besondere Position einnimmt.

	Interpret - *„Titel“*	mean pitch in selection
1	Andy Strauß – *„Stefanie Hitzer“*	268 Hz
2	Wolfgang Lüchtrath – *„Der Leuchtwart“*	222 Hz
3	Pauline Füg – *„N' Orientierungslos“*	280 Hz
4	Anke Fuchs – *„Was wisst ihr denn schon davon“*	195 Hz

Tabelle 6. Durchschnittliche Grundfrequenz aller Slam-Beispiele.

Möglicherweise ist es daher sinnvoll, im Fall einer weiteren Studie eine Differenzierung nach verschiedenen Textsorten vorzunehmen, ähnlich, wie sie nach Westermayr im Kapitel 3.3 ausgearbeitet wurde. Auf der einen Seite könnten so genrespezifische prosodische Merkmale besser erfasst werden wie zum Beispiel die eines komischen Textes, auf der anderen Seite ist die Vielfalt ein Charakteristikum von Slam Poetry. Es stellt sich daher die Frage, ob dadurch nicht eine Sammlung von Texten aufgeteilt wird, die im Grunde zusammengehören.

Abschließend zur Untersuchung der durchschnittlichen Grundfrequenz ist zu sagen, dass bei drei von vier Slam-Hörbeispielen von einer erhöhten durchschnittlichen Grundfrequenz und damit von einer hohen Stimmlage gesprochen werden kann. Damit wurde die Hypothese bestätigt, dass Slam-Poeten meist ungewöhnliche sprachliche Mittel verwenden und ihre Sprache einen erhöhten Aktivierungsgrad zeigt.

4.2.2 Regelmäßigkeit der Grundfrequenzmaxima

Es wurde zu jedem 10-Sekunden-Abschnitt ein Strichdiagramm erstellt. Auf der y-Achse haben die Hörbeispiele aus der ersten Beispielgruppe folgende Anordnung:
0) Slam: Andy Strauß, 1) Lyrik: Peter Reimers, 2) Essay: Harald Eggebrecht.

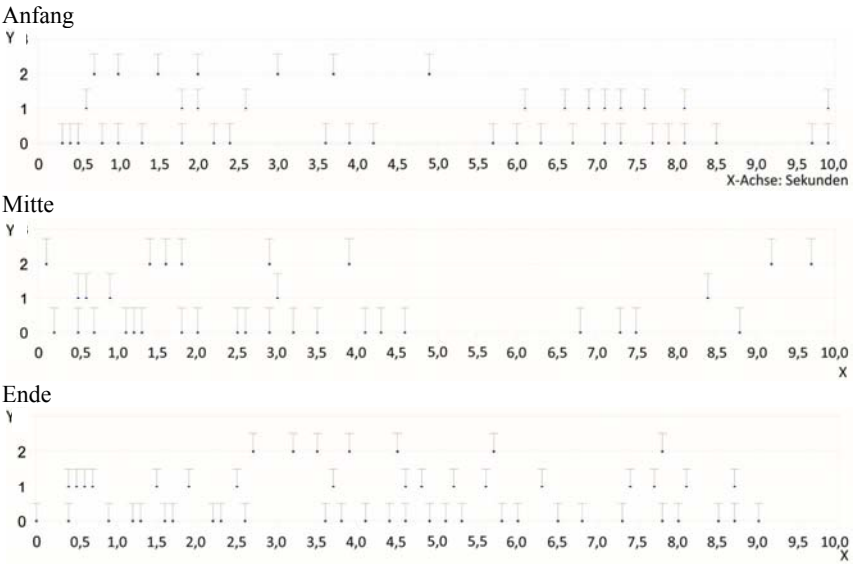


Abbildung 8. Regelmäßigkeit der Grundfrequenzmaxima Beispielgruppe 1.

Auf den ersten Blick fällt auf, dass das Slam-Beispiel über wesentlich mehr Grundfrequenzmaxima verfügt als das Lyrik- oder Essay-Beispiel. Gefolgt von dem Gedicht, welches vor allem im dritten Abschnitt über viele Grundfrequenzmaxima verfügt. Das Essay enthält am wenigsten Akzentuierungen durch den Tonhöhenverlauf und zeigt bei diesen nur bedingt Regelmäßigkeit.

Im Gegensatz dazu scheint das Gedicht sehr gleichmäßig-rhythmisiert zu sein, weil die Abstände zwischen den Grundfrequenzmaxima in den akzentuierten Teilen des Signals ähnlich lang sind. Im mittleren Abschnitt des Lyrik-Beispiels kommt es zu einer langen Phase ohne Grundfrequenzmaximum, was auf eine lange Stille hindeuten kann.

Es kommen nur wenige nicht akzentuierte Teile des Signals vor und die Grundfrequenzmaxima folgen regelmäßig, aber schnell aufeinander. Es ist erstaunlich zu sehen, wie rhythmisch Andy Strauß in dem Bericht über Stefanie Hitzer mit seiner Stimme umgeht und wie ausgeglichen die Abstände zwischen den Grundfrequenzmaxima sind.

In der theoriebasierten Exploration wurde die liedartige Struktur und die damit einhergehende zu vermutende Rhythmik von Slam Poetry herausgestellt. Ein spezielles Reimschema kann hier nicht zugeordnet werden. Parallelen zur mündlichen Dichtung,

wie von Ong dargestellt, bestehen insofern, als dass diese als stark rhythmisch beschrieben wird.

Insgesamt kann festgehalten werden, dass bei der ersten Beispielgruppe eine Übereinstimmung mit den aus der Literatur geschlossenen Hypothesen zu verzeichnen ist. Miosga (2006) macht in ihren Ausführungen deutlich, dass es durch den Einsatz von vermehrt oder vermindert vielen Grundfrequenzmaxima „zu situativen Missverständnissen oder Persönlichkeitszuschreibungen von monoton, affektlos, antriebsarm oder langweilig bis zu übertrieben, affektiv, affektiert, überdreht oder cholerisch kommen“ kann (S.180). Gerade bei einem Vortrag wie dem von Andy Strauß liegt die Vermutung nahe, dass hier ein überdrehter oder cholerischer Eindruck entstehen sollte.

In der zweiten Beispielgruppe wurde für die Diagramme folgende Reihenfolge gewählt: 0) Slam: Wolfgang Lüchtrath, 1) Lyrik: Jürgen von der Lippe, 2) Essay: Reinhard Kahl.

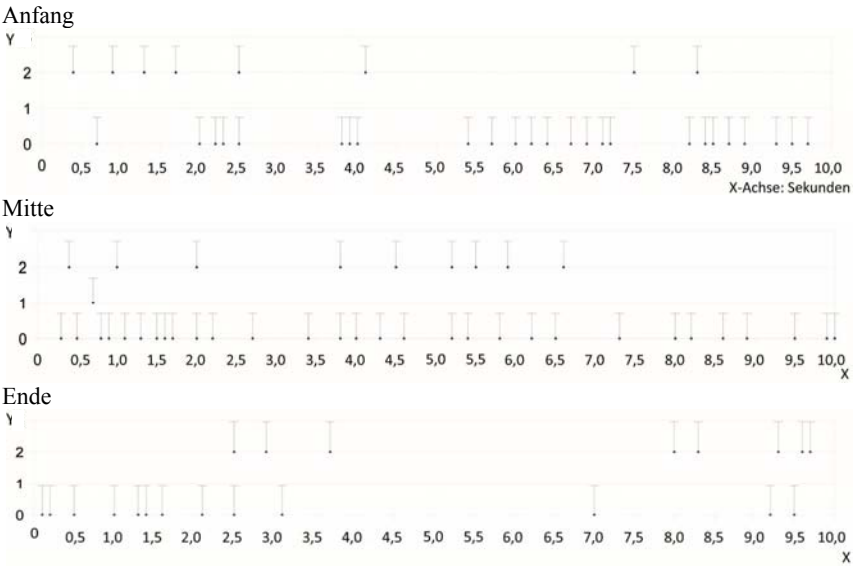


Abbildung 9. Regelmäßigkeit der Grundfrequenzmaxima Beispielgruppe 2.

Bei der ersten Rezeption der Diagramme dominiert in der Menge der Grundfrequenzmaxima auch in dieser Beispielgruppe das Poetry Slam-Beispiel. Gerade im mittleren Abschnitt verfügt es über deutlich mehr Grundfrequenzmaxima als die Vergleichsbeispiele, während das Gedicht kaum solche vorweisen kann und das Essay im Verhältnis wesentlich weniger akzentuiert wird. Kommen im Essay-Beispiel Grundfrequenzmaxima vor, wie zum Beispiel im ersten Abschnitt, sind sie relativ gleichmäßig voneinander entfernt und deuten auf den gezielten Einsatz der Tonhöhe zur Strukturierung des lautlichen Signals hin. Im letzten Abschnitt kommt es zu einer Häufung der Grundfrequenzmaxima.

Dem Essay-Beispiel ähnlich können bei dem Slam-Beispiel unterschiedliche Phasen im Signal angenommen werden. Anfänglich mit unbetonten Phasen versehen, wird das Slam Poetry Beispiel im mittleren Abschnitt stark rhythmisch und im letzten Abschnitt wieder mit mehreren Lücken versehen. In Phasen häufiger Grundfrequenzmaxima besteht in großen Teilen eine Regelmäßigkeit in den Abständen zwischen den Grundfrequenzmaxima und deuten auf den bewussten Einsatz derer hin.

Im Vergleich zu der vorangegangenen Beispielgruppe zeigt sich ein ähnliches Bild und eine Übereinstimmung mit den herangezogenen Hypothesen. Unterschiedlich ist, dass das Gedicht in der zweiten Beispielgruppe deutlich weniger rhythmisch ist und in dem Slam-Beispiel häufiger unbetonte Phasen eingesetzt werden. Unter Umständen dient dies der Strukturierung des Textes und der Markierung von wichtigen Textpassagen. Außerdem wird hier der Eindruck des Cholerischen abgeschwächt, wie ihn Miosga für vermehrten Einsatz von Tonhöhenschwankungen beschrieb (vgl. Miosga 2006:180).

Die y-Achse der dritten Beispielgruppe wurde folgendermaßen eingeteilt: 0) Slam: Pauline Füg, 1) Lyrik: Weibliche Sprecherin, 2) Essay: Cora Stephan.

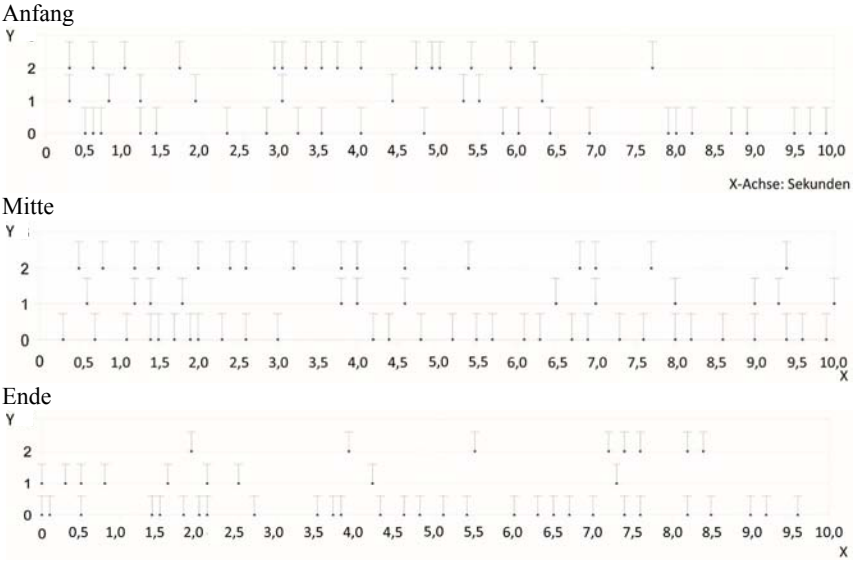


Abbildung 10. Regelmäßigkeit der Grundfrequenzmaxima Beispielgruppe 3.

Der Gesamteindruck zeigt eine deutliche Dominanz des Slam-Beispiels in der Anzahl der Grundfrequenzmaxima. Allerdings verfügen in dieser Beispielgruppe das Essay- und das Lyrik-Beispiel ebenfalls über im Vergleich zu den anderen Beispielgruppen viele Grundfrequenzmaxima. Die Beispiele der ersten beiden Beispielgruppen werden von Männern gesprochen, deren Sprache laut Miosga (2006:141) verminderte emotionale Expressivität zuzuordnen ist: *„Frauen realisieren einen durchschnittlichen Stimmumfang von vier bis fünf Tönen, Männer nutzen nur ca. drei Töne ihres Stimmumfangs, sodass ihnen […] mehr „informierende Autorität zugesprochen wird“*.

Zu der erhöhten Stimmlage von Frauen kommt also die größere Tonhöhenvariation und damit der vermehrte Einsatz von Grundfrequenzmaxima, die *„hier häufig mit weiblichem, unkontrolliertem Emotionsausdruck in Verbindung gebracht [werden], der Unterordnung, Nähe und Wertschätzung auf der Beziehungsebene signalisiert“* (Miosga 2006:140). Diese klischeehafte Interpretation erlebt laut Miosga zurzeit eine Neubewertung und die Tatsache, dass eine Gedichtinterpretation eines männlichen Lyrikers von einer Frau inszeniert wird, illustriert dies. Nichtsdestotrotz kann festgestellt werden, dass die weiblichen Sprecherinnen mehr Grundfrequenzmaxima verwenden als die männlichen.

Zwischen den einzelnen Beispielen in der dritten Beispielgruppe bestehen ebenfalls Unterschiede. Das Lyrik- und das Essay-Beispiel distanzieren sich insofern voneinander, als dass sie in unterschiedlichen Phasen ähnlich viele, aber verschiedenverteilte Grundfrequenzmaxima aufweisen. Betrachtet man die einzelnen Abschnitte, so ist zu sagen, dass das Gedicht anfangs weniger Grundfrequenzmaxima besitzt, welche aber rhythmischer angeordnet sind als bei dem Essay. Im mittleren Abschnitt wird das Essay wesentlich rhythmischer und de Grundfrequenzmaxima kommen in regelmäßigen Abständen. Im letzten Abschnitt kann von etwa gleich vielen Grundfrequenzmaxima ausgegangen werden, welche spiegelverkehrt-angeordnet sind. Es werden mehr unbetonte Phasen eingesetzt als in den vorherigen beiden Abschnitten.

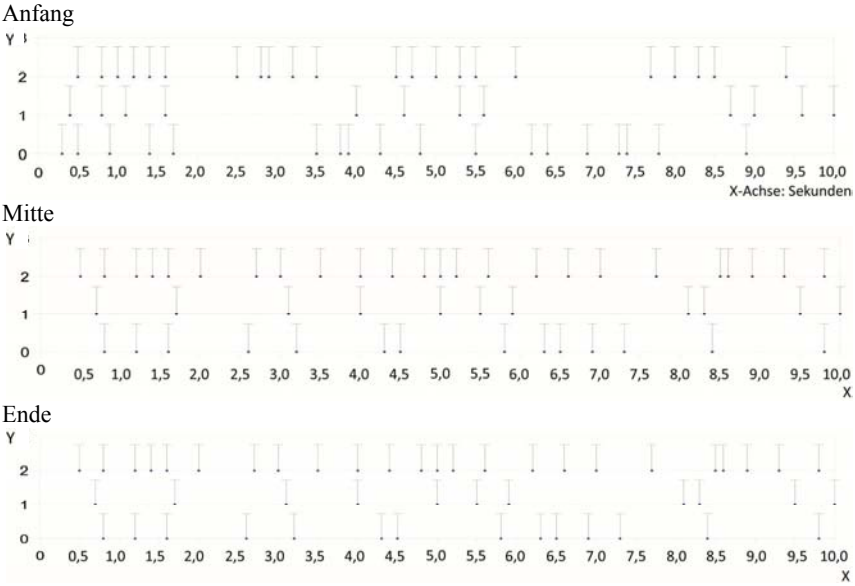


Abbildung 11. Regelmäßigkeit der Grundfrequenzmaxima Beispielgruppe 4.

Das Slam Poetry Beispiel zeigt eine durchgängig-gleichmäßige, häufige Akzentuierung und kann möglicherweise bereits als isoton angesehen werden. Pauline Füg nutzt in allen drei Abschnitten ihre Tonhöhe zur Rhythmisierung ihres Textes und zieht so

einen roten Faden durch ihren Vortrag. Auch hier können die in der theorisbasierten Exploration generierten Vermutungen belegt werden. Die vermehrte Akzentuierung deutet auf die Basis von Slam Poetry hin: Lieder, Gedichte und Rap.

Die vierte Beispielgruppe in Abbildung 11 wird wie folgt aufgeteilt: 0) Essay: Christiane Grefe, 1) Lyrik: Ute Beckert, 2) Slam: Anke Fuchs.

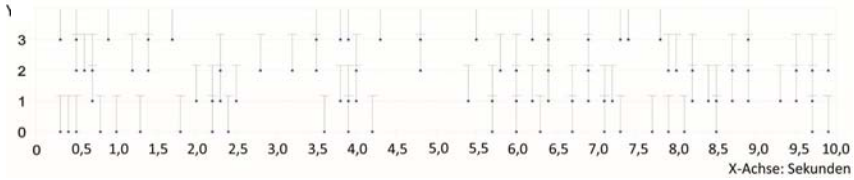
Auch bei der Analyse der Regelmäßigkeit der Grundfrequenzmaxima hat die vierte Beispielgruppe eine exponierte Position. Auf den ersten Blick erscheint die Anzahl der Grundfrequenzmaxima in allen Abschnitten bei allen Beispielen annähernd gleich. Bei näherem Hinsehen ist festzustellen, dass das Lyrik-Beispiel die wenigsten, das Slam Poetry Beispiel mehr und das Essay-Beispiel die meisten Grundfrequenzmaxima innehat.

Betrachtet man die Ergebnisse der empirischen Erhebung, ist zu sagen, dass die Probanden möglicherweise nach der Häufigkeit der Grundfrequenzmaxima ihre Zuordnung vollzogen haben. So wurde das Gedicht-Beispiel am häufigsten für Slam Poetry gehalten, das Essay-Beispiel am häufigsten für das Gedicht und das Slam Poetry Beispiel am häufigsten für das Essay. Dies kann daran liegen, dass die Vergleichsbeispiele Essay und Gedicht beide über ungewöhnlich viele Grundfrequenzmaxima verfügen und eine klare Zuordnung dadurch erschwert wurde. Wahrscheinlicher ist aber, dass Anke Fuchs ihren ernsten Slam Poetry Text mit dem Titel „Was wisst ihr denn schon davon“ eher ruhig und wenig-akzentuiert vorgetragen hat, sodass der Eindruck von Schwermütigkeit und Ernst entsteht, welcher eher bei einem Essay zu erwarten ist, das meist ernste oder wissenschaftliche Themen behandelt. Daher wurden die Beispiele mit mehr Grundfrequenzmaxima als Slam Poetry identifiziert. An dieser Beispielgruppe ist abzulesen, welche Erwartungen seitens der Probanden in Bezug auf die Realisierung eines Poetry Slam-Vortrags bestehen. Der Text von Anke Fuchs ist anders strukturiert und verwendet entgegen des ursprünglich kämpferischen Wesens von Slam Poetry, wie es von Ong und Smith proklamiert wurde, andere Stilmittel zur Akzentuierung bzw. Vermittlung der Inhalte.

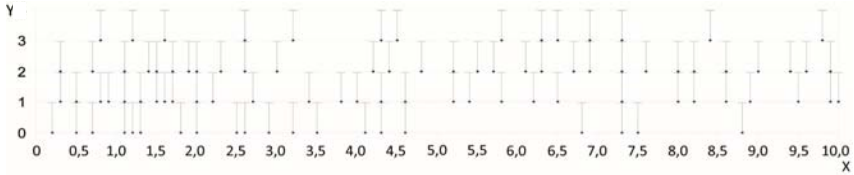
Dieser Eindruck erhärtet sich bei der Betrachtung der Grundfrequenzmaxima aller Slam Poetry Beispiele, welche in den Diagrammen in Abbildung 12 in folgender Reihenfolge dargestellt sind: y-Achse: 0) Andy Strauß, 1) Wolfgang Lüchtrath, 2) Pauline Füg, 3) Anke Fuchs.

Abschließend zur Analyse der Regelmäßigkeit der Grundfrequenzmaxima kann festgehalten werden, dass die meisten Slam-Beispiele die im Theorieteil festgehaltenen Hypothesen bestätigen. Slam Poetry wird stark durch Grundfrequenzmaxima rhythmisiert und es kommen nur wenige unbetonte Phasen vor. Der Vielfältigkeit von Slam Poetry ist zuzuschreiben, dass dies nicht immer zutreffen muss und ernste Texte, wie zum Beispiel der von Anke Fuchs, schwerer zu erkennen sind als die den typischen Merkmalen entsprechenden Präsentationen. Hier wird erneut deutlich, dass eine Aufteilung der Beispiele in unterschiedliche Gruppen zweckdienlich sein könnte, um solche „Ausreißer“ zu vermeiden und eindeutige Ergebnisse zu generieren. Möglicherweise kann eine Analyse der Grundfrequenzmaxima in Hinblick auf die tatsächlich-realisierten Worte eine Zuordnung vereinfachen. Dies war in dieser Studie nicht möglich, weil mit technisch-delexikalisierten Hörproben gearbeitet wurde.

Anfang



Mitte



Ende

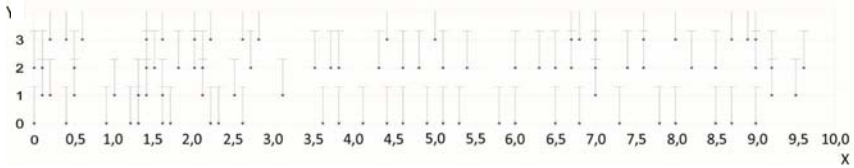


Abbildung 12. Regelmäßigkeit der Grundfrequenzmaxima aller Slam-Beispiele.

4.2.3 Variabilität der Intensität

	Zeitpunkt der Messung →	Anfang			Mitte			Ende			
	Wert →	Min. Int.	Max. Inte.	Variabilität	Min.	Max.	Var.	Min.	Max.	Var.	Ø Var.
1	Slam: Andy Strauß	24,27 dB	33,69 dB	9,42 dB	24,62 dB	32,80 dB	8,18 dB	24,27 dB	32,44 dB	8,17 dB	8,59 dB
2	Lyrik: Peter Reimers	23,20 dB	28,00 dB	4,80 dB	23,87 dB	29,45 dB	5,58 dB	21,78 dB	29,24 dB	7,46 dB	5,94 dB
3	Essay: Harald Eggebrecht	24,45 dB	29,60 dB	5,15 dB	22,67 dB	29,24 dB	6,57 dB	22,67 dB	27,64 dB	4,97 dB	5,56 dB

Tabelle 7. Variabilität der Intensität Beispielgruppe 1.

Bei der Betrachtung der Tabelle zur Variabilität der Intensität, also der Lautstärkevariation, fällt auf, dass Andy Strauß im Durchschnitt ein wesentlich größeres Spektrum anwendet als Peter Reimers oder Harald Eggebrecht. Dies kann ein Hinweis auf die starke Verzerrung in Strauß Stimme sein, aber auch auf die im Slam als Ausdruckswerkzeug verwendete Lautstärke.

	Zeitpunkt der Messung →	Anfang			Mitte			Ende			
	Wert →	Min. Int.	Max. Int.	Variabilität	Min.	Max.	Var.	Min.	Max.	Var.	Ø Var.
1	Slam: Wolfgang Lüchtrath	22,14 dB	32,44 dB	10,30 dB	22,49 dB	34,40 dB	11,91 dB	22,31 dB	31,91 dB	9,60 dB	10,60 dB
2	Lyrik: Jürgen von der Lippe	21,25 dB	32,80 dB	11,55 dB	19,65 dB	27,64 dB	7,99 dB	20,54 dB	24,45 dB	3,91 dB	7,81 dB
3	Essay: Reinhard Kahl	22,67 dB	34,57 dB	11,90 dB	23,20 dB	29,78 dB	6,58 dB	22,85 dB	30,49 dB	7,64 dB	8,70 dB

Tabelle 8. Variabilität der Intensität Beispielgruppe 2.

Wolfgang Lüchtrath verfügt beim Vergleich der drei Beispiele ebenfalls über die größte durchschnittliche Variationsbreite der Lautstärke.

	Zeitpunkt der Messung →	Anfang			Mitte			Ende			
	Wert →	Min. Int.	Max. Int.	Variabilität	Min.	Max.	Var.	Min.	Max.	Var.	Ø Var.
1	Slam: Pauline Füg	22,67 dB	32,44 dB	9,77 dB	22,67 dB	32,26 dB	9,59 dB	22,85 dB	32,26 dB	9,41 dB	9,59 dB
2	Lyrik: Weibliche Sprecherin	21,96 dB	29,07 dB	7,11 dB	22,31 dB	32,09 dB	9,78 dB	21,25 dB	28,18 dB	6,93 dB	7,94 dB
3	Essay: Cora Stephan	19,65 dB	31,55 dB	11,90 dB	23,02 dB	32,09 dB	9,07 dB	22,49 dB	29,78 dB	7,29 dB	9,42 dB

Tabelle 9. Variabilität der Intensität Beispielgruppe 3.

Bei der Analyse der dritten Beispielgruppe ist festzustellen, dass Pauline Füg hier zwar die höchste Intensitätsvariation verwendet, diese sich aber nur gering von Cora Stephans durchschnittlichem Wert unterscheidet. Die vierte Beispielgruppe eröffnet ebenfalls Interpretationsspielraum.

	Zeitpunkt der Messung →	Anfang			Mitte			Ende			
	Wert →	Min. Int.	Max. Int.	Variabilität	Min.	Max.	Var.	Min.	Max.	Var.	Ø Var.
1	Slam: Anke Fuchs	23,38 dB	30,49 dB	7,11 dB	23,38 dB	32,26 dB	8,88 dB	23,56 dB	30,31 dB	6,75 dB	7,58 dB
2	Lyrik: Ute Beckert	23,91 dB	34,04 dB	10,13 dB	23,74 dB	32,44 dB	8,70 dB	23,91 dB	34,04 dB	10,13 dB	9,65 dB
3	Essay: Christiane Grefe	22,49 dB	31,91 dB	9,42 dB	22,85 dB	32,26 dB	9,41 dB	22,49 dB	34,74 dB	12,25 dB	10,36 dB

Tabelle 10. Variabilität der Intensität Beispielgruppe 4.

Anke Fuchs verwendet in dieser Beispielgruppe die geringste durchschnittliche Lautstärkevariation. Christiane Grefes Wert dagegen ist für ein Essay bei Betrachtung der Beispiele der männlichen Sprecher erstaunlich hoch. Dies kann bedeuten, dass Frauen generell über ein größeres Intensitätsspektrum verfügen als Männer - sowohl in der Lyrik als auch im Essay - und das jeweilige Slam-Beispiel daher nicht heraussticht. In diesem Fall wäre eine Differenzierung zwischen Merkmalen männlicher und weiblicher Sprecher im Poetry Slam sinnvoll. Vergleicht man allerdings alle Slam-Beispiele, so wird deutlich, dass die ersten drei Beispiele eine hohe Intensität aufweisen. Lediglich das Beispiel von Anke Fuchs stellt eine Besonderheit dar.

	Zeitpunkt der Messung →	Anfang			Mitte			Ende			
	Wert →	Min. Int.	Max. Int.	Variabilität	Min.	Max.	Var.	Min.	Max.	Var.	Ø Var.
1	Andy Strauß	24,27 dB	33,69 dB	9,42 dB	24,62 dB	32,80 dB	8,18 dB	24,27 dB	32,44 dB	8,17 dB	8,59 dB
2	W. Lüchtrath	22,14 dB	32,44 dB	10,30 dB	22,49 dB	34,40 dB	11,91 dB	22,31 dB	31,91 dB	9,60 dB	10,60 dB
3	Pauline Füg	22,67 dB	32,44 dB	9,77 dB	22,67 dB	32,26 dB	9,59 dB	22,85 dB	32,26 dB	9,41 dB	9,59 dB
4	Anke Fuchs	23,38 dB	30,49 dB	7,11 dB	23,38 dB	32,26 dB	8,88 dB	23,56 dB	30,31 dB	6,75 dB	7,58 dB

Tabelle 11. Variabilität der Intensität aller Slam Beispiele.

Es kann also festgehalten werden, dass der Parameter Lautstärke bzw. Lautstärkevariation im Poetry Slam eine große Rolle spielt. Inwiefern ein Unterschied zwischen männlichen und weiblichen Vertretern besteht, müsste intensiver geprüft werden.

4.2.4 Sprechgeschwindigkeit

In der ersten Beispielgruppe hat das Slam-Beispiel mit 522 die meisten Silben pro Minute. Danach folgt das Essay mit 402 Silben pro Minute. Das langsamste Hörbeispiel ist das Gedicht mit 348 Silben pro Minute. Nach Miosgas Einteilung (vgl. 2006:90) ist die Sprechgeschwindigkeit der ersten beiden Beispiele sehr rasch (>350 Silben/Minute) und die des Gedichts rasch (350 Silben/Minute).

Das höchste Sprechtempo in der zweiten Beispielgruppe weist das Lyrik-Beispiel mit 624 Silben pro Minute auf. Darauffolgend verfügt das Slam-Beispiel über 486 Silben pro Sekunde. Zuletzt ist das Essay mit 378 Silben pro Minute zu nennen. Sie sind nach Miosga alle als sehr rasch zu bezeichnen.

Die dritte Beispielgruppe zeigt ebenfalls eine andere Reihenfolge. Hier ist das Essay mit 450 Silben pro Minute am schnellsten gesprochen, an zweiter Stelle das Lyrik-Beispiel mit 384 Silben pro Minute und an letzter Stelle das Slam-Beispiel mit 366 Silben pro Minute. Mit 450 Silben pro Minute übersteigt das Essay hiermit die übliche Variationsbreite im Deutschen, welche nach Miosga zwischen 100 und 400 Silben pro Minute liegt. Laut Aussage von O. Niebuhr liegen plausible Werte allerdings zwischen

180 und 720 Silben pro Minute, sodass die Werte im üblichen bzw. möglichen Bereich liegen.

In der vierten Beispielgruppe dominiert das Slam-Beispiel mit 420 Silben pro Minute. Damit ist es schneller gesprochen als das Essay mit 396 Silben pro Minute und das Gedicht mit 258 Silben pro Minute. Das Gedicht ist nach Miosga daher als übermittel (300 Silben/Minute) zu kategorisieren.

Beim Vergleich aller Slam-Beispiele ergibt sich folgende Reihenfolge: Andy Strauß spricht mit 522 Silben pro Minute am schnellsten. Wolfgang Lüchtrath verwendet als zweiter 486 Silben pro Minute. Anke Fuchs positioniert sich mit 420 Silben pro Sekunde an dritter Stelle. Pauline Füg ist mit 366 Silben pro Minute letzte. Alle Beispiele sind allerdings als sehr rasch eingeteilt.

Da die Reihenfolge in den jeweiligen Beispielgruppen variiert, ist keine Tendenz hinsichtlich des besonders schnellen oder langsamen Wesens von Slam Poetry im Vergleich zu anderen mündlichen Genres nachweisbar. Deswegen ist das Sprechtempo wahrscheinlich nicht ausschlaggebend für die Wahl der Probanden gewesen. Möglicherweise ist eine hohe Sprechgeschwindigkeit und die damit einhergehenden Funktionen wie zum Beispiel dem Ausdruck von Freude oder Angst oder der Kennzeichnung von Nebeninformationen, ein Merkmal nicht natürlicher bzw. spontaner Sprache (vgl. Miosga 2006:196f.). Die Hypothese, dass Slam Poetry einen Sachverhalt schnell auf den Punkt bringt als Indiz für eine hohe Sprechgeschwindigkeit, war durch diese Messung nicht zu erkennen.

4.2.5 Pausen im Signal

Bei einer Analyse der Pausenanzahl und der Pausenlänge anhand von technisch-delexikalisierten Hörbeispielen ist es nur bedingt möglich, eine Unterscheidung zwischen gefüllten oder ungefüllten Pausen vorzunehmen. Leere Sprechpausen entstehen durch die physiologische Notwendigkeit des Atmens. Gefüllte Pausen erfüllen Funktionen wie zum Beispiel lexikalische Unterscheidungen „yellow stone vs. Yellowstone“ (Miosga 2006: 153) oder als Grenzsignal zwischen Sinnabschnitten.

Genre	Slam	Lyrik	Essay
Anzahl der Pausen	12	30	13
Pausenlänge insgesamt	2,67 sek	10,45 sek	3,97 sek
Ø Pausenlänge einzeln	0,22 sek	0,35 sek	0,31 sek

Tabelle 12. Pausen im Signal Beispielgruppe 1.

Die erste Beispielgruppe beginnt mit dem Slam-Vortrag von Andy Strauß, der in dieser Gruppe zwar nicht die meisten, allerdings die im Durchschnitt kürzesten Pausen macht. Er spricht daher relativ hektisch, was laut Miosga Aufgeregtheit oder Engagement bedeuten kann (vgl. Miosga 2006:159). In Kombination mit dem hohen Sprechtempo kann dies bedeuten, dass der Sprecher nur wenig Zeit zur Atemergänzung hat und das Auditorium nur wenig Zeit, um die Informationen zu verarbeiten (vgl. Miosga

2006:160). Es ist festzustellen, dass das Slam-Beispiel in der ersten Beispielgruppe die wenigsten Pausen aufweist.

Genre	Slam	Lyrik	Essay
Anzahl der Pausen	20	32	34
Pausenlänge insgesamt	12,55 sek	15,12 sek	8,18 sek
Ø Pausenlänge einzeln	0,63 sek	0,47 sek	0,24 sek

Tabelle 13. Pausen im Signal Beispielgruppe 2.

Wolfgang Lüchtraths Slam-Text aus der zweiten Beispielgruppe weist allerdings andere Merkmale auf. Unter den aufgeführten Beispielen liegt sein Wert für die durchschnittliche Pausenlänge bei 0,63 Sekunden, was im Vergleich zu dem vorangegangenen Beispiel ausgedehnt ist. Anscheinend steht Wolfgang Lüchtrath während seines Vortrages weniger unter Zeitdruck als Andy Strauß oder er setzt Sprechpausen bewusst zur Herstellung von Spannung oder Markierung von Priorität ein. Mit einer Gesamtpausenlänge von 12,55 Sekunden liegt er in der Mitte zwischen dem Essay- und dem Gedichtbeispiel. Anhand der Länge der Sprechpausen ist es also nur bedingt möglich, Slam-Texte von anderen Genres zu unterscheiden. Auch in der zweiten Gruppe sind in dem Poetry Slam-Beispiel die wenigsten Pausen zu verzeichnen. Diese Beobachtung wir durch die folgenden Gruppen bestätigt.

Genre	Slam	Lyrik	Essay
Anzahl der Pausen	19	22	21
Pausenlänge insgesamt	4,79 sek	11,82 sek	7,44 sek
Ø Pausenlänge einzeln	0,25 sek	0,54 sek	0,35 sek

Tabelle 14. Pausen im Signal Beispielgruppe 3.

Genre	Slam	Lyrik	Essay
Anzahl der Pausen	10	27	15
Pausenlänge insgesamt	4,35 sek	10,7 sek	4,4 sek
Ø Pausenlänge einzeln	0,44 sek	0,4 sek	0,29 sek

Tabelle 15. Pausen im Signal Beispielgruppe 4.

Pauline Füg macht in ihrem Slam Vortrag 19 Pausen, Anke Fuchs nur zehn. Die Anzahl der Pausen kann daher für die Probanden ein Anhaltspunkt für die Unterscheidung der Beispiele gewesen sein. Gerade in Hinblick darauf, dass das Lyrik Beispiel stets die längste Pausenlänge aufweist und im Vergleich viele Pausen gemacht werden. Dies wird in der Tabelle 16 sichtbar. Es ist zu erkennen, dass die Slam-Interpreten insgesamt eher kürzere Pausen machen und sich in der Summe weniger Zeit für Pausen nehmen als die Lyriker. Das kann auf das Veranstaltungsformat hinweisen: Die Slam-Poeten

haben nur begrenzt Zeit, ihre Inhalte zu vermitteln und müssen sich daher unter Umständen beeilen, um ihren Text in der Kürze der Zeit zu performen. Auch soll das Publikum mitgerissen werden. Daher könnten diese Eigenschaften charakteristisch für Slam-Texte sein. Bei den Gedichten besteht kein Zeitdruck und der Hörer bekommt die Möglichkeit, die Informationen in Ruhe zu verarbeiten.

Slam-Interpret	Andy Strauß	W. Lüchtrath	Pauline Füg	Anke Fuchs
Anzahl d. Pausen	12	20	19	10
Pausenlänge insg.	2,67 sek	12,55 sek	4,79 sek	4,35 sek
Ø Pausenlänge einz.	0,22 sek	0,63 sek	0,25 sek	0,44 sek
Lyrik Interpret	P. Reimers	J.v.d. Lippe	Weibl. Sprecherin	Ute Beckert
Anzahl d. Pausen	30	32	22	27
Pausenlänge insg.	10,45 sek	15,12 sek	11,82 sek	10,7 sek
Ø Pausenlänge einz.	0,35 sek	0,47 sek	0,54 sek	0,4 sek

Tabelle 16. Pausen im Signal aller Slam- und Gedichtbeispiele.

Die Werte für Essay und Slam weisen dagegen häufiger Ähnlichkeiten auf. Es ist anhand der Anzahl der Pausen und der Pausenlänge schwer zu unterscheiden, ob es sich um ein Essay oder einen Slam-Text handelt, wie der Tabelle 17 zu entnehmen ist. Dies könnte darauf zurückzuführen sein, dass in einem Essay ähnlich viel Inhalt in kurzer Zeit zu vermitteln ist wie in einem Poetry Slam-Vortrag.

Slam-Interpret	Andy Strauß	W. Lüchtrath	Pauline Füg	Anke Fuchs
Anzahl d. Pausen	12	20	19	10
Pausenlänge insg.	2,67 sek	12,55 sek	4,79 sek	4,35 sek
Ø Pausenlänge einz.	0,22 sek	0,63 sek	0,25 sek	0,44 sek
Essay-Sprecher	H. Eggebrecht	Reinhard Kahl	Cora Stephan	Christiane Grefe
Anzahl d. Pausen	13	34	21	15
Pausenlänge insg.	3,97 sek	8,18 sek	7,44 sek	4,4 sek
Ø Pausenlänge einz.	0,31 sek	0,24 sek	0,35 sek	0,29 sek

Tabelle 17. Pausen im Signal aller Slam- und Essaybeispiele.

Vergleicht man die Poetry Slam-Beispiele unter sich, ist zu sagen, dass sie sehr unterschiedliche Merkmale aufweisen. Die Pausenanzahl variiert zwischen zehn und zwanzig Pausen und auch die Pausenlänge ist teilweise unterschiedlich. Klare Unterscheidungskriterien zeigen sich folglich nur im direkten Vergleich zu dem Lyrik-Beispiel.

5. Zusammenfassung

Diese Pilotstudie hat einige Ergebnisse und Hypothesen bezüglich der prosodischen Eigenschaften von Slam Poetry zutage getragen, die im Folgenden zusammenfassend dargestellt werden.

Eine Umfrage hat ergeben, dass Slam Poetry erkennbar bzw. von anderen mündlichen Genres zu unterscheiden ist. Die Prosodie von Slam Poetry unterscheidet sich also von anderen mündlichen Genres. Anscheinend bestanden bei den Teilnehmern der Umfrage bereits bestimmte Vorannahmen zu den prosodischen Eigenschaften von Slam Poetry. So wurden Ausnahmen zwar wahrgenommen, jedoch wurde nach dem vorangegangenen Muster gewählt. Daher stellt sich die Frage, inwiefern sich Slam Poetry von anderen mündlichen Genres wie den in dieser Studie verwendeten vorgelesenen Essays und Gedichten unterscheidet.

Um dies zu ermitteln, wurden die in der Umfrage benutzten delexikalisierten Hörbeispiele prosodisch mit dem Programm PRAAT untersucht. Durch diese Analyse konnten einige Merkmale festgestellt werden, die Slam Poetry eigen sind und möglicherweise als charakteristisch gelten können. Gestützt werden die Hypothesen mit Ergebnissen der theoriebasierten Exploration.

Bei Slam Poetry Vorträgen wird eine erhöhte Stimmlage verwendet. Dies ist wahrscheinlich durch den erhöhten Aktivierungsgrad der Slam-Poeten auch wegen des öffentlichen Veranstaltungsformats zu begründen. Es wird an mehreren Stellen erwähnt, dass in Poetry Slam-Texten ein kämpferischer Ton vorherrscht.

Slam Poetry enthält vermehrt Grundfrequenzmaxima, welche in regelmäßiger Abfolge vorkommen. Daher kann sie als stark rhythmisch bezeichnet werden. Die Tonhöhe wird im Poetry Slam als Werkzeug zur Akzentuierung und Rhythmisierung verwendet. In der Literatur ist dieses Phänomen nicht nur als Merkmal von Mündlichkeit nach Ong zu verzeichnen, sondern beruht auch auf der musikalisch-kulturellen Vorgeschichte des Formats. Unter Umständen ist dies auch ein Indiz für die Verwendung von Komik durch übermäßige Akzentuierung. Es ist feststellbar, dass Frauen ihre Tonhöhe häufiger als Ausdrucksmittel verwenden als Männer, daher ist eventuell eine Differenzierung zwischen männlichen und weiblichen Sprechern interessant.

Es konnte nicht festgestellt werden, dass Slam Poetry über eine im Vergleich zu den anderen Genres besonders hohe Sprechgeschwindigkeit verfügt. Mit durchschnittlich über 350 Silben pro Minute schneller Sprache ist die Sprechgeschwindigkeit bei Slam Poetry allerdings als sehr rasch zu bezeichnen.

Slam Poetry wird unter anderem durch die Lautstärke gestaltet. In der Analyse belegt und in der Literatur angedeutet, verwenden Slam-Poeten eine breite Intensitätsvariation, was auf einen virtuosen Umgang mit Sprache hindeuten kann.

Im Vergleich zu der Pausenanzahl und Pausenlänge der Essaybeispiele weisen Poetry Slam-Texte kaum deutliche Abstufungen auf. Gedichte scheinen generell mehr und längere Pausen zu haben als Essays oder Slam-Texte und können dadurch gut von ihnen unterschieden werden. Dies kann ein Hinweis auf das Veranstaltungsformat sein, weil Poeten bei einem Poetry Slam nur wenig Zeit für ihre Performance haben. Die Poetry Slam-Beispiele selbst unterscheiden sich in Pausenanzahl und Pausenlänge stark.

6. Fazit

Obwohl bei dieser Studie einige Ausprägungen prosodischer Parameter festgestellt werden konnten, die in Literatur und Analyse übereinstimmend sind und auch durch die Umfrage im Groben bestätigt wurden, sind eindeutige Aussagen über die prosodischen Eigenschaften von Slam Poetry schwierig festzumachen. Ausnahmen wie zum Beispiel der Vortrag von Anke Fuchs, der auch den Probanden der empirischen Studie Probleme bereitete, machen deutlich, dass es Slam-Poeten möglich ist, von der Erwartung abweichende Gestaltungsmittel zu verwenden. Es stellt sich die Frage, ob es sich bei diesen Ausnahmen um Einzelphänomene handelt oder um eigene Sub-Genres von Slam Poetry. Mittels eines umfangreicheren Korpus könnte festgestellt werden, ob diese Ausnahmefälle eventuell nur in dieser Studie in der Unterzahl sind.

7. Ausblick

Die mit Hilfe dieser Studie generierten Hypothesen über die prosodische Gestaltung von Slam Poetry sind erste Schritte in ein weites Forschungsfeld und lassen viele Forschungsfragen offen. So bestehen möglicherweise prosodische Unterschiede zwischen Vorträgen männlicher und weiblicher Poeten, sowie zwischen Slam-Texten mit ernstem oder komischem Inhalt. Die Delexikalisierung der Hörbeispiele hatte zur Folge, dass einige Parameter wie zum Beispiel die Lautdauer oder die Stimmqualität nicht für die empirische Studie verwendet werden konnten. Eine Untersuchung der unbearbeiteten Hörbeispiele erschien hier nicht sinnvoll, weil diese auch den Probanden nicht zur Verfügung standen. Interessant wären auch Interviews mit den Poeten selbst zur prosodischen Gestaltung bzw. ihrer Intention dahinter. Dies sind Aspekte des Forschungsbereichs, die im Rahmen einer Master Arbeit auch aufgrund des begrenzten Umfangs nicht zu erforschen sind und weiterer Studien bedürfen.

Abschließend ist zu sagen, dass Poetry Slam nicht nur ein neues und revolutionäres Veranstaltungsformat ist, sondern auch Poesie erzeugt, die auf eine spannende Art und Weise vorgetragen wird. Aktuelle Entwicklungen wie zum Beispiel die sogenannten Poetry Clips machen neugierig auf die moderne Dichtkunst, die nicht nur auf junge Rezipienten Einfluss nehmen kann.

8. Diskussion

Im Verlauf dieser wissenschaftlichen Arbeit haben sich viele methodische Entscheidungen bewährt, es gab aber auch einiges, was im Fall einer Folgestudie zum Beispiel in Hinblick auf den Einsatz der Methoden verbesserungswürdig ist. Die theoriebasierte Exploration ging aufgrund der passend-gewählten Literaturquellen relativ unkompliziert vonstatten. Die Ansätze gaben Anregungen und waren für die Hypothesengenerierung bezüglich der prosodischen Eigenschaften von Slam Poetry fruchtbar.

Die Auswahl der Hörbeispiele für die Umfrage wurde weitgehend willkürlich vorgenommen, und es hätten durchaus andere Vergleichsbeispiele gewählt werden

können. Die hier verwendeten erschienen der Autorin besonders einschlägig in Hinblick auf die prosodischen Eigenschaften. Es ist allerdings zu bedenken, dass hier nicht nur eine unterschiedliche Form vorliegt - Slam, Lyrik oder Essay - sondern auch unterschiedliche Vortragssituationen und Zielgruppen gegeben sind. Das Essay wurde zum Beispiel in einem Studio aufgenommen und richtet sich an die Hörer des Radiosenders NDR Kultur. Slam Poetry ist dagegen ein Vortrag vor Publikum in einem Club, in dem sich vornehmlich junge Menschen aufhalten. Die Kriterien zur Auswahl der Hörbeispiele waren in dieser Studie geeignet, können aber noch durch Eliminierung der genannten überflüssigen Variablen optimiert werden, sodass sie nicht Teil der Untersuchung werden. Zu den Hörbeispielen ist außerdem zu sagen, dass sie in ihrer Qualität unterschiedlich waren. Optimal wäre eine standardisierte Qualität, sodass eine stärkere Vergleichbarkeit möglich wäre.

Die Umfrage zur Master Arbeit verlief schnell und reibungslos. Dank des Einsatzes von Internetmedien wie Youtube oder Facebook konnten in kurzer Zeit viele Probanden gewonnen werden. In einer weiteren Studie wäre es sinnvoll, die Einsicht der Probanden in die Ergebnisse der Vorgänger zu unterbinden, sodass eine Beeinflussung ausgeschlossen wird. Dazu ist es nötig, einen größeren Stichprobenumfang zu gewährleisten und die Probanden genau zu instruieren, welche Gruppen auszuwählen sind. Das Indiz für eine solche Anpassung, die Verringerung der Antwortstreuung, konnte nicht festgestellt werden und somit sind die Ergebnisse für die Studie brauchbar. Die Anpassung birgt allerdings eine gewisse Gefahr, welche zu verhindern wäre. Die Auswertung der Studie verlief dank der Hilfe von Herrn Förster problemlos und die Leitfragen konnten beantwortet werden. Ebenso war die Unterstützung von Herrn Prof. Dr. Niebuhr bei der Auswahl der Parameter und den nötigen Analyseschritten sehr hilfreich.

Die Analyse selbst stellte sich als relativ aufwendig und kleinschrittig dar. Den Vorstellungen zu Genauigkeit und Umfang der Analyse seitens der Autorin, die vor Beginn der vertiefenden Beschäftigung bestanden, konnte sie nur bedingt gerecht werden. Es wäre möglicherweise sinnvoll, für eine weitere Beschäftigung mit Prosodie ein Programm zu verwendet, das keine oder weniger Fehler zulässt und übersichtlicher ist. Für die Auswertung der prosodischen Analyse sowie die in der theoriebasierten Exploration aufgestellten Hypothesen ließ sich das Werk von Miosga gut anwenden. Es ist vorstellbar, dass die Auswertung noch reichhaltiger wäre, wenn der Theorieteil umfangreicher ausgearbeitet wäre, was aber den Umfang dieser Master Arbeit stark ausgedehnt hätte.

Rückblickend ist festzuhalten, dass die Autorin auch aufgrund der vielen interessanten Ergebnisse diese Pilotstudie für sich als erfolgreich bewertet. Dank der Unterstützung von Experten der Universität Flensburg und der Universität Kiel konnte prosodisch bislang noch nicht erforschtes Material analysiert und an Probanden getestet werden.

9. Referenzen

ALTMANN, H. / A. BATLINER / W. OPPENRIEDER (Hrsg.) 1989. Zur Intonation von Modus und Fokus im Deutschen. Tübingen: Max Niemeyer Verlag.

ANDERS, P. 2010. Poetry Slam im Deutschunterricht. Aus einer für Jugendliche bedeutsamen kulturellen Praxis Inszenierungsmuster gewinnen, um das Schreiben, Lesen und Zuhören zu fördern. Hohengehren: Schneider Verlag.

ANDERS, Y. 2001. Merkmale der Melodisierung und des Sprechausdrucks ausgewählter Dichtungsinterpretationen im Urteil von Hörern. Frankfurt am Main: Peter Lang GmbH.

BOERSMA, P. 2001. Praat, a system for doing phonetics by computer. Glot International 5, 341-345.

BORTZ, J. / N. DÖRING. 2006. Forschungs-Methoden und Evaluation für Human- und Sozialwissenschaftler. [4. Auflage] Heidelberg: Springer Medizin Verlag.

DIEKMANN, A. 2005. Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. [13. Auflage] Hamburg: Rowohlt Verlag.

DUDENREDAKTION. 2009. Duden. Die deutsche Rechtschreibung. [24. Auflage] Mannheim: Bibliographisches Institut.

JAKOBSON, R./ R.L. WAUGH. 1986. Die Lautgestalt der Sprache. New York: de Gruyter.

LADD, D.R. 1996. Intonational phonology. New York: Cambridge University Press.

LUCKMANN, T. 1986. Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen. In: F. Neidhardt / M.R. Lepsius / J. Weiß (Hrsg.), Kultur und Gesellschaft (S. 191-211). Opladen: Westdeutscher Verlag.

KOCH, P./OESTERREICHER, W. 1994. Schriftlichkeit und Sprache. In: H. Günther / O. Ludwig (Hrsg.), Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung (S. 587-604). Berlin/New York: de Gruyter.

MIOSGA, C. 2006. Habitus der Prosodie. Die Bedeutung der Rekonstruktion von personalen Sprechstilen in pädagogischen Handlungskontexten. Frankfurt am Main: Peter Lang.

NEUBER, B. 2002. Prosodische Formen in Funktion. Frankfurt am Main: Peter Lang.

PRECKWITZ, B. 2002. Poetry Slam. Nachhut der Moderne: eine literarische Bewegung als Anti-Avantgarde. Norderstedt: Books on Demand.

PRECKWITZ, B. 2005. Spoken Word und Poetry Slam. Kleine Schriften zur Interaktionsästhetik. Wien: Passagen Verlag.

RABSAHL, S. 2011. Jung und wild. Sebastian23. In: Sprachnachrichten 50 05/2011.

SCHNEIDER, J. N. 2004. Ins Ohr geschrieben. Lyrik als akustische Kunst zwischen 1750 und 1800. Göttingen: Wallenstein Verlag.

SCHÖNHERR, B. 1997. Syntax- Prosodie- nonverbale Kommunikation. Empirische Untersuchungen zur Interaktion sprachlicher und parasprachlicher Ausdrucksmittel im Gespräch. Tübingen: Niemeyer.

SONNTAG, G. P. 1999. Evaluation von Prosodie. Aachen: Shaker Verlag.

STOCK, E. 1999. Deutsche Intonation. [4. Auflage] Leipzig: Langenscheidt.

STOCK, E. 1996. Deutsche Intonation. Klangbeispiele zur Akzentuierung im Wort und in der Wortgruppe, zur Gliederung und Rhythmisierung. [Audiocassette] Berlin und München: Langenscheidt.

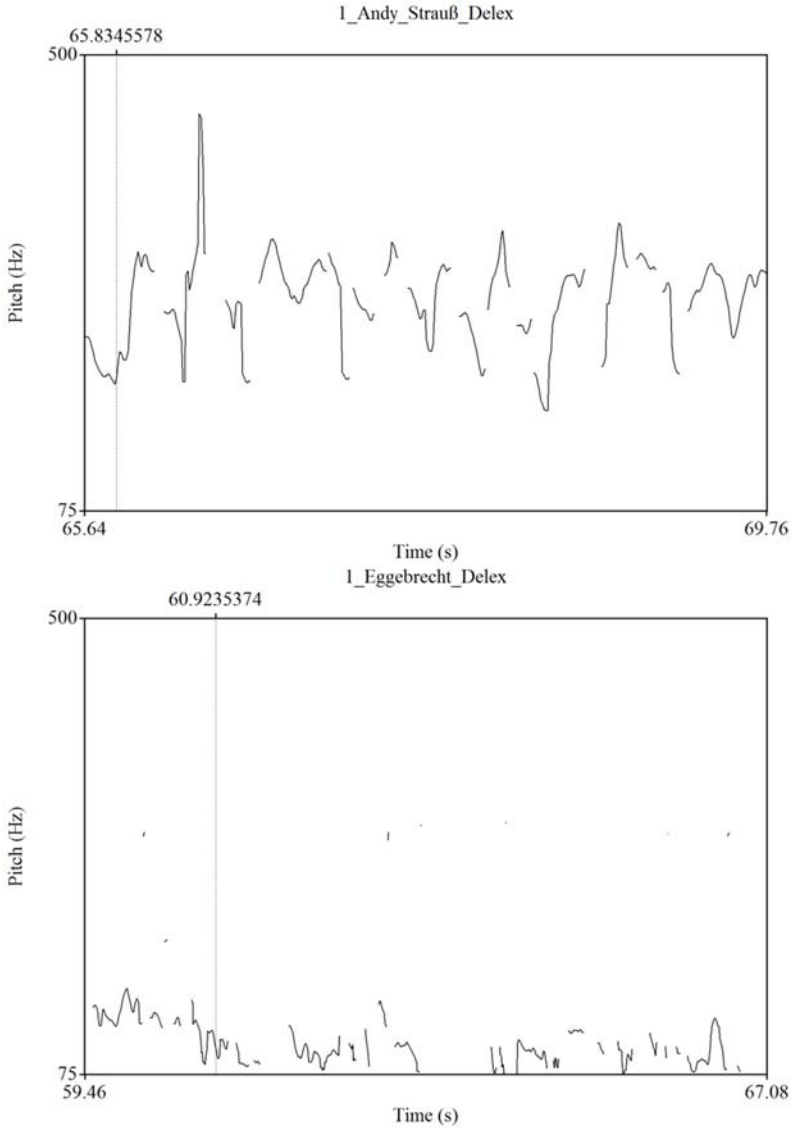
TRAVKINA, E. 2010. Sprechwissenschaftliche Untersuchungen zur Wirkung vorgelesener Prosa. Frankfurt am Main: Peter Lang.

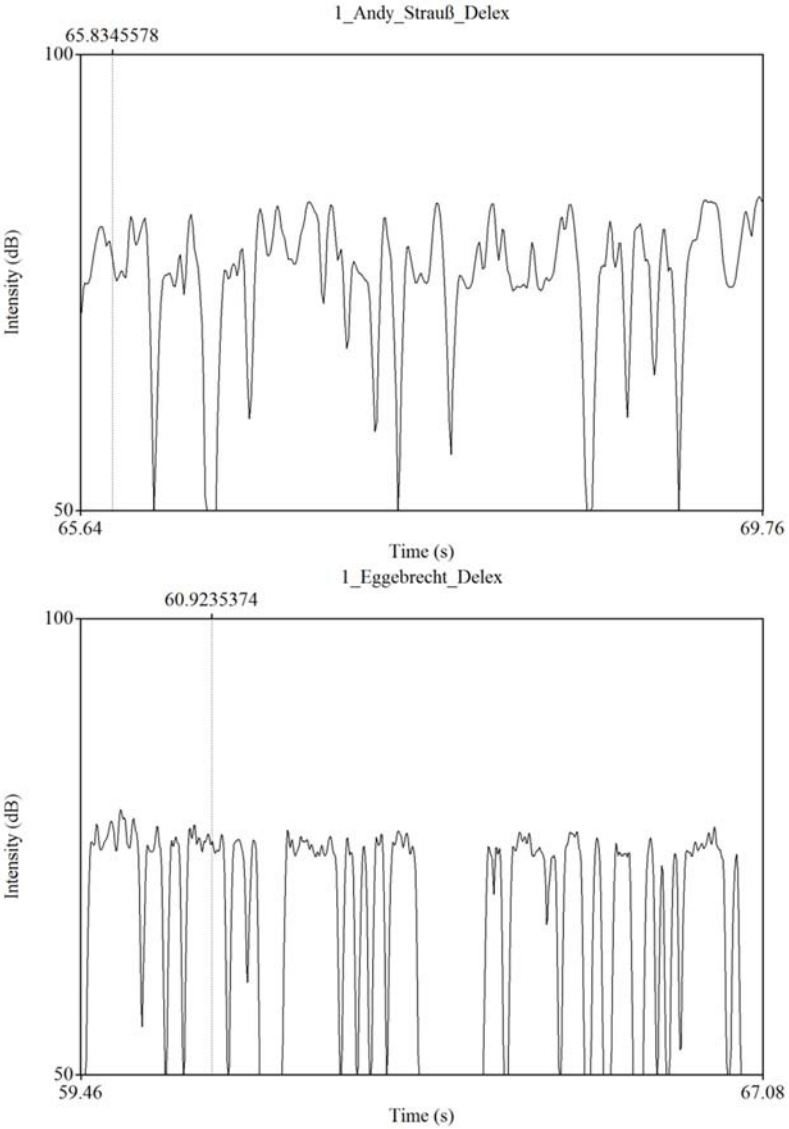
WESTERMAYR: 2010. Poetry Slam in Deutschland. Theorie und Praxis einer multimedialen Kunstform. Marburg: Tectum.

WILLRICH, A. 2010. Poetry Slam für Deutschland. Paderborn: Lektora.

10. Anhang

Analyseabbildungen aus PRAAT




1_Andy_Strauß_Delex
65.8345578
100
50
Intensity (dB)
65.64
69.76
Time (s)
1_Eggebrecht_Delex
60.9235374
100
50
Intensity (dB)
59.46
67.08
Time (s)

Identifizierung von Response-Adaptation (RA).

Response-Adaptation

RA bedeutet die Anpassung der individuellen Antworten in Befragungen. Unsere Situation ist eine Befragung mit einem dichotomen Item. N Individuen werden nacheinander befragt, wobei jeder Befragte die Antworten aller vor ihm Befragten als Liste sieht. RA ist dann eine Verzerrung, die dadurch zustande kommt, dass die Wahrscheinlichkeit für eine L-Antwort mit der empirischen Wahrscheinlichkeit (relativen Häufigkeit) dieser Antwort zunimmt. Je mehr Befragte also vorher eine L-Antwort gegeben haben, desto größer wird – ceteris paribus – die Wahrscheinlichkeit einer L-Antwort.
RA stellt ein Problem für die Reliabilität dar, insofern die individuellen Ergebnisse bei einer anderen Befragungsreihenfolge nicht bzw. schlecht reproduziert werden könnten.

Identifizierung

Wir teilen die Datenreihe in zwei Hälften: die ersten n/2 Fälle einerseits (R1) und die letzten n/2 Fälle andererseits (R2). Bei Vorliegen von RA müsste die Streuung der Antworten in den R2-Daten kleiner sein als in R1, da eine Konzentration auf eine bestimmte Antwort stattgefunden hätte.
Allein genügt dieser Test allerdings noch nicht, denn die Antworten könnten sich ja in R2 z.B. um 0 Konzentrieren, obwohl die Wahrscheinlichkeit in R1 für L größer war.
Um abzusichern, dass eine Konzentration auf RA verweist, muss also festgestellt werden, dass die dichotomen Erwartungswerte (0 oder L) der Streuungen in R1 und R2 gleich sind. Die dichotomen Erwartungswerte ergeben sich aus der Richtung der Abweichung von 0.5, wenn diese Abweichung signifikant ist.

Schritt 1

H $p_{R1} \neq 0.5$

Falls die Hypothese verworfen wird, können wir davon ausgehen, dass RA nicht vorliegt. Die Annahme dahinter ist, dass eine Gleichverteilung der Antworten die individuelle Antwort nicht nach der einen oder anderen Richtung verzerrt. Demnach führt erst eine deutliche Abweichung zu einer Wahrnehmung und kann mithin das Antwortverhalten beeinflussen.

Schritt 2

H $\begin{cases} p_{R2} < p_{R1} & \text{, wenn } \hat{p}_{R1} < 0.5 \quad \text{(0-RA)} \\ p_{R2} > p_{R1} & \text{, wenn } \hat{p}_{R2} > 0.5 \quad \text{(L-RA)} \end{cases}$

Die Bestätigung der Hypothese verweist auf RA. Schritt 3 sichert den Befund.

Schritt 3

H $\sigma_{R1}^2 > \sigma_{R2}^2$

Da es für kleine Fallzahlen nicht gelingen kann, sowohl die Kondition (Schritt 1) als auch die RA-Identifikation aus Schritt 2 nachzuweisen, ist in solchen Situationen eine RA-Identifikation auch alternativ zu Schritt 2 nur mit Schritt 3 zulässig (vgl. RA-Identifikation für x4 in folgendem Beispiel).

Beispiele

Artifizielle Daten: 30 Fälle, 5 dichotome Items (x1, x2, x3, x4, x5).

ti	x1	x2	x3	x4	x5
1	1	0	0	1	1
2	1	0	1	1	1
3	1	0	0	1	1
4	1	0	1	1	1
5	1	0	0	1	1
6	1	0	1	1	1
7	1	0	0	1	1
8	1	0	1	1	1
9	1	0	0	1	1
10	1	0	1	1	1
11	1	0	0	1	0
12	1	0	1	0	0
13	1	0	0	0	0
14	1	0	1	0	0
15	1	0	0	0	0
16	1	1	1	1	0
17	1	1	0	1	0
18	1	1	1	1	0
19	1	1	0	1	0
20	1	1	1	1	0
21	1	1	0	1	0
22	1	1	1	1	0
23	1	1	0	1	0
24	1	1	1	1	0
25	1	1	0	1	0
26	1	1	1	1	0
27	1	1	0	1	0
28	1	1	1	1	0
29	1	1	0	1	0
30	1	1	1	0	1

Schritt	x1	x2	x3	x4	x5
1: Kondition	***	***	ns	+	ns
2: 0L-RA	ns	ns	ns	ns	ns
3: Streuung	ns	ns	ns	**	***
Befund	kein RA	kein RA	kein RA	L-RA	kein RA

Gruppe 1, Beispiel 1

	Slam	Gedicht	Essay	DN
Kondition	***	***	***	***
0L-RA	0 ns	- ns	0 ns	- ns
Streuung	-	ns	ns	ns
Befund	?	kein RA	kein RA	kein RA

Gruppe 1, Beispiel 2

	Slam	Gedicht	Essay	DN
Kondition	***	**	ns	***
0L-RA	- ns	0 ns	- ns	- ns
Streuung	ns	ns	ns	ns
Befund	kein RA	kein RA	kein RA	kein RA

Gruppe 1, Beispiel 3

	Slam	Gedicht	Essay	DN
Kondition	**	***	**	***
0L-RA	- ns	0 ns	- ns	- ns
Streuung	ns	-	ns	-
Befund	kein RA	?	kein RA	kein RA

Gruppe 2, Beispiel 1

	Slam	Gedicht	Essay	DN
Kondition	ns	ns	ns	**
0L-RA	- ns	0 ns	0 ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 2, Beispiel 2

	Slam	Gedicht	Essay	DN
Kondition	ns	ns	*	**
0L-RA	0 ns	- ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 2, Beispiel 3

	Slam	Gedicht	Essay	DN
Kondition	ns	*	ns	**
0L-RA	0 ns	- ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 3, Beispiel 1

	Slam	Gedicht	Essay	DN
Kondition	*	ns	ns	**
0L-RA	0 ns	0 ns	L ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 3, Beispiel 2

	Slam	Gedicht	Essay	DN
Kondition	**	ns	ns	**
0L-RA	- ns	L ns	0 ns	0 ns
Streuung	-	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 3, Beispiel 3

	Slam	Gedicht	Essay	DN
Kondition	*	**	**	**
0L-RA	- ns	0 ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 4, Beispiel 1

	Slam	Gedicht	Essay	DN
Kondition	ns	ns	**	***
0L-RA	- ns	0 ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 4, Beispiel 2

	Slam	Gedicht	Essay	DN
Kondition	*	**	ns	***
0L-RA	- ns	0 ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

Gruppe 4, Beispiel 3

	Slam	Gedicht	Essay	DN
Kondition	ns	ns	**	***
0L-RA	0 ns	L ns	- ns	0 ns
Streuung	ns	ns	ns	-
Befund	kein RA	kein RA	kein RA	?

3 Zusammenhänge zwischen Stimmbildung und Stimmwahrnehmung – physiologische, akustische und perzeptorische Analysen

Evelin Graupe, M.A.
Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
evelin_graupe@yahoo.de

Die experimentell-empirischere Arbeit sucht nach Einflussfaktoren auf den empfundenen Stimmwohlklang und führt hierzu signalphonetische und perzeptive Untersuchungen von SprecherInnen mit unterschiedlich ausgeprägtem Gesangstraining durch. Im Einzelnen werden die experimentellen Untersuchungen durch die Hypothesen geleitet, dass sich eine wohlklingende Stimme signalphonetisch durch bestimmte Charakteristika auszeichnet und dass diese vor allem bei trainierten bzw. professionellen SängerInnen zu finden sind. Darüber hinaus wird angenommen, dass trainierte SängerInnen als solche von Hörern über ihrer Stimme erkannt werden können und dass diese Erkennungsleistung unter anderem mit Wohlklang in Verbindung steht. Hierzu wurden Sprachdaten von insgesamt 23 Sprechern mittleren Alters, die je nach Gesangshintergrund in Gesangsunerfahrene, Laiensänger und Profisänger unterteilt wurden, hinsichtlich ausgewählter akustischer Parameter (F0, LTF2, HNR, Jitter, OQ) analysiert und in einem mehrteiligen Wahrnehmungsexperiment von 70 Hörern beurteilt.

1. Einleitung

> “Mehr als die Schönheit selbst
> bezaubert die liebliche Stimme;
> jene zieret den Leib;
> sie ist der Seele Gewalt.”
> (Johann Gottfried Herder, 18.Jh.)

> "Bei der Natur der Stimme fragt man nach der Quantität und der Qualität. Bei der Quantität ist die Sache einfacher. Denn in der Hauptsache ist die Stimme entweder stark oder schwach. […] Die Qualität ist manchfaltiger. Denn die Stimme ist entweder rein oder heiser, voll oder fein, weich oder rauh, beengt oder fließend, hart oder geschmeidig, hell oder dumpf."
> (QUINTILIANUS, Anleitung zur Beredsamkeit, 1.Jh. n.Chr.)

Wie aus den Zitaten ersichtlich wird, reicht die Beschäftigung mit der menschlichen Stimme unter ästhetischen Gesichtspunkten mindestens bis in das antike Rom zurück. Die Sprechstimme, die ein komplexes Phänomen aus prosodischen Mustern, artikulatorischen Eigenheiten und stimmqualitativen Eigenschaften darstellt und in unserer tagtäglichen Kommunikation auf vielen Ebenen eine wichtige Rolle spielt, ist längst Gegenstand nicht nur ästhetisch-philosophischer Abhandlungen, sondern auch psychologischer und phonetischer Forschung.

Wir bilden uns, ob wir wollen oder nicht, auch anhand der Stimme eines Menschen eine Meinung über seine Charaktereigenschaften und fällen Urteile über seine Intelligenz, Kompetenz oder sozialen Status (vgl. z.B. Scherer 1972, Berry 1992 und Miyake/Zuckerman 1993). Wir dekodieren Emotionen (vgl. z.B. Scherer/Banse/Wallbott/Goldbeck 1991 oder Banse/Scherer 1996) und Sprechereinstellungen wie Ironie oder Sarkasmus (vgl. Scharrer/Christmann/ Knoll 2011 oder Trede 2011). Die Entschlüsselung der zugrundeliegenden Mechanismen hat sich die akustische und perzeptive Phonetik zur Aufgabe gemacht. Nicht zuletzt in der forensischen Phonetik sind außerdem die Zusammenhänge zwischen Sprechstimme und physischen Eigenschaften eines Sprechers wie Alter, Gewicht, Gesundheitszustand, Intoxikation, Geschlecht oder sogar sexuelle Orientierung Thema wissenschaftlicher Studien (vgl. z.B. Ferrand 2002 und Moos 2010).

Die vorliegende Arbeit widmet sich der psychophonetischen Betrachtung der Sprechstimme unter ästhetischen Gesichtspunkten. Die Suche nach der „wohlklingenden" Stimme soll dabei auf Basis empirischer Forschungsmethoden mit Fokus auf die Stimmperzeption vollzogen werden. Die Motivation zur Durchführung der eingebetteten Produktions- und Perzeptionsstudie liegt in den folgenden Fragen begründet: Gibt es eine objektiv „wohlklingende" Stimme? Wenn ja, lässt sich Wohlklang messen? Ist eine wohlklingende Stimme angeboren oder erlernbar?

Der ersten Frage widmet sich das in der vorliegenden Arbeit beschriebene Perzeptionsexperiment. Die zweite Frage hingegen zielt darauf ab, einen möglichen Zusammenhang zwischen den innerhalb der Produktionsstudie zu ermittelnden, akustischen Stimmprofilen und deren perzeptorischer Bewertung zu belegen. Die dritte Frage nach der Erlernbarkeit ist naturgemäß nur durch eine Langzeitstudie, z.B. die Betrachtung von Sprechern vor und nach umfangreichen stimmbildnerischen Maßnahmen zu beantworten. Da dies im Rahmen einer Masterarbeit zeitlich nicht möglich ist, soll eine erste Annäherung über eine Querschnittstudie geschehen. Somit ist der dritten Frage die Zusammensetzung der Produktionsstichprobe aus Probanden mit unterschiedlichem Stimmbildungslevel gewidmet. Als Repräsentanten für Sprecher mit ausgebildeten Stimmen werden hier sowohl Laien- als auch Profisänger betrachtet. Dies setzt allerdings die Annahme voraus, dass eine sängerische Stimmbildung auch Einfluss auf die Sprechstimme hat.

Die vorliegende Arbeit wagt somit einen Einblick in das Spannungsfeld zwischen der physiologischen, akustischen und perzeptiven Seite der menschlichen Stimme. Sie wird sich dementsprechend im theoretischen Teil mit der Theorie der Stimmbildung innerhalb des Gesangsunterrichts (Abschnitt 4) sowie mit der Perzeption (Abschnitt 3) und Beschreibung von Stimmen und Stimmqualität (Abschnitt 2) beschäftigen. Weiterhin werden einige den oben aufgezeigten Fragen und den theoretischen Erkenntnissen entsprechende Hypothesen zu den Ergebnissen des experimentellen Teils der Arbeit aufgestellt (Abschnitt 5). Der Beschreibung der Methode der Sprachaufnahmen inklusive akustischer Analyse (Abschnitt 6 und 8) und des Perzeptionsexperiments (Abschnitt 7) folgt in der Ergebnisauswertung die Betrachtung der Zusammenhänge von Stimme in Produktion und Perzeption (Abschnitt 9). Der letzte Abschnitt der Arbeit beschäftigt sich mit der Interpretation der Ergebnisse und ihrer Bedeutung für die aufgestellten Hypothesen. Zudem bietet er Platz für Methodenkritik und einen Ausblick auf potentielle weiterführende Untersuchungen im Themengebiet Stimme, Stimmbildung und Perzeption.

2. Stimme, Stimmklang und Stimmqualität

2.1 Begriffsbestimmungen

Obwohl eine klare Definition von Stimme für ihre Untersuchung unerlässlich ist, stellt sich die Formulierung einer einzelnen, für alle Zwecke geeigneten Definition aufgrund der Vielfalt der Funktionen, denen die menschliche Stimme dient, als schwierig dar. Jeder hat wohl eine Vorstellung davon, was mit dem Begriff „Stimme“ gemeint ist. Eine eindeutige, kontextinsensitive Definition scheint jedoch unmöglich und letztlich muss der Forschungskontext die Begriffsdefinition bestimmen. Somit dienen die folgenden Abschnitte der Festlegung des begrifflichen Rahmens der Arbeit.

2.1.1 Enges und weites Verständnis von Stimme

Stimme lässt sich einerseits in einem sehr engen Verständnis des Begriffes als den Klang, der durch die Vibration der Stimmlippen erzeugt wird, definieren. Dieses Verständnis bedeutet jedoch für die phonetische Forschung, dass die Filtereffekte des Ansatzrohres bei der Messung von Stimme exkludiert werden müssen, wofür spezielle Messverfahren wie die Elektro- oder Photoglottographie zum Einsatz kommen. Dementsprechend wird zwischen Stimme (engl. *voice*) und „Sprachklang“ oder „Sprechweise“ (engl. *speech*) unterschieden. (Vgl. Kreiman/Vanlancker-Sidtis/Gerrat 2005: 343)

Ein sehr weites Verständnis von Stimme nutzt andererseits die beiden Begriffe fast synonym, sodass hier neben der reinen Phonation durch die Stimmlippen auch die physiologische Konfiguration des Ansatzrohres (also des Filters) sowie artikulatorische Details (auch stimmloser Sprachabschnitte) und prosodische Muster (Intonation, Intensitätsverläufe und temporale Eigenschaften wie Rhythmus und Sprechgeschwindigkeit) mit eingeschlossen werden.

Für die vorliegende Arbeit ist das enge Verständnis nicht geeignet, da sich der perzeptive Aspekt der Forschungsfrage nicht mit einem Konzept von Stimme vereinbaren lässt, welches eine auditive Wahrnehmung ausschließt oder zumindest sehr schwierig macht. Das weite Verständnis hingegen ist für die Forschungsfragen geeignet, ihm kann jedoch aufgrund seiner hohen Parameterkomplexität nicht gänzlich entsprochen werden, da dies die gleichzeitige Beobachtung unzähliger prosodischer, stimmqualitativer und artikulatorischer Parameter erfordern würde. Zur Einschränkung der Parameterkomplexität, muss das Stichwort Habitus in der Begriffsbestimmung mit berücksichtigt werden. Somit soll sich das hier verwendete Verständnis von Stimme auf globale, d.h. regelmäßig wiederkehrende und/oder nahezu ständig vorhandene physiologische Konfigurationen des Vokaltrakts und die daraus resultierenden akustischen Charakteristika beziehen. Dem hier verwendeten Verständnis von Stimme kommt somit die nachstehende Definition[1] von Abercrombie am nächsten: *„those characteristics which are present more or less all the time that a person is talking: it is a quasi-permanent quality running through all the sound that issues from his mouth."* (Abercrombie 1967: 91)

Die methodische Umsetzung der Isolation der relevanten Eigenschaften zur Analyse der Stimme im Gegensatz zum weiter gefassten Phänomenbereich Sprechweise oder Sprachklang stellt jedoch sowohl bei der akustischen Analyse als auch im Perzeptionsexperiment eine Herausforderung dar.

2.1.2 Stimmklang, Stimmqualität und Sprachklang

Abgrenzend zum physiologisch orientiert definierten Begriff Stimme bietet es sich an, den Begriff „Stimmklang" für das perzeptorisch orientierte Konzept zu nutzen. Dieses baut auf Abercrombies Definition auf und soll ergänzt werden um eine hörerseitige *"cumulative abstraction over a period of time of a speaker-characterizing quality […]"* (Laver 1980: 1).

Der Begriff „Stimmqualität" hingegen soll hier aus rein phonetischer Perspektive betrachtet werden. Er umfasst dabei nicht nur Phonationstypen (siehe Abschnitt 2.3), sondern auch die durch Larynxposition und Pharynxkonfiguration bedingten Parameter. (Vgl. Catford 1988: 49 f.)

Es werden verschiedene stimmqualitative Kategorien unterschieden, deren Charakterisierung und Bezeichnung allerdings wiederum auf der Perzeption von Stimme beruhen (vgl. z.B. den Begriff *harsh voice*). Auf physiologischer Seite sind diese Kategorien durch unterschiedliche Konfigurationen der Muskeln des Larynx (vor allem Höhe des Larynx und adduktive/abduktive Spannung der Stimmlippen) und der Atemmuskulatur konstituiert und können mit physiologischen und akustischen Korrelaten beschrieben werden. Die Abgrenzung zu den prosodischen Faktoren wie Tonhöhe und Intensität fällt jedoch schwer, da auch diese durch die Einstellung der Larynxmuskeln und Veränderungen des Luftstroms bedingt sind.

[1] Die Beschreibung bezieht Abercrombie eigentlich auf den englischen Begriff *voice quality*. Die dt. Entsprechung *Stimmqualität* wird hier jedoch anders verstanden.

Analog zu Abercrombie (1967: 89) werden in der vorliegenden Arbeit grundsätzlich drei Komponenten von Sprachklang[2] unterschieden:

(a) artikulatorische Eigenschaften (bei Abercrombie *„segmental features“)*
(b) prosodische Eigenschaften (bei Abercrombie *„features of voice dynamics“*)
(c) stimmqualititative Eigenschaften (bei Abercrombie *„features of voice quality“*)

Bei dem perzeptorischen Konzept des Stimmklangs handelt es sich dementsprechend immer noch um einen Komplex aus den drei Eigenschaftsbereichen. Im Gegensatz zum hier weiter gefassten Begriff Sprachklang geht es dabei aber, wie oben beschrieben, um nichttemporäre, global auffindbare und damit um sprecheridentitätsstiftende Eigenschaften.

Zur Entwirrung der Begriffsvielfalt wurde in Abbildung 1 nochmals ein Schema der Begriffsverhältnisse dargestellt. Dabei nimmt die Größe des eingeschlossenen Phänomenbereichs bzw. des sich daraus ergebenden Parametersets in einer entsprechenden Untersuchung von oben nach unten ab.

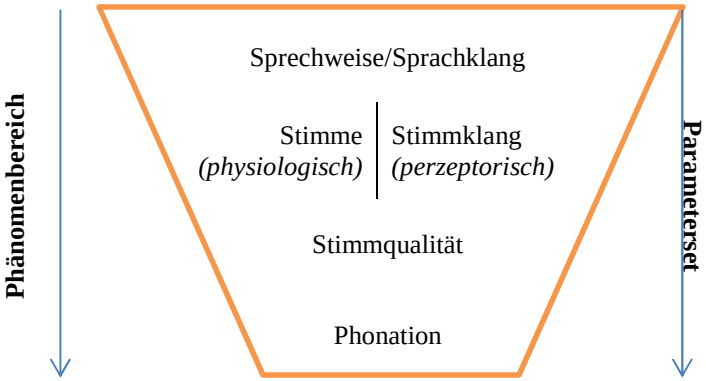


Abbildung 1. Darstellung zu den in der Arbeit zugrunde gelegten Begriffsverhältnissen

2.2 Parametrisierung von Stimmklang

Gemäß der Quelle-Filter-Theorie der Sprachproduktion (vgl. Fant 1960) wird Stimmklang durch zwei Komponenten bestimmt: die Quelle des Sprachschalls (in der vorliegenden Arbeit vornehmlich die vibrierenden Stimmlippen) und das Filter (Ansatzrohr/Vokaltrakt). Beide Komponenten können durch eine Reihe akustisch messbarer Parameter beschrieben werden, die einen Einfluss auf den Stimmklang haben.

Abhängig von der Quelle sind dabei[3]:

[2] Als analoger Begriff zu Sprechweise oder engl. *speech.*

1. Durchschnittliches F0 (zur Beschreibung der habituellen Sprechlage)
2. F0-Verlaufsmuster (inklusive F0-Standardabweichung zur Beschreibung der Tonhöhenvariabilität eines Sprechers und F0-Umfang)
3. Maße für Mikropertubationen des Quellsignals (*Jitter* und *Shimmer* zur Beschreibung der zeitlichen und energetischen Regelmäßigkeit der Stimmlippenschwingungen)
4. Öffnungsphase der Stimmlippen (*Open Quotient* bzw. *Closed Quotient* zur Beschreibung der Form der Stimmlippenschwingung)
5. Intensitätsverlaufsmuster und durchschnittliche Intensität (zur Beschreibung der habituellen Sprechlautstärke)
6. Phonationsinitiation (zur Beschreibung des Stimmansatzes)

In Abhängigkeit vom Filter stehen:

1. Absolute Werte der Formantfrequenzen und Formantenbandbreite (zur Beschreibung der habituellen Artikulation von beispielsweise Vokalen)
2. Durchschnittliches Langzeitspektrum (zur Beschreibung der physiologisch gegebenen und habituellen Resonanzeigenschaften des Ansatzrohres)

Abhängig von der Kombination aus Quelle und Filter ist nicht zuletzt die Form der spektralen Hüllkurve und die spektrale Neigung.

Welche Parameter primär zum Perzept Stimmklang beitragen und wie dieses beschrieben werden kann, wird im wissenschaftlichen Diskurs untersucht (siehe Abschnitt 3). Der experimentelle Teil der vorliegenden Arbeit soll ebenso einen Beitrag zur Beantwortung dieser Fragen leisten. Dabei liegt, wie bereits erwähnt, der Fokus auf stimmqualitativen Eigenschaften von Stimme. Entsprechend gibt der folgende Abschnitt einen noch genaueren Einblick in die phonetische Beschreibung von Stimmqualität.

2.3 Phonetische Beschreibung von Stimmqualität

Abgeleitet von den physiologischen Gegebenheiten des Stimmapparates können verschiedene Stimmqualitätskategorien beschrieben werden. Laver (1980) unterscheidet hierzu grundlegend zwischen der Konfiguration der Stimmlippen während der Phonation (*phonatory settings*) und der supralaryngalen Konfiguration (*supralaryngal settings*). Erstere bestimmt den Phonationstyp und betrifft damit die Quelle des Sprachschalls, letztere bestimmt resonanzbezogene, globale Eigenschaften der Stimme und betrifft dementsprechend das Filter. Die Kombination beider Konfigurationen ist im Sprachschall als Stimmklang wahrnehmbar und (teilweise) akustisch messbar.

2.3.1 Physiologische Parameter

Phonation ist bekanntermaßen ein myoelastisch-aerodynamischer Prozess. Aus der Betrachtung der Physiologie des Kehlkopfes lassen sich drei Parameter muskulärer

[3] Die Liste der Parameter erhebt keinen Anspruch auf Vollständigkeit.

Spannung, die mit den aerodynamischem Faktoren interagieren, aufzeigen (vgl. Laver 1980: 108 f.):

(a) adduktive Spannung, die über die Stellung der Arytenoiden (Stellknorpel) geregelt wird und für das Zusammenpressen (Adduktion) oder Auseinanderhalten (Abduktion) der Stimmlippen verantwortlich ist;
(b) longitudinale Spannung oder Längsspannung, die durch das Kippen des Thyroid (Schildknorpel) kontrolliert wird und für die Dehnung der Stimmlippen eingesetzt wird;
(c) mediale Kompression, die ähnlich der adduktiven Spannung für das Zusammenpressen der Stimmlippen verantwortlich ist, es jedoch ermöglicht, nur den vorderen Teil der Glottis (*ligamental glottis*) zu schließen, was zum Beispiel beim Flüstern der Fall ist.

Jeder Phonationstyp hat hinsichtlich der drei Spannungstypen unterschiedliche Konfigurationen, die für die spezifische Form der Stimmlippenschwingungen verantwortlich sind. Die Form der Stimmlippen-schwingung hat wiederum Einfluss auf die akustischen Merkmale des Quellsignals und ist im Stimmklang wahrnehmbar. Laver (1980: 109 f.) beschreibt sechs Phonationstypen und zusätzlich deren mögliche Kombinationen. Vier der sechs Phonationstypen sollen hier näher erläutert werden, da sie im weiteren Verlauf der Darstellung eine Rolle spielen:

(a) Modalstimme (*modal voice*), die physiologisch durch moderate adduktive Spannung und mediale Kompression und in der Indifferenzlage[4] ebenso durch moderate Längsspannung gekennzeichnet ist. Akustisch ist sie durch hohe Periodiziät und ohne wahrnehmbare glottale Friktion charakterisiert.
(b) Behauchte Stimme (*breathy voice*), die durch geringe muskuläre Spannung und damit durch unvollständige Schließung der Stimmlippen gekennzeichnet ist. Dies bedingt eine leichte glottale Friktion, die als Behauchung wahrgenommen wird.
(c) Knarrstimme (*creaky voice*), bei der eine hohe adduktive Spannung mit einer geringen Längsspannung kombiniert ist, was zur Verdickung der Stimmlippen und zu unregelmäßigen, tieffrequenten Stimmlippenschwingungen führt.
(d) Gespannte Stimme (*tense voice*), bei der generell eine hohe muskuläre Spannung in Larynx und auch Pharynx zu beobachten ist. Sie zeichnet sich durch eine hohe adduktive Spannung und mediale Kompression aus. Diese Stimmqualität geht auditiv mit dem Eindruck einer gepressten Stimme einher.

Von den supralaryngalen Konfigurationen soll an dieser Stelle nur eine vorgestellt werden, die in der vorliegenden Untersuchung von Bedeutung ist. Es handelt sich um

[4] Indifferenzlage bezeichnet denjenigen Tonhöhenbereich der Sprechstimme, in dem mit dem geringsten Kraftaufwand über einen langen Zeitraum mühelos gesprochen werden kann.

das Sprechen mit gesenktem Larynx (*lowered larynx*): Das muskulär kontrollierte Herabsenken des Kehlkopfes geht physiologsich mit der Verlängerung Vokaltraktes einher, was akustisch mit der Absenkung der unteren Formanten verbunden ist und perzeptiv zu einem Eindruck einer „dunkleren" Stimmfarbe beitragen kann. (Vgl. Fant 1960: 64)

2.3.2 Abgeleitete Parameter

Während für viele prosodische Merkmale der Stimme akustische Korrelate weitgehend etabliert sind (z.B. Stimmlage und Grundfrequenz, Stimmlautstärke und Energie im Spektrum), ist die Parametrisierung der Stimmqualität noch nicht hinreichend geklärt. Für die Darstellung, wie die Stimmlippen sich öffnen und schließen, werden verschiedene Methoden zugrunde gelegt. Eine Methode zur Quantifizierung der Stimmlippenfunktion ist das inverse Filtern, wodurch der glottale Luftstrom im Zustand vor der Beeinflussung durch das Filter des Vokaltrakts rekonstruiert wird (vgl. zB. Alku/Bäckström/Vilkman 2002). Hierbei wird auf das akustische Signal ein Filter mit einer Übertragungsfunktion invers zu der des Vokaltrakts angewendet. Das so entstehende komplexe Signal bildet den Schallfluss an der Glottis über der Zeit ab (*volume velocity waveform*, vgl. Klatt 1990: 825). Hierzu wird jedoch die genaue Bestimmung der Formanten in Frequenz und Bandbreite aus dem akustischen Signal benötigt.

Eine etwas direktere Methode ist die Elektroglottographie (EGG), bei der über zwei Elektroden, die in Höhe des Kehlkopfes angebracht werden, der elektrische Wechselstromwiderstand bzw. die Leitfähigkeit im Gewebe des Kehlkopfes gemessen wird. Ein größerer Kontakt der Stimmlippen in der Verschlussphase geht dabei mit höherer Leitfähigkeit einher. Wie Sherer/Druker/Titze (1988) zeigen konnten, verhält sie sich linear zur Kontaktfläche der Stimmlippen, wodurch das EGG-Signal als Funktion von Glottisöffnung über der Zeit interpretiert werden kann. An dieser ist wiederum der *Open Quotient* ablesbar, welcher als Kennwert für das Dauerverhältnis von Öffnungs- und Verschlussphase der Stimmlippen während einer Schwingung dient.

In beiden Verfahren erweist sich die exakte Lokalisierung der Anfangs- und Endzeitpunkte der gesuchten Phasen im Signal als problematisch. Das methodische Vorgehen zur Bestimmung des *Open Quotient* und verwandter Kennwerte ist daher umstritten, insbesondere da die Öffnung der Glottis graduell sein kann und dadurch der Beginn der offenen Phase im Signal teilweise schwer erkennbar ist. Dies bedeutet, dass Schwellenwerte festgelegt (z.B. die Amplitudenschwelle von 35% bei Rothenberg/Marshie (1988)) und experimentell im Vergleich mit anderen Verfahren (z.B. Photoglottographie) überprüft werden müssen. Die hier gewählte Methode der Abstandsmessung zwischen Minima und Maxima im differenzierten Signal beruht u.a. auf Baer/Löfqvist/McGarr (1983). Eine genauere Beschreibung erfolgt im Abschnitt 8.

Für die beschriebenen Phonationstypen finden sich in der Literatur typische *Open Quotient*-Werte von ca. 0,5 für modale Stimme, ca. 0,3 für gepresste Stimme oder auch Knarrstimme in Kombination mit Aperiodizität und ca. 0,65 für behauchte Stimme. Die exakten Wertebereiche sind jedoch vom Messmodel abhängig. (Vgl. Childers/Lee 1991: 2405)

Weitere stimmqualitative Parameter sind *Jitter* und *Shimmer*, die über die F0-Analyse bestimmt werden und dazu dienen, die Regelmäßigkeit der Stimmlippenschwingungen zu erfassen. Diese sogenannten Mikrovariationen in stimmhaften Ab-

schnitten werden bei Frequenzunregelmäßigkeiten als *Jitter*, bei Amplitudenunregelmäßigkeiten als *Shimmer* bezeichnet und können den wahrgenommenen Stimmklang beeinflussen. Bei der akustischen Analyse unterscheidet man eine Vielzahl von Maßen für *Jitter* und *Shimmer*. Häufig genutzte sind die *Relative Average Perturbation* (RAP) für *Jitter* und der *Average Perturbation Quotient* (APQ) für *Shimmer*. (Vgl. Boersma & Weenink 2012). Abschnitt 8 enthält genauere Angaben zur angewandten Messmethode in der vorliegenden Arbeit.

Durch Irregularitäten in der Stimmlippenschwingung oder wenn diese nicht richtig schließen, wie dies z.B. bei behauchter Stimme der Fall ist, kommt es zu Turbulenzen im Luftstrom, die sich in einem erhöhten spektralen Rauschen äußern. Maße hierfür sind z.B. HNR (*Harmonics-to-Noise-Ratio*) und NHR (*Noise-to-Harmonics-Ratio*). Beide Maße treffen eine Aussage über das Energieverhältnis von periodischen zu aperiodischen Signalanteilen. Sie erfassen allerdings nicht nur glottale Friktion, sondern auch anderweitige Rauschquellen im Vokaltrakt (oder sogar außerhalb). Daher sollten diese Verhältnismaße nur in rein vokalischen Signalabschnitten gemessen und auf einen optimalen Signal-Rausch-Abstand in der Aufnahmequalität geachtet werden. (Vgl. de Krom 1994: 14)

Auch die Form der spektralen Hüllkurve ist ein aussagekräftiges Maß für die Stimmqualität eines Sprechers. Die spektrale Neigung ist typischerweise am flachsten für geknarrte und am steilsten für behauchte Stimme, d.h. die Abnahme der Energie von niedrigeren zu höheren Harmonischen ist bei Knarrstimme am geringsten (ca. 6 dB/Oktave), bei behauchter Stimme am größten (ca. 18 dB/Oktave). Modale Stimme findet sich dazwischen mit einer Energieabnahme von ca. 12 dB/Oktave. (Vgl. Klatt 1990: 825)

Zur Quantifizierung der spektralen Neigung werden verschiedene Verhältnismaße herangezogen, die sich als unterschiedlich zuverlässige (d.h. studienabhängige) akustische Korrelate für Stimmqualität erwiesen haben, beispielsweise die Differenz der Amplituden zwischen der ersten und der zweiten Harmonischen oder zwischen F0 und F1. (Vgl. Gordon 2001: 15)

3. Perzeption von Stimme

Bei der Vielzahl an akustischen Parametern und Messmethoden und der Komplexität der experimentellen Untersuchungen der Wahrnehmungspsychologie ist es nicht verwunderlich, dass es eine scheinbar unüberschaubare Anzahl an Studien zum Thema Perzeption und Akustik von Stimme gibt, deren Ergebnisse ebenso vielfältig sind. Daher sollen an dieser Stelle nur beispielhaft einige im wissenschaftlichen Diskurs relevante Studien vorgestellt werden.

3.1 Perzeptive Beurteilung von Stimmqualität

Insbesondere in der klinischen Phonetik und Logopädie ist die Evaluation von Stimmeigenschaften von Bedeutung, wofür sowohl objektive als auch subjektive Maße

eine Rolle spielen. Als perzeptives Bewertungssystem hat sich hier GRBAS (entsprechend der Bewertungsskalen *Grade, Roughness, Breathiness, Aestehnia* und *Strain*[5]) etabliert, für die Evaluation auf Grundlage akustischer Messungen das Sprachanalyseprogramm MDVP (Multi-Dimensional Voice Program), welches bis zu 34 akustische Maße für Stimmqualität berechnet. Bhuta/Patrick/Garnett (2003) untersuchten mögliche Korrelationen zwischen beiden Bewertungssystemen und konnten über Regressionsanalysen drei signifikante Prädiktoren auf akustischer Seite für die perzeptiven Parameter feststellen. In ihrer Studie korrelierten zum einen das Verhältnis von tieffrequenter zu hochfrequenter harmonischer Energie mit der wahrgenommenen Abnormität und Behauchtheit der Stimme und einer generell wahrgenommenen Stimmschwäche. Weiterhin konnte aus dem Maß der Aperiodizität die Bewertung einer Stimme als abnorm und rau vorhergesagt werden. Dementsprechend korrelierte auch das Maß für das Verhältnis von aperiodischer und harmonischer spektraler Energie mit der Wahrnehmung der Abnormität der Stimme. (Vgl. Ebd.: 301)

Den Zusammenhang zwischen wahrgenommener behauchter Stimme und Eigenschaften der spektralen Hüllkurve konnte Klatt (1990) belegen. Hierfür erwies sich der Spitzenpegel der ersten Harmonischen im Verhältnis zur zweiten Harmonischen oder zum ersten Formanten als aussagekräftiges Maß, wobei eine relativ höhere erste Harmonische mit behauchter Stimme einhergeht. (Vgl. Klatt 1990: 824, 836). Weiterhin konnte Klatt sowohl akustisch als auch perzeptorisch einen Unterschied für männliche und weibliche (amerikanische) Sprecher im generellen Behauchtheitsgrad der Stimme belegen. (Vgl. ebd.: 852 f.)

3.2 *Vocal attractiveness* und Stimmstereotype

In Studien zur Stimmattraktivität konnten geschlechtsspezifisch Stimmeigenschaften aufgezeigt werden, die mit hörerseitigen Stimmpräferenzen oder der assoziierten physischen Attraktivität der Sprecher korrelierten. Liu/Xu (2011) fanden für die Beurteilung weiblicher Stimmen durch männliche Sprecher heraus, dass sowohl Stimmqualität als auch die Vokaltraktlänge (repräsentiert durch das Verhältnis der Langzeitdistribution von Formanten im Spektrum[6]) und durchschnittliches F0 mit der Stimmattraktivität zusammenhängen. Dabei werden behauchtere und höhere Stimmen und eine geringere Vokaltraktlänge als attraktiver wahrgenommen. Collins (2000) untersuchte hingegen das Verhältnis zwischen männlichen Stimmeigenschaften und Urteilen von weiblichen Hörerinnen über die Sprecher. Ihr Interesse bezog sich dabei vor allem auf individuumsübergreifende Urteile der Hörerinnen zu physischen Eigenschaften der männlichen Sprecher anhand ihrer Stimmen. Ihre Studie weist darauf hin, dass männliche Sprecher mit tieferen und engeren Formantlagen (was mit einem längeren Vokaltrakt korreliert) und starken tieffrequenten Energieanteilen als attraktiver, älter und muskulöser wahrgenommen werden. Zu ähnlichen Ergebnissen kommen auch Fein-

[5] *Grade* = genereller Grad der Abnormität der Stimme, *Roughness* = Irregularitäten in der Grundfrequenz, *Breathiness* = Luftstromturbulenzen durch Luftverlust an der Glottis, *Aesthenia* = generelle Schwäche der Stimme, *Strain* = Eindruck von Gespanntheit oder erhöhtem Aufwand (Vgl. Bhuta/Patrick/Garnett 2003: 300).
[6] engl. *Long-Term Formant Distribution.*

berg/Jones/Little/Burt/Perret (2004). Allerdings korreliert diese Wahrnehmung nicht mit den tatsächlichen Eigenschaften der Sprecher. (Vgl. Collins 2000: 777)

Die Suche nach den akustischen Korrelaten einer (geschlechtsunabhängigen) schönen Stimme beschäftigte u.a. Zuckerman/Miyake (1993). Zusätzlich untersuchten sie den Zusammenhang zwischen Stimmattraktivität und subjektiven Maßen von Stimmklang. Als Prädiktoren für hohe Stimmattraktivitätswerte konnten Wahrnehmungsgrößen wie Lautheit und Resonanz[7] aufgezeigt werden. Auf akustischer Seite erwies sich die durchschnittliche Grundfrequenz als Einflussgröße auf die Stimmattraktivität, hier allerdings wiederum nur für männliche Sprecher. Weiterhin wurde das Verhältnis von attraktiven Stimmen zu vermuteten Charaktereigenschaften untersucht und Zusammenhänge zwischen der Stimmattraktivität und assoziierten Eigenschaften wie ehrlich, offen, sympathisch oder auch dominant aufgezeigt. (Vgl. Berry 1992 und Zuckerman/Driver 1989)

K. R. Scherer (1972) führte eine groß angelegte Studie zur Beurteilung von Persönlichkeitsmerkmalen anhand der Stimme mit amerikanischen und deutschen Sprechern und Hörern durch. Hierbei wurden zunächst Persönlichkeitsprofile von insgesamt 59 Sprechern anhand von Selbst- und Fremdeinschätzung (durch Freunde der Sprecher) erfasst. Sodann wurden mit einer Auswahl der Sprecher Sprachaufnahmen im Rahmen einer simulierten Diskussion von Geschworenen bei Gericht erstellt. Die daraus entstandenen Stimuli wurden wiederum durch amerikanische und deutsche Hörer hinsichtlich ihrer Assoziationen in bestimmten Charakterdimensionen wie z.B. emotionale Stabilität oder Gewissenhaftigkeit (*emotional stability* und *conscientiousness*, vgl. Scherer 1972: 194) bewertet. Dabei zeigte sich eine signifikante Übereinstimmung der Hörerurteile mit den Fremdeinschätzungen (allerdings nicht mit den Selbsteinschätzungen).

All diese Studien haben gemeinsam, dass sie sich auf die Suche nach Stimmstereotypen begeben, sei es für physische Attraktivität oder für charakterliche Eigenschaften, und tatsächlich einschlägige Ergebnisse zugunsten des Vorhandenseins solcher Stimmstereotype liefern. Die Frage zu stellen, wieviel die vorhandenen Stereotype mit der Realität zu tun haben, ist ein nachvollziehbarer zweiter Schritt und deren Antwort abhängig vom Phänomenbereich. Die vorliegende Arbeit ordnet sich ebenso in dieses Forschungsgebiet ein und stellt beide Fragen mit Blick auf den Phänomenbereich Stimmbildung. Daher soll das abschließende Theoriekapitel zur Darstellung der Stimmbildungsaufgaben im Gesangsunterricht dienen.

4. Stimmbildung und Gesangsunterricht

Bereits seit Jahrhunderten werden Abhandlungen zur Theorie der Formung einer ausgereiften, wohlklingenden Stimme mithilfe von Gesangsunterricht bzw. Stimmbildung

[7] Resonanz als Wahrnehmungsgröße wird hier folgendermaßen beschrieben: *„Resonant voices are smooth, strong, and without a very wide range of pitch“* (Zuckerman/Miyake 1993). Leider geben die Autoren keine Erklärung, warum der Tonhöhenumfang ausschlaggebend für eine wahrgenommene Resonanz sein soll.

verfasst, deren Ansichten sich im Laufe der Zeit immer wieder wandelten. Der folgende Abschnitt gibt einen groben historischen Überblick über die Ansprüche an die ausgebildete Stimme des Sängers und an den Stimmbildungsunterricht. Hiernach wird vor allem P.-M. Fischers (1993) Theorie zur Stimmbildung dargestellt und Aufgaben der Stimmbildung beschrieben. Dabei werden auch die Studien zur Singstimme von J. Sundberg (1987) herangezogen.

4.1 Entwicklung der Stimmbildungstheorie

„Dionysos von Halikarnass (1. Jahrh. v. Chr.) benannte die Laute nach ihrer ästhetischen Wirkung. Er ging von den Vokalen aus und bezeichnete die langen, bei denen der Atem voll verströmen könne, als die anmutigsten und ausdrucksvollsten." (Fischer 1993: 20) Es ist also nicht verwunderlich, dass Vokale und Vokalisen zu jeder Zeit Hauptträger von stimmbildenden Gesangsübungen waren und dies auch heute noch sind.

Über die Jahrhunderte hinweg änderten sich jedoch die Klanganspräche an den Sänger. Das Klangideal des Mittelalters strebte eine sanfte Tongebung an. Die Expansionsmöglichkeit der Stimme war eingeschränkt, die hohe und tiefe Stimmlage wenig ausgeprägt. Erst im Übergang vom späten Mittelalter zur Renaissance verschob sich die Klangvorstellung zu einem resonanzreichen, vollen Klang. (Vgl. ebd.: 26)

Bis in das 16. Jahrhundert hinein wurde eine *„raue, tremolierende Tongebung"* (ebd.) noch verschmäht. Der mehrstimmige Gesang stand im Vordergrund, die Stimme hatte sich also in den Gesamtklang einzufügen. Erst gegen Ende des 16. Jahrhunderts wich die Anonymität der Einzelstimme als Teil polyphoner Musik langsam solistisch eingesetzter Stimmvirtuosität. (Vgl. ebd.: 28) Insbesondere mit dem Aufkommen der Oper im 17. Jahrhundert wurde der Anspruch an die Tragfähigkeit der Solostimme erhöht.

> „Die Aufgaben in der Gesangspädagogik wurden in dieser Zeit durch die Tatsache geprägt, dass der virtuose Solist gänzlich aus dem Verband der Ensemblesänger heraustrat. Er hatte mit seiner Stimme nicht nur Bühnenräume, sondern auch mit Prunk ausgestattete Kirchen zu füllen. Damit wurden jedoch auch die Mängel der stimmtechnischen Bildung deutlich offenbar, und der Wunsch kam auf, diese Unzulänglichkeiten zu beseitigen." (Ebd.: 29)

Im 19. Jahrhundert änderten sich die Anforderungen an den Sänger erneut. Die Oper wurde neu konzipiert, sodass den Sängern eine Gesangstechnik abverlangt wurde, die auch ein schauspielerisches Agieren möglich machte. Zusätzlich wurde infolge des Einsatzes größerer Orchester eine erhöhte Stimmleistung gefordert. Neue wissenschaftliche Entdeckungen in der Stimmwissenschaft ermöglichten nun eine physiologisch fundierte Gesangs-pädagogik. Neu ist ebenfalls die Aufteilung der Stimme in die sogenannten „Register". Hierbei wurden Brust-, Mittel- (Falsett-) und Kopfstimme unterschieden und ihre bewusste Mischung durch den Sänger gefordert. (Vgl. ebd.: 34)

Ebenso hielt die Theorie des „gestützten" Singens, also der Aktivierung der Einatmungsmuskeln während der Stimmgebung, und die gezielte Aussprache-schulung Ein-

zug in die Gesangspädagogik. Müller-Brunow konstatierte, dass nicht der Sprechton der richtige Sington sei, sondern umgekehrt:

> „Es giebt [sic!] nur einen Weg, alles, was die Natur dem Menschen sing- und klingbares in Kehle und Brust gelegt hat, zum Klingen zu bringen. In erster Linie durch die Hinwegräumung aller Hindernisse, geschaffen durch die Natur der Gewohnheit – Sprache, Sprechen, Dialekt – und falsches Erfassen des Begriffes „singen". Dieses Hinwegräumen aller Kulturanhängsel nennen wir das Suchen nach dem primären Ton – das heisst [sic!] also, nach dem Grundton eines jeden Menschen, der seiner Natur – nicht nach Gewohnheit – sondern nach Bau und Stimmenlage am nächsten, am bequemsten liegt, [...]." (Müller-Bruno w 1890: 9)

Müller-Brunow fordert also die optimale Einstellung des Kehlkopfes und des Ansatzrohres. Im 20. Jahrhundert wurde der Standpunkt vertreten, dass die Stimme durch das Verhältnis der Register zueinander bestimmt werde. Es gäbe nur zwei Funktionen der Stimmlippen, die Vollschwingung (Bruststimme) und die Randschwingung (Kopfstimme, Falsett) mit den daraus resultierenden Resonanzen in Brust und Kopf. Alles andere seien Mischverhältnisse dieser beiden Hauptfunktionen. Wesentlich bei diesem Ansatz war, dass „*die unverdorbene, unberührte Naturstimme die Registerunterschiede nicht kenne*" (Fischer 1993: 42). Der Sänger in der Ausbildung müsse also das „Einregister" anstreben. In stimmphysiologischen Untersuchungen wurde hinsichtlich der Kehlkopfbewegung herausgefunden, dass bei Berufssängern beim Auf- und Abwärtssingen diese Bewegungen umso kleiner sind, je besser die Stimmen geschult waren. Grundsätzlich liege der Kehlkopf bei Kunst-(Berufs-)sängern tiefer als bei Laiensängern. (Vgl. ebd: 43)

Von grundlegender Bedeutung für die Stimmbildungsforschung war auch die Entdeckung der „Sängerformanten". Es wurden zwei Formanten bestimmt, zum einen der „Tiefformant" im Bereich von 400 bis 600 Hz für Männer und Frauenstimmen, und zum anderen der „Hochformant", der für die Tragfähigkeit der Stimme mitverantwortlich ist. Dieser Hochformant liegt bei Männerstimmen im Bereich von 2400 Hz bis 3200 Hz, bei Frauenstimmen um 3200 Hz. (Vgl. Sundberg 1987: 118)

4.2 Singstimme und Sängerstimme

Die Stimmbildung ist wesentlicher Bestandteil des Gesangsunterrichts. Heute beinhaltet sie das Trainieren der richtigen Atmung, der Stütze, das Erreichen einer Intensitätssteigerung durch die Erzeugung der Sängerformanten, den bruchlosen Übergang zwischen den Registern durch Herunterziehen der Kopfstimme bzw. Hochziehen der Bruststimme und deren Mischung (Einregister) sowie den weichen Stimmansatz.

Fischer (1993: 71) unterscheidet die sogenannte „Singstimme" von der „Sängerstimme" nach ihrer Funktion und Ausdrucksvermögen:

> „Die Singstimme ist dadurch definiert, dass sie nicht die optimale oder gar maximale Leistung des Stimmorgans darstellt, selbst dann nicht, wenn das Singen, wie bei vielen Menschen, mit einer erheblichen Anstrengung verbunden ist. […] Die Sing-

> stimme ist also diejenige Stimme, die nicht mehr einer vollen Leistung fähig ist. […] Physikalisch-akustisch ist die Singstimme dokumentiert durch einen völlig „gerade" verlaufenden, linearen Ton, der sich ohne jede Regung in Brust und Kehle bewegt und zwar in einem Umfang der Stimme, der zwei Oktaven nur selten überschreitet."

Die Sängerstimme hingegen sei (nicht nur nach Fischers Ansicht) von der Natur gegeben. Die durchschnittliche Singstimme hebe sich dann von der Sängerstimme ab als ein *„durch jahrelange und umständebedingte Nichtübung atrophiertes, funktionsgeschwächtes Phänomen."* (Ebd.: 1) Bernd Weikl spricht in seinem gesangstheoretischen Buch von der „Denaturierung" des Stimmorgan:

> „Die Neugeborenen können noch alle richtig atmen und Töne „produzieren". Wird dann der Rede- oder Singdrang im Elternhaus, in der dünnwandigen Hochhauswohnung, in der Schule oder im sonstigen öffentlichen Leben eher gehemmt als gefördert, so führt dies zu einer Denaturierung dieses Organs (Phonasthenie) und zum Verlust der natürlichen Fähigkeiten, damit artspezifisch „lautstark" umgehen zu können." (Weikl 1998: 38)

Die Sängerstimme, also das Produkt eines durch Stimmbildung wieder zu voller Leistungsfähigkeit rückgeführten Stimmorgans, zeichne sich gegenüber der Singstimme durch das Vibrato, einen größeren Stimmumfang, erweiterte Dynamik, die Herausbildung des Tief- und Hoch-(Sänger-)formanten, Tonintensität (Durchdringungskraft), Stimm- und Atemführung aus.[8] Die laut Fischer entscheidenden Merkmale der Sängerstimme seien *„ein ebenmäßiges Vibrato und der obere Singformant"*. (Fischer 1993: 73)

Akustisch entspricht das Vibrato einer periodischen, sinusartigen Modulation der Grundfrequenz und ist durch Modulationsrate und -umfang charakterisiert. Diese periodische Tonhöhenschwankung ist durch eine im Vergleich zum *Jitter* geringe, sängerabhängige Modulationsrate von ca. 5 bis 10 Hz und einen wiederum sängerabhängigen Modulationsumfang von etwa einem Halbton deutlich auditiv wahrnehmbar. Physiologisch geht es mit einer Vibration der laryngalen Muskeln (insbesondere des Cricothyroid, des Musculus vocalis und des Cricoarytenoid) und der an der Atmung beteiligten Muskeln einher, was eine periodische Modulation des subglottalen Drucks hervorruft. (Vgl. Sundberg 1987: 163ff)

Die Erzeugung des hohen Sängerformanten geschieht durch Bündelung der oberen Formanten (F3, F4 und F5). Dies wird durch Modulation des Ansatzrohres realisiert, indem zum einen der Kehlkopf abgesenkt und zum anderen der untere Rachenraum geweitet wird. Somit ändern sich sowohl die Länge als auch die Form des Vokaltraktes und dementsprechend seine Resonanzfrequenzen ohne Einfluss auf die für die Vokalqualität entscheidenden Formanten. (Vgl. Seidner/Wendler 1997: 120)

[8] Es sei an dieser Stelle darauf hingewiesen, dass die gängige Stimmbildungsliteratur nicht vorwiegend empirischer Natur ist und somit die vorhergehenden Ausführungen impressionistischen Beobachtungen entspringen.

Singen erfordert bei der Atmung den Gebrauch eines größeren Anteils der Vitalkapazität der Lunge als dies beim Sprechen der Fall ist. Sänger streben daher an, ihre Vitalkapazität zu erhöhen. Durchschnittlich besitzen Berufssänger eine um ca. 20% höhere Vitalkapazität als Nichtsänger. Dies wird jedoch nicht durch eine Vergrößerung der Totalkapazität der Lunge realisiert, sondern durch eine Verringerung des Residualvolumens. (Vgl. Sundberg 1987: 35) Ebenso werden die Atemmuskeln und die adduktive Spannung der Stimmlippen trainiert, um den subglottalen Luftdruck, der die Lautstärke bestimmt, sehr präzise steuern zu können. Dies ist nicht nur für die Umsetzung des dynamischen Ausdrucks (Lautstärkegestaltung) eines Stückes von Bedeutung, sondern auch für die Mikrovariationen des subglottalen Drucks, die nötig sind, um die Lautstärke in unterschiedlichen Tonhöhen anzupassen.

Das Singen in unterschiedlichen Registern geht mit einem jeweils registerspezifischen Schwingungsverhalten der Stimmlippen und damit mit unterschiedlichen Frequenzspektren einher. Die angestrebte Mischung der Register zielt also physiologisch darauf ab, das gesamte Schwingungspotential der Stimmlippen auszunutzen und ein stabiles Frequenzspektrum zu erzeugen, was sich perzeptorisch in einer gleichbleibenden Stimmqualität für verschiedene Bereiche des Stimmumfangs äußert.

4.3 Sprecherziehung in der Gesangsausbildung

Die Gesangausbildung umfasst neben dem Training der Singstimme auch sprecherzieherische Maßnahmen. Dies beinhaltet zweierlei: einen ökonomischen Stimmgebrauch beim Sprechen und den Einsatz stimmlicher Mittel, um bestimmte Wirkungen zu erreichen. Letzteres betrifft variable Stimmkonfigurationen bezüglich der Stimmqualität und prosodischer Muster, aber auch das Trainieren einer deutlichen Artikulation (Bühnensprache). Zum ökonomischen Stimmgebrauch gehören laut Seidner/Wendler (1997: 157) *„die Orientierung auf die Indifferenzlage, das Sprechen mit optimal angepaßtem Anblasedruck und das Vermeiden häufiger harter Einsätze“*. Weiterhin gelte es, Verengungen im Rachenraum und geringe Kieferöffnung zu vermeiden, um das Resonanzvermögen des Vokaltrakts optimal zu nutzen. (Vgl. ebd.) Die (künstlerische) Sprecherstimme unterscheide sich dabei vom alltäglichen Stimmgebrauch. *„Durch die bewußte Führung und Kontrolle der Stimme während des Sprechvorgangs nähern sich einzelne Leistungen der Sprecherstimme den Singstimmfunktionen an.“* (Ebd.: 156)

5. Hypothesen zum Zusammenhang zwischen Stimmbildung, Akustik und Perzeption

Mit der Stimmbildung im Rahmen von Gesangsunterricht wird u.a. das Erlernen der folgenden Fertigkeiten angestrebt, die Einfluss auf die Quelle- und Filterkonfiguration haben:

- Kontrolle des Atemflusses
- Kontrolle der muskulären Spannungen im Kehlkopf
- Tiefe Kehlkopflage und erweiterter Rachenraum
- deutliche Aussprache für eine hohe Ausdruckskraft

Durch Aneignung dieser Fertigkeiten können wiederum die folgenden Stimmeigenschaften erreicht werden:

- → gleichmäßige und ganzheitliche Stimmlippenschwingung
- → großer Dynamikumfang und Dynamikkontrolle
- → weicher Stimmansatz
- → resonanzverstärkende Konfiguration des Vokaltraktes und damit große Tragfähigkeit der Stimme
- → keine artikulatorischen Verschleifungen und deutliche Stimmhaftigkeit
- → hohe Expressivität der Sprache

Diese Eigenschaften gehen theoretisch mit bestimmten akustischen Eigenschaften der Stimme einher. Für die vorliegende Untersuchung ist besonders der Aspekt der gleichmäßigen und ganzheitlichen Stimmlippenschwingungen von Bedeutung, die mit sehr geringen Mikrovariationen in Dauer und Amplitude der Schwingungen korrelieren und sich dementsprechend in niedrigen *Jitter*- und *Shimmer*-Werten und hohen Harmonizitätswerten ausdrücken. Weiterhin kann die resonanzverstärkende Konfiguration des Ansatzrohres durch Tieflage des Kehlkopfes zu verschobenen Formantlagen führen.

Beide Eigenschaften sollten in Anlehnung an die in Kapitel 3.2 erwähnten Studien zu einer höheren wahrgenommenen Stimmattraktivität führen. Insofern lassen sich für den experimentellen Teil der Arbeit folgende verkettete Hypothesen formulieren:

A) Eine als wohlklingend wahrgenommene Stimme ist mit bestimmten stimmqualitativen Eigenschaften verbunden.
B) Bevorzugt (sängerisch) ausgebildete Sprecher haben Stimmen mit solchen Eigenschaften.
C) Ihre Stimmen werden dementsprechend als wohlklingender empfunden als nicht ausgebildete Stimmen.
D) Es gibt hörerseitige Stimmstereotype für ausgebildete Stimmen.
E) Dieser Stereotyp überschneidet sich mit dem Stereotypen für wohlklingende Stimmen.

Natürlich können in der Querschnittsstudie keine Rückschlüsse auf kausale Zusammenhänge zwischen Stimmbildung und Stimmattraktivität gezogen werden. Dennoch sollte die Ergebnisanalyse zur Sichtung erster Hinweise darauf, ob die formulierten Hypothesen verworfen werden müssen oder eine weitere Prüfung lohnenswert erscheint, beitragen.

6. Sprachaufnahmen

6.1 Sprecher

Zunächst wurde ein Sprachkorpus aufgebaut, für welches ca. 25-minütige Aufnahmen von insgesamt 23 Sprechern erstellt wurden. Als Sprecher wurden hier Personen mit

unterschiedlichem Stimmbildungslevel ausgewählt, wobei von drei Kategorien ausgegangen wurde:

1. Gesangsunerfahrene (Kategorie GU): Die Sprecher dieser Kategorie durften keinerlei angeleitete Gesangserfahrung besitzen, d.h. auch Chorgesang war nur zu einem Mindestmaß erlaubt (beispielsweise Kinderkirchenchor).
2. Laiensänger (Kategorie LS): Hier wurde mindestens ein Jahr Einzelgesangsunterricht vorausgesetzt und regelmäßige solistische oder chorische Tätigkeit, die jedoch nicht über das Maß einer Freizeitbeschäftigung hinaus ausgeübt wird.
3. Professionelle Sänger (Kategorie PS): Für diese Kategorie wurde ein abgeschlossenes Gesangsstudium und die berufliche Tätigkeit als Sänger vorausgesetzt.

In allen drei Kategorien sollten die Sprecher Nichtraucher und stimmlich gesund sein. Zudem wurde das Alter auf höchstens 50 Jahre für männliche und 45 Jahre für weibliche Sprecher beschränkt. Die Wahl der oberen Altersgrenze begründet sich auf mehrere Studien zum Einfluss des Alters auf die Stimme. Untersuchungen zu diesem Thema konnten zeigen, dass die profilstiftenden physiologischen und akustischen Werte bei Männern von etwa 20 bis 55 Jahren, bei Frauen von etwa 16 Jahren bis zum Eintreten der Wechseljahre (ca. zwischen 40 und 58 Jahren) stabil bleiben. (Vgl. Xue/Deliyski 2001, Brückl/Sendlmeier 2003 und Schötz/Müller 2007)

Die Sprecher wurden durch Aushang und persönliche Anfragen (insbesondere bei den professionellen Sängern) geworben. Termin und Aufwandsentschädigung wurden jeweils individuell abgesprochen. Weiterhin wurden die Sprecher gebeten, zu den Aufnahmen mindestens zwei Stunden nach dem Aufstehen und nicht nach 18 Uhr zu erscheinen, um den Einfluss der Tageszeit auf die Stimme möglichst gering zu halten. (Vgl. Görs 2011)

Es wurden vier männliche und vier weibliche Sprecher in der Kategorie Gesangsunerfahrene im Alter von 23 bis 45 Jahren aufgenommen. In der Kategorie Laiensänger wurden vier männliche und sechs weibliche Sprecher aufgenommen, wobei zwei der weiblichen Sprecher mit bilingualem Hintergrund aus den weiteren Analysen ausgeschlossen und im Perzeptionsexperiment nur für „Dummy"-Stimuli herangezogen wurden, sodass auch in der Kategorie Laiensänger insgesamt acht Sprecher zwischen 26 und 46 Jahren erfasst wurden. Die Akquisition der professionellen Sänger erbrachte zwei männliche und drei weibliche Sprecher (32 bis 43 Jahre). Anhang A enthält eine Übersicht über die Angaben aller Sprecher.

6.2 Aufnahmen

Die Aufnahmen wurden im Tonstudio des Instituts für Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft durchgeführt. Nach der Aufklärung über den Experimentverlauf und der Unterzeichnung der freiwilligen Einwilligung wurden die Probanden an das EGG (Glottal Enterprises EG2-PCX) angeschlossen, wobei hier das Signal des Headsetmikrofons (linker Kanal „Mic") und der auf Höhe des Kehlkopfes befestigten Elektroden (rechter Kanal „EGG") im Verstärker zu einem Stereosignal zusammen-

gefasst und über ein portables Aufnahmegerät (Zoom H2 Handy Recorder) mit 48 kHz Samplingrate und 24 Bit Samplingtiefe aufgenommen wurden. Insbesondere für die Messung stimmqualitativer Parameter ist eine gute Aufnahmequalität von Bedeutung, daher sollten durch den Einsatz des akkubetriebenen Aufnahmegeräts Interferenzen durch Geräteeigenschwingungen so gering wie möglich gehalten werden, um so einen guten Signal-Rauschabstand der Aufnahmen zu erzielen.

Die Aufnahmen umfassten sowohl Lese- als auch Spontansprache. Der Ablauf der Aufnahmen war für alle Probanden derselbe. Die Texte konnten von den Probanden vor Betreten der Aufnahmekabine ohne Zeitbeschränkung gelesen und Unklarheiten durch Rückfragen beseitigt werden. Da im aufbauenden Perzeptionsexperiment nur ein kleiner Teil des erhobenen Sprachmaterials genutzt wurde, soll hier eine kurze Aufzählung der Texte genügen. Die Auswahl wurde jedoch mit dem Blick auf mögliche zukünftige akustische Untersuchungen zu Stimmqualität, Prosodie und Artikulationscharakteristika getroffen. Die Texte beinhalteten eine Liste von 19 nach phonetischen Kriterien ausgewählten Phrasen, die einmal in neutraler und einmal in expressiver Sprechweise zu lesen waren. Hinzu kamen sieben kurze Gedichte und ein zweiseitiger humoristischer Prosatext. Die Probanden wurden darauf hingewiesen, dass es nicht um fehlerfreies Lesen ginge und Versprecher jederzeit von ihnen korrigiert werden können. Weiterhin wurden die Probanden gebeten, die Wörter <ja>, <so>, <nee>, <du> und <nie> jeweils mit fallender, steigender und ebener Intonation zu produzieren. Für die spontansprachliche Aufnahme erhielten die Probanden die Anweisung, ca. 5 min aus ihrem Werdegang zu berichten. Dabei saß die Versuchsleiterin als Adressatin den Probanden gegenüber, beschränkte sich jedoch auf nonverbales Feedback.

Nach der Aufnahme füllten die Probanden einen personenbezogenen Fragebogen aus, in dem sie neben demographischen Angaben die Gesundheit ihres Stimmapparates bestätigten. Weiterhin wurden sie gebeten, so ausführlich wie möglich ihre Stimmbildungs- und anderweitige Gesangserfahrung aufzulisten.

6.3 Datenaufarbeitung

In der Nachbearbeitung wurde das Aufnahmematerial zunächst in je 15 Stereo-wav-Dateien pro Sprecher geschnitten, wobei aus den Lesetexten Häsitationen und Versprecher entfernt wurden.[9] Der Prosatext wurde in drei Abschnitte unterteilt. Aus den 5-minütigen spontansprachlichen Aufnahmen wurden für jeden Sprecher all diejenigen Sätze ausgewählt, deren Inhalt nicht auf die gesangliche Ausbildung des Sprechers schließen lässt (zwischen 5 und 11 Sätze pro Sprecher). Danach wurden die Kanäle getrennt und orthographische Transkripte der Texte erstellt. Über das Webtool WEBMAUS des Bayerischen Archivs für Sprachsignale (http://clarin.phonetik.uni-muenchen.de/BASWebServices) wurden in Praat (Boersma/Weenink 2014) lesbare TextGrids mit orthographischer, phonemisch-kanonischer und phonetischer Etikettenebene erstellt. Diese automatische Segmentierung diente als Grundlage für die halbautomatische Messung der akustischen Parameter, deren methodische Beschreibung sich

[9] Dies war ein für die spätere automatische Segmentierung und Etikettierung notwendiger Schritt und für weitere Untersuchungen unerheblich, da nicht die Lesekompetenz der Sprecher im Fokus steht.

aus praktischen Gründen an die Beschreibung der Methode des Perzeptionsexperimentes anschließt. Für alle 21 in die Analyse mit einfließenden Sprecher liegen also 60 Dateien vor: 45 Wav-Dateien (48 kHz, 24bit), davon 15 Stereo (EGG +Mic)[10], 15 Mono (Mic) und 15 Mono (EGG), und 15 TextGrids.

7. Perzeptionsexperiment

Der Konzeption des Perzeptionsexperiments und der akustischen Analyse liegen zwei Fragenkomplexe zugrunde, die sich aus den einleitenden Basisfragen und den in Kapitel 5 aufgestellten Hypothesen ableiten.

Fragenkomplex 1: Gibt es individuumsübergreifende Stimmpräferenzen? Wenn ja, sind diese Stimmpräferenzen mit bestimmten sprachlich erfassbaren oder sogar akustisch messbaren klanglichen Eigenschaften verbunden? Haben ausgebildete Stimmen eher solche Eigenschaften und sind damit objektiv „schöner“ als unausgebildete Stimmen?

Hierfür ist im Perzeptionsexperiment zum einen die Bewertung der Schönheit einer Stimme einzuflechten. Zum anderen muss die sprachliche Beschreibung der Stimme mit abgefragt werden, wofür eine offene Frageform naheliegend ist. Da die entsprechende Auswertung jedoch sehr komplex und fehleranfällig ist, sollte auf eine geschlossene Form z.B. durch ein semantisches Differential zurückgegriffen werden. Zusätzlich sollen für eine Untersuchung außerhalb des Rahmens der vorliegenden Arbeit das Vorhandensein von individuumsübergreifenden „cross-modalen Korrespondenzen“[11] anhand von Texturen, also einer außersprachlichen Möglichkeit zur Charakterisierung von Stimmen, überprüft werden.

Fragenkomplex 2: Existieren auf Hörerseite stereotype Klangvorstellungen von ausgebildeten Stimmen? Inwieweit entsprechen diese Vorstellungen der Realität und lassen dadurch ausgebildete Stimmen auditiv identifizierbar werden? Für den Fragenkomplex 2 sollen im Perzeptionsexperiment intuitive Urteile über das Stimmbildungslevel eines Sprechers abgefragt werden.

Es sei an dieser Stelle nochmals erwähnt, dass die Ergebnisse des hier vorliegenden Experiments bestenfalls erste Tendenzen zur Beantwortung der Fragenkomplexe aufzeigen können.

Zur Annäherung an die Beantwortung der Fragenkomplexe sind also vier Aufgaben an die Hörer zu richten:

---

[10] Ohne High-Pass-Filter und Phasenverschiebungsausgleich zwischen EGG und Mic, d.h. es zeigen sich im EGG-Signal noch die tieffrequenten Artefakte des Verstärkers und der linke (Mic) Kanal ist zum rechten (EGG) aufgrund des Abstandes zwischen Stimmlippen und Abschallöffnung um ca. 0,5 ms verschoben.

[11] Als cross-modale Korrespondenzen (manchmal auch schwache Synästhesie) werden Assoziationen in einer bestimmten Sinnesdimension zu Reizen einer anderen Sinnesdimension bezeichnet. Dieses Phänomen ist sehr häufig auch in der Alltagssprache zu finden, z.B. in der Wortverbindung „heller Klang“.

1. die Einschätzung des Stimmbildunglevels,
2. die Bewertung der Schönheit einer Stimme,
3. die Zuweisung von Adjektiven zu einer Stimme und
4. die Zuweisung von Texturen zu Stimmen.

Es ist erstrebenswert, eine gegenseitige Beeinflussung der einzelnen Aufgaben auszuschließen. Beispielsweise könnte das Bewusstmachen von stimmlichen Vorlieben bei den Hörern zu verschobenen Adjektivzuweisungen führen, da diese für den Hörer selbst positiv oder negativ konnotiert sein könnten und widersprüchliches Antwortverhalten womöglich bewusst vermieden wird. Gleiches gilt in entgegengesetzter Richtung und ein ähnliches Risiko besteht auch zwischen dem Urteil des Stimmbildungslevels und den anderen Aufgaben. Um in der statistischen Auswertung genügend große Stichproben nutzen zu können, ist es jedoch praktisch nicht umsetzbar, für jede der vier Aufgaben ein einzelnes Experiment anzusetzen. Eine Mindestbedingung sollte jedoch sein, dass die obige Aufgabe 1 intuitiv beantwortet werden kann, also unabhängig von den anderen Aufgaben getestet wird. Dementsprechend wurde diese Aufgabe an den Anfang des Perzeptionsexperiments gestellt. Ebenso sollten die Aufgaben 2 und 3 unabhängig voneinander getestet werden, wofür die Stichprobe der Hörer in zwei Gruppen aufgeteilt wurde.

Der Fokus bei der Untersuchung von Stimmbewertungen liegt auf natürlicher Sprache. Dementsprechend wurde im Perzeptionsexperiment größtenteils spontansprachliches Material als Stimuli genutzt. Um einen ersten Einblick in das Verhältnis von stimmqualitativen zu prosodischen und artikulatorischen *Cues* in der Beurteilung von Stimmen zu erhalten, soll jedoch auch expressive Lesesprache, in denen letztgenannte *Cues* vermutlich stärker hervortreten, mit eingesetzt werden. Das Experiment besteht daher aus drei Hauptteilen. Teil 1 und Teil 2 gliedern sich jeweils in zwei Subteile A und B, wobei für die Teile 2A und 2B die Hörerstichprobe zufällig aufgeteilt wurde (siehe Abbildung 2).

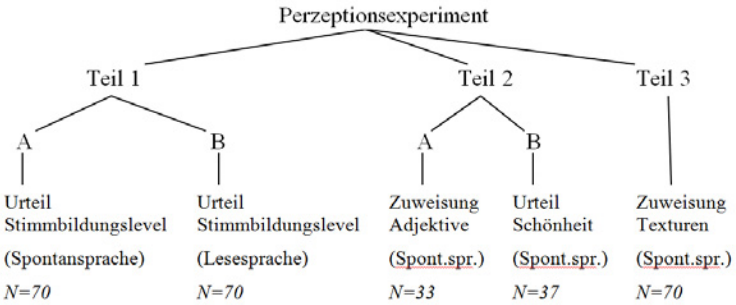


Abbildung 2. Struktur des Perzeptionsexperiments.

7.1 Stichprobe

Am Perzeptionsexperiment nahmen 70 Personen zwischen 20 und 60 Jahren freiwillig teil. Bis auf die technische Ausstattung mit eigenem PC, Internetzugang und Kopfhörern gab es keine weiteren Teilnahmevoraussetzungen. Von den 70 Probanden waren

47 weiblich und 23 männlich. Zum Zeitpunkt der Teilnahme befanden sich 42 noch in der Ausbildung (Studium/berufliche Ausbildung), 28 waren berufstätig. Weiterhin sind vier Probanden nicht-deutsche[12] und 66 deutsche Muttersprachler, von denen der Großteil (43) norddeutscher Herkunft ist. Keiner der Probanden hat sich in der Selbstauskunft als Synästhet bezeichnet, 20 gaben „Ich bin nicht sicher“ an, was vermutlich eher auf eine begriffliche als auf eine faktische Unsicherheit zurückzuführen ist. Eine Ausbildung auf einem Musikinstrument weisen 37 von 70 Probanden auf, Gesangsunterricht haben oder hatten hingegen nur sieben Probanden. Anderweitige regelmäßige Gesangserfahrungen (z.B. Singen im Chor) gaben 27 Probanden an.

Die Altersverteilung ist aufgrund der Art und Weise der Probandenakquisition bimodal mit einem ersten Maximum zwischen 20 und 24 und einem zweiten zwischen für 50 Jahre und älter, also an den jeweiligen Rändern der Altersspanne.

Die Stichprobe zeichnet sich durch große Heterogenität aus, was ihre Repräsentativität potentiell erhöht. Dennoch hat die Stichprobe natürlich strukturelle Eigenheiten, die beispielsweise von einer zugrunde gelegten Population deutscher erwachsener Muttersprachler abweichen. Mindestens die Variablen Geschlecht und musikalische Vorbildung sowie Alter aufgrund ihrer bimodalen Verteilung müssen daher als Zwischensubjektfaktoren getestet werden.

7.2 Stimuli

Aus den Mikrofonaufnahmen wurden für jeden Sprecher drei Stimulustypen (ST1 bis ST3) erstellt, wobei ST1 und ST2 Spontansprache und ST3 Lesesprache repräsentieren.

Der Stimulustyp 1 ist eine Abwandlung des Stimulustyp 2 und seine Beschreibung wird daher hintenan gestellt. Stimulustyp 2 besteht aus je vier spontansprachlichen Sätzen oder Phrasenverbindungen mit jeweils zweimal final fallender und zweimal final steigender Intonation, die abwechselnd hintereinander geschnitten wurden, beginnend mit steigender Intonation. An Anfang und Ende der Stimuli sowie zwischen den Sätzen wurden jeweils 0,5 Sekunden Stille eingefügt. Zweck der intonatorischen Einschränkung war, im ST2 jeden Sprecher potentiell mit seinem spontansprachlich habituellen F0-Umfang zu repräsentieren. Anhang B enthält die Transkripte der ausgewählten Phrasen.

Für die Lesesprache wurden drei Sätze aus dem Prosatext ausgewählt, bei denen die Sprechenden jeweils unterschiedliche Rollen einnehmen mussten:

> Satz 1: Erzähler: „Kossonossow kam nach Haus und begab sich gleich am Tag seiner Ankunft zum Dorfsowjet.“
>
> Satz 2: Nebencharakter: „Sie, Genosse, müssen etwas volkstümlicher sprechen, bitte, dass Sie die Masse auch versteht.“

[12] Zwei Probanden mit russischer, und je einer mit ukrainischer und arabischer Muttersprache, die jedoch seit mehreren Jahren in Deutschland leben. Da es sich um so eine geringe Zahl handelt und sich das Antwortverhalten nicht signifikant von den deutschen Muttersprachlern unterschied, wurde davon abgesehen, die Probanden aus der Stichprobe auszuschließen.

> Satz 3: Hauptcharakter: „Da ist einmal eine Kuh bei uns in den Propeller gekommen!“

Dabei nimmt die Expressivität zu. Satz 3 ist gleichfalls einer der intonatorischen Höhepunkte des Textes, was in den Aufnahmen der Sprecher mit dem höchsten F0-Gipfel einherging.

Für den Stimulustyp 3 wurden die oben genannten Sätze hintereinander geschnitten. Zwischen den einzelnen Sätzen sowie am absoluten Anfang und Ende des Stimulus wurden wiederum je 0,5 Sekunden Stille eingefügt.

Ebenso wie Stimulustyp 3 wird der Typ 1 im ersten Teil des Perzeptionsexperimentes zur Abfrage der Erkennbarkeit von ausgebildeten Stimmen dienen. Bei Stimulustyp 3 stehen dem Hörer dabei sowohl stimmqualitative als auch prosodische und artikulatorische Cues zur Verfügung. Der Fokus der Arbeit liegt jedoch auf stimmqualitativen Parametern, sodass es gilt, die anderen Faktoren weitgehend auszuschließen bzw. deren Einfluss einschätzen zu können. Die Abfrage anhand der Lesesprache dient also wie bereits erwähnt als Vergleichsbasis zur besseren Beurteilung des Einflusses der Stimmqualität. Für den Stimulustyp 1 galt es daher, eine Möglichkeit zu finden, die Aufmerksamkeit des Hörers während der Beurteilung auf den globalen Stimmklang zu konzentrieren. Eine naheliegende und in der Forschung gängige Methode ist der Einsatz von gehaltenen Vokalen als Stimuli. (Vgl. de Krom 1994: 40) Da die Idee, sowie die Fragestellung und Hypothesen der Arbeit sich jedoch auf den natürlichen Gebrauch von Sprache beziehen, scheint diese Methode in der vorliegenden Arbeit wenig aussagekräftige Ergebnisse zu ermöglichen. Daher wurde die Spontansprache aus ST2 als Ausgangsbasis genutzt und in Anlehnung an Scherer (1972: 196) weiter verarbeitet. Hierfür wurden jeweils eine final steigende und eine final fallende Phrase aus den ST3 ausgewählt und in Abhängigkeit zur Sprechgeschwindigkeit der Sprecher mithilfe eines Praatscripts in 0,6 bis 0,8 Sekunden lange Abschnitte zerteilt, die in zufälliger Reihenfolge wieder zusammengesetzt wurden. Dabei wurden die Nulldurchgänge berücksichtigt, um Klickgeräusche an den Schnittstellen zu vermeiden. Weiterhin wurden deutliche Phrasierungspausen vorher entfernt, um einen kontinuierlichen Sprachstrom zu erzeugen. Die Ergebnisse der Zerstückelung wurden auditiv auf Störgeräusche durch ungünstige Schnittstellen (beispielsweise durch Konsonantenclusterung) überprüft und gegebenenfalls neu zusammengefügt. Am Ende wurden wiederum 0,5 Sekunden Stille an Anfang und Ende der Stimuli eingefügt. Die Länge der Signalstücke wurde nach auditiver Beurteilung und Absprache mit einem Testhörer so festgelegt, dass genügend zusammenhängender Sprachschall vorhanden bleibt, um einen Eindruck von der Stimme des Sprechers zu bekommen, jedoch Inhalt und Intonationsverläufe der Phrasen nicht mehr erkennbar sind und dadurch nicht vom globalen Stimmklang ablenken. Natürlich ist aber jeder Hörer in seinem Hörverhalten individuell und die hier gewählte Methode bleibt diskussionswürdig und hat ebenso wie die gehaltenen Vokale gewisse Nachteile. Es lässt sich definitiv nicht von einer Beseitigung der prosodischen und artikulatorischen sowie inhaltlichen Einflussfaktoren sprechen, aber zumindest von einer starken Dämpfung.

Es wurden 63 Stimuli (21 Sprecher x 3 Stimulustypen) erstellt, wobei die Stimuli vom Typ 2 sowohl in Teil 2 als auch in Teil 3 des Experiments zu hören waren. Zusätzlich wurden vier „Dummy“-Stimuli von drei weiteren Sprechern erstellt – zwei für den

Experimentteil 1A und je einer für die restlichen Teile. Jeder Proband hatte insgesamt also 89 Stimuli zu beurteilen.

7.3 Experimentablauf

Für die Abwicklung des Perzeptionsexperiments wurde ein selbstständiges Programm geschrieben, welches die Probanden nach einer kurzen Installations-routine lokal auf ihrem PC laufen ließen. Das Installationspaket stand online als Download für einen Zeitraum von zweieinhalb Wochen zur Verfügung. Die Probandenakquisition erfolgte über persönliche Kontakte und die Nutzung von sozialen Netzwerken. Dabei wurden die potentiellen Probanden über die technischen Details (Systemvoraussetzungen, Ausstattung mit Kopfhörern), den Umfang des Experiments (30 bis 40 Minuten) und das Thema (Bewertung von Stimmen) aufgeklärt. Als Teilnahmemotivation wurde ein Gutschein verlost.

Beim Starten des Programms wurden die Probanden nach dem Zufallsprinzip in zwei Gruppen eingeteilt, sodass sie im Teil 2 des Experiments entweder nur den Subteil A oder nur den Subteil B absolvierten. Diese Einteilung blieb für die Probanden unsichtbar. Zunächst bekamen sie die Möglichkeit, die Funktionsfähigkeit ihrer Kopfhörer zu testen und die Lautstärke individuell anzupassen. Weiterhin wurden sie dazu aufgefordert, Stimuli zu überspringen, wenn Sie der Meinung waren, dass sie den Sprecher kennen.

Im Teil 1A waren die Probanden dazu aufgefordert, auf einer 5-Punkt-Skala die gesangliche Ausbildung der Sprecher anhand der Stimuli vom Typ 1 (zerstückelte Spontansprache) einzuschätzen. Die Skala war an den Enden mit „keine Gesangsausbildung“ und „professionelle Gesangsausbildung“ etikettiert. Der Anleitungstext enthielt ausführliche Informationen dazu, dass Sprecher mit unterschiedlichem Stimmbildungslevel zu hören sein werden und wie die Endpunkte der Skala zu verstehen sind, sowie eine explizite Vorbereitung auf die zerstückelten Sprachsignale und die Aufforderung, sich allein auf den Stimmklang zu konzentrieren und intuitiv zu entscheiden. Die Stimuli durften einmal wiederholt werden.

Teil 1B enthielt dieselbe Bewertungsaufgabe, diesmal auf Grundlage von Lesesprache (ST3). Auch hier wurden die Hörer im Anleitungstext auf die Art des Stimulus vorbereitet. Zudem wurde suggeriert, es handle sich um ein neues Set an Sprechern. Die Aufnahmen durften wieder einmalig wiederholt werden.

Zu Beginn von Teil 2 wurde den Hörern eine gefälschte Trefferquote von rund 50% für den ersten Experimentteil angezeigt, um ihre bis dahin herausgebildete Bewertungsstrategie und ihre interne Kategorisierung der Sprecher in Frage zu stellen und damit das Risiko von Ausstrahlungseffekten auf den nachfolgenden Experimentteil zu verringern.

Die Probanden, die Teil 2A absolvierten, hatten die Aufgabe, den Stimmklang in den spontansprachlichen Stimuli (ST2) anhand von sechs bipolaren VA-Skalen zu bewerten. Das semantische Differential beinhaltete die Adjektivpaare

„dunkel“ – „hell“,
„rau“ – „weich“,
„kräftig“ – „schwach“,
„dumpf“ – „brillant“,

„unebenmäßig“ – „ebenmäßig“ und
„angespannt“ – „entspannt“,

die sowohl in Anlehnung an die Studie von Moos/Simmons/Simner/Smith (2013) als auch mithilfe der Kollokationsanalyse des Wortes <Stimme> im DWDS (http://www.dwds.de) ausgewählt wurden.[13]

Die Probanden, die Teil 2B durchliefen, sollten bewerten, wie schön sie Stimme fanden und konnten dies auf einer einzelnen VA-Skala mit den Polen „überhaupt nicht schön“ und „sehr schön“ angeben. Sowohl in Teil 2A als auch in Teil 2B konnten die Aufnahmen beliebig oft abgespielt werden. Den Positionen des Schiebereglers auf den VA-Skalen wurden die Werte 0 bis 100 zugeordnet. Die Mittelkategorie (50) kann dabei als „weder A noch B“ oder als „Eigenschaft konnte nicht beurteilt werden“ interpretiert werden.

Nach dem zweiten Experimentteil wurden die personenbezogenen Angaben der Probanden abgefragt. Dies beinhaltete die Generierung eines Kürzels, Angaben zu Alter, Geschlecht, Herkunft, Muttersprache und Beruf sowie Aussagen zur musikalischen Vorbildung (Gesangserfahrung allgemein, Gesangsunterricht speziell und Ausbildung auf einem Instrument). Zudem wurde eine Selbsteinschätzung zur synästhetischen Veranlagung abgefragt.

Teil 3 des Experiments beinhaltete die Abfrage von cross-modalen Korrespondenzen anhand von 16 Texturen, die aus Moos/Simmons/Simner/Smith (2013) übernommen wurden. In dieser Studie untersuchten die Autoren durch menschliche Stimmen evozierte Farb- und Texturassoziationen bei Synästheten, Phonetikern und naiven Hörern, die keine synästhetische Veranlagung berichteten, und konnten einige individuumsübergreifende Korrespondenzen zwischen bestimmten Stimmqualitäten und Textureigenschaften sowie Farbbereichen feststellen. Für ihre Auswertungen überführten sie die Texturauswahl durch Quantifizierung auf acht bipolaren Skalen (semantisches Differential) in einem separaten Experiment vom Nominalskalenniveau auf Intervallskalenniveau, wodurch Korrelationsanalysen mit akustischen Parametern ermöglicht wurden. Das Bildmaterial wurde für das Perzeptionsexperiment übernommen, um somit vergleichbare Ergebnisse zu ermöglichen.[14] Die 16 quadratischen Bitmaps wurden dabei in vier Reihen und vier Spalten angeordnet und konnten durch Anklicken nach Hören des Stimulus ausgewählt werden. Die Anordnung der Bilder war für jeden Probanden zufällig. Aufgabe war, dasjenige Bild auszuwählen, welches die gehörte Stimme am besten beschreibt. Nach Durchlaufen des dritten Experimentteils wurden die Daten online auf einen privaten Server übertragen.

Zur Vermeidung des Einflusses von Kontexteffekten wurden die Stimuli der 21 Sprecher für das gesamte Experiment in zwei feste Blöcke von 10 und 11 Sprechern unterteilt. In den einzelnen Experimentteilen blieb die Reihenfolge der Blöcke gleich. Innerhalb der Blöcke wurden die Stimuli jedoch randomisiert, sodass ein möglichst

[13] Ursprünglich waren sieben Adjektivpaare vorgesehen. Das Paar *ruhig – hektisch* wurde nach dem Probedurchlauf mit dem Testprobanden wegen Redundanz zu *angespannt – entspannt* ausgeschlossen.

[14] Im Rahmen der vorliegenden Arbeit kann leider nur ein kurzer Einblick in die Ergebnisse des dritten Experimentteils gegeben werden. Dieser erfolgt im Abschnitt 10.

großer Abstand zwischen zwei Stimuli desselben Sprechers lag, ohne eine feste Reihenfolge zu bestimmen. Den Blöcken vorangestellt waren jeweils die „Dummy"-Stimuli, wobei es vermieden wurde, zweimal denselben „Dummy"-Sprecher zu Beginn eines Experimentteils einzusetzen. Ziel dieser Strategie ist es, die Erinnerbarkeit bzw. Wiedererkennbarkeit der Sprecher möglichst gering zu halten und so die Beeinflussung durch vorhergehende Urteile zu vermeiden. Dies war insbesondere für den Teil 1B von Bedeutung. Alle getroffenen Vorkehrungen in den Anweisungstexten und der Experimentstruktur lassen eine ernstzunehmende Beeinflussung zwischen den einzelnen Experimentteilen ausschließen.

8. Messmethoden der akustischen Parameter

Wie in Abschnitt 2 dargelegt, stehen für die akustische Analyse diverse segmentelle, prosodische und stimmqualitative Parameter zur Auswahl. Da der Fokus der vorliegenden Arbeit auf der Perzeption der Stimmqualität liegt, wurden aus den prosodischen Parametern einzig durchschnittliche Grundfrequenz, die Standardabweichung der Grundfrequenz als Maß der Tonhöhenvariabilität und Grundfrequenzumfang ausgewählt. Die Entscheidung zur Miteinbeziehung dieser Parameter ist ebenso wie die Einbettung von Lesesprache in das Perzeptionsexperiment der Versuch, neben den Effekten der Stimmqualität erste Hinweise zum Einfluss der Prosodie auf Stimmvorlieben zu erhalten. Zudem ist die habituelle Stimmlage eines Sprechers, welche hier über das durchschnittliche F0 parametrisiert wird, eine wichtige Komponente des Stimmklangs. Eine umfangreichere prosodische Analyse, die beispielsweise auch auf Intensität bezogene Parameter mit einschließt, würde jedoch den Rahmen der Arbeit sprengen.

Zur Beschreibung der Stimmqualität im Hinblick auf die individuelle Stimmlippenfunktion dienen der Öffnungsgrad der Stimmlippen (*Open Quotient* = OQ) und *Jitter*. Somit kann zum einen die generelle Form und zum anderen die Regelmäßigkeit der Stimmlippenschwingungen beschrieben werden. Im Hinblick auf die Spektrumsanalyse wurden *LTF2* (Langzeitanalyse des 2. Formanten) und HNR ausgewählt. Im Folgenden werden die Messmethoden der einzelnen Parameter dargelegt.

8.1 Mittleres F0, F0-Standardabweichung und F0-Umfang

Da die akustischen Messungen mit dem Ziel geschehen sollten, Zusammenhänge zwischen Akustik und Perzeption zu ergründen, wurden die akustischen Messungen nur in dem Sprachmaterial vorgenommen, das auch im Perzeptionsexperiment zum Einsatz kam - konkret die gelesene Prosa und die Ausschnitte aus der Spontansprache. Dabei wurden alle oben aufgelisteten Parameter in beiden Sprachstilen erhoben.

Die F0-Analyse in Praat (Boersma/Weenink 2014) erwies sich als unzureichend, indem sich die Messwerte der F0-Maxima und -Minima als stark abhängig von der Angabe des zulässigen F0-Bereiches (*pitch floor* und *pitch ceiling*) zeigten und somit keine pauschalen Parametereinstellungen festgelegt werden konnten. Daher wurde der F0-Umfang durch auditive Identifikation der Grundfrequenzmaxima und -minima in den Stimulustypen 2 und 3 des Perzeptionsexperimentes und lokale Messung mithilfe von WaveSurfer (Sjölander/Beskow 2013) ermittelt. Dabei wurden geknarrte Bereiche

ausgeschlossen, da hier aufgrund der Aperiodizität keine zuverlässigen Aussagen zur Grundfrequenz getroffen werden können. Der Anteil an geknarrter Stimme ist für den Höreindruck und die Bewertung von Stimmen mit Sicherheit von großer Bedeutung, kann jedoch in der vorliegenden Arbeit aufgrund von fehlenden Messmethoden nur auf ohrenphonetisch-impressionistischer Grundlage in die Analyse mit einfließen.

Für die Messung von Median und arithmetischem Mittel der Grundfrequenz und der jeweiligen Standardabweichung als Maß der F0-Variabilität in ST2 und ST3 wurde ein Praatskript genutzt, dass die vorher bestimmten sprecherindividuellen Grundfrequenzbegrenzungen mit berücksichtigte. Das bedeutet jedoch, dass ebenso wie beim F0-Umfang Signalbereiche mit Knarrstimme gänzlich ausgeschlossen wurden.

8.2 HNR, *Jitter* und LTF2

Der HNR (*Harmonics-to-Noise-Ratio*) gibt das Energieverhältnis von additivem Rauschen zu harmonischen Frequenzbereichen im Spektrum an. In Praat wird der HNR direkt über die mittlere Autokorrelation in den Analysefunktionen zur Harmonizität ermittelt. Er ist definiert als HNR = -10*log(1 – *meanAC*)[15] und wird in dB angegeben. Die Messwerte und deren Interpetation sind dabei stark abhängig von der Vokalqualität (bzw. Filterfunktion des Ansatzrohres). Ein HNR von 20 dB beim Vokal [a:] kommt zustande durch HNR = 20 = -10*log(1-0,99) und ist so zu interpretieren, dass 1 % der spektralen Energie in Rauschanteilen liegt bzw. das Signal zu 99 % periodisch ist. (Vgl. Boersma 2004: 247 f.)

Für die Berechnung von *Jitter*-Werten existieren verschiedene Algorithmen. Grundlegend handelt es sich jedoch dabei um Maße der Mikroschwankungen im Grundfrequenzbereich, also von Dauerabweichungen von benachbarten Perioden. In der vorliegenden Analyse wird das relative Maß Jitter(ddp) ausgewertet, welches als die durchschnittliche absolute Differenz zwischen konsekutiven Längendifferenzen benachbarter Perioden geteilt durch die durchschnittliche Periodendauer eines Signalabschnittes definiert ist:

$$DDP = \frac{\sum_{i=2}^{N-1} \frac{|(T_{i+1} - T_i) - (T_i - T_{i-1})|}{N-2}}{\frac{\sum_{i=1}^{N} T_i}{N}}$$

Der Wertebereich des DDP liegt dabei zwischen 0 und 600. Es gilt, je geringer der Wert, desto gleichmäßiger die Stimmlippenschwingungen.

Da dieser Parameter abhängig von der F0-Analyse ist, zeigt sich auch hier eine gewisse Unzuverlässigkeit in der automatischen Analyse. Aus diesem Grund wurden hier ebenfalls die sprecherindividuellen *pitch floor*- und *pitch ceiling*-Werte eingesetzt, d.h., dass auch bei der Bestimmung der *Jitter*-Werte geknarrte Sprachabschnitte ausgeschieden sind. Da dieses Analyseverfahren für alle Sprecher gilt, bleibt die Vergleichbarkeit erhalten.

[15] Die Autokorrelation dient zur Ermittlung der Ähnlichkeit zwischen zwei benachbarten Perioden. Bei einer perfekten Periodizität ist die mittlere Autokorrelation in Praat = 1 (vgl. Boersma 2004: 247).

Der LTF2 (*long term formant 2*) gibt die Langzeitverteilung des zweiten Formanten an. Die Lage des zweiten Formanten lässt Rückschlüsse auf die Länge des Vokaltrakts zu und wird insbesondere in der forensischen Phonetik als aussagekräftiger Parameter für Sprecheridentität und Sprecheridentifikation angesehen. (Vgl. Nolan/Grigoras 2005: 143) Insbesondere die Studie von Moos/Simmons/Simner/Smith (2013) konnte Korrelationen zwischen LTF2 und cross-modalen Stimmassoziationen beschreiben. Sie zeigten beispielsweise einen Zusammenhang zwischen höheren LTF2-Werten und helleren Farbassoziationen sowohl bei Synästheten als auch bei nicht synästhetisch veranlagten Hörern.

Die drei stimmqualitativen Parameter HNR, Jitter und LTF2 wurden anhand derselben Datenbasis mithilfe der Analysetools in Praat über scriptbasierte Halbautomatisierung ausgewertet. Hierfür wurden vorher zu den Stimuli ST2 und ST3 wiederum mit WEBMAUS (BAS 2014) TextGrids erstellt, deren Segmentgrenzen praatscript-gestützt zunächst auf die Nulldurchgänge verschoben wurden, um die Erkennung von Perioden in geschnittenem Material zu verbessern. Hiernach wurden alle als Vokal gelabelten Segmente (inkl. Diphtonge) extrahiert und konkateniert. Dabei blieb die Segmentierung der Vokale in einem neuen TextGrid nachvollziehbar. Für jedes Vokalsegment wurde daraufhin HNR und Jitter(ddp) gemessen. Wie oben bereits erwähnt, wurden als Analyseparameter die sprecherindividuellen F0-Minima und -Maxima berücksichtigt. Weiterhin wurden das TextGrid-Label, die Anzahl der Perioden und der Anteil der stimmlosen Frames ausgegeben, um im Nachhinein für die Analyse ungeeignete Segmente ausschließen zu können. Segmente mit einem Anteil von stimmlosen Frames > 0 und mit einer Periodenzahl < 4 zeigten sich sowohl in der HNR als auch im *Jitter*-Wert durchgängig als Ausreißer und wurden daher pauschal für alle Sprecher aus der Analyse ausgeschlossen. Zur Überprüfung der Plausibilität der *Jitter*- und HNR-Werte wurden zusätzlich die Aufnahmen aus der Intonationsübung herangezogen und die Parameter nochmals für jeden Sprecher in den hinsichtlich F0, F1 und F2 stabilen Vokalabschnitten der Wortproduktionen [ja:], [zo:], [ni:], [du:] und [ne:] mit ebener Intonation gemessen.

Für die LTF2-Bestimmung wurden über das jeweils gesamte Signal konkatenierter Vokale an äquidistanten Messzeitpunkten (Schrittweite 6,25 ms) F1, F2 und F3 bestimmt und das aktuelle Segmentlabel herausgelesen. Der Parameter LTF2 ergibt sich danach als Mittel über alle plausiblen Messwerte von F2. Bei der Beurteilung der Plausibilität (= Ausschluss von Messfehlern) wurden zum einen Messzeitpunkte, an denen F1 größer als 1000 Hz oder F2 kleiner als 600 Hz war, pauschal ausgeschlossen. Zum anderen wurde die Plausibilität der F2-Werte manuell in Relation zur gelabelten Vokalqualität und zum gemessenen F1- und F3-Wert am selben Messzeitpunkt und den Formantwerten an den benachbarten Messzeitpunkten beurteilt. Die Lesesprache im Stimulustyp 3 enthält naturgemäß für alle Sprecher dieselben Vokale als Analysegrundlage. In den spontansprachlichen Daten besteht jedoch die Gefahr, dass die Vokalqualitäten über die einzelnen Sprecher nicht gleichverteilt sind und somit die Lage des LTF2 beeinflussen. Daher wurden für die Spontansprache nur die Instanzen des gespannten zentralen Vokals [a:] in die Analyse mit einbezogen.

8.3 *Open Quotient*

In der vorliegenden Studie wurde der *Open Quotient* nach Spontan- und Lesesprache getrennt als Mittelwert aus allen messbaren vokalischen Perioden im EGG-Signal berechnet. Für die Spontansprache wurden zunächst mit Praat skriptbasiert alle mit [iː], [eː], [aː], [oː] oder [uː] gelabelten Segmente aus den extrahierten spontansprachlichen Phrasen isoliert und alle Vokale unter 50 ms und mit aperiodischen Anteilen[16] herausgefiltert. Vor der Extrahierung wurden die EGG-Signale bei 40 Hz hochpassgefiltert, um tieffrequente Signalartefakte, die beispielsweise durch Bewegungen des Kehlkopfes entstehen, zu entfernen. Nach diesem Verfahren standen als Datengrundlage für die Spontansprache pro Sprecher im Durchschnitt 20 Vokale zur Verfügung.

Da die Lesesprache im Stimulustyp 3 nur aus jeweils drei Sätzen bestand und bei der Vokalisolierung nach dem oben beschriebenen Muster teilweise weniger als fünf Vokale pro Sprecher für die Analyse gewonnen werden konnten, wurde als Ausgangssignal jeweils die gesamte Prosaaufnahme gewählt. Dabei wurden wie bei der Spontansprache die EGG-Signale gefiltert und die Mindestdauer der Segmente aufgrund der jetzt großen Datenmenge auf 80 ms hochgesetzt. Dadurch standen für die Berechnung des OQ in der Lesesprache pro Sprecher durchschnittlich 89 Vokale zur Verfügung. Ein direkter Vergleich der OQs in Spontan- und Lesesprache ist also wegen der unterschiedlich großen Datengrundlage nicht unproblematisch, wird jedoch in der weiteren Analyse keine Rolle spielen.

Die isolierten Signalabschnitte wurden hiernach (nach Lese- und Spontansprache getrennt) wiederum skriptgestützt differenziert. Dabei markieren die jeweiligen Minima im differenzierten Signal den Beginn und die Maxima das Ende der Glottisöffnung (vgl. Baer/Löfqvist/McGarr 1983). Über die Abstandsmessung zwischen Minima und Maxima kann so die Dauer der Öffnungsphasen bestimmt werden, die zur jeweiligen Periodendauer ins Verhältnis gesetzt den *Open Quotient* ergibt. Anhang C enthält ein Messbeispiel.

9. Ergebnisse

9.1 Struktur der Analysen

Teil des Forschungsfragenkomplexes ist der Vergleich von Sprechern mit unterschiedlichem Stimmbildungslevel im Hinblick auf Unterschiede in den akustischen Eigenschaften und in der Wahrnehmung ihrer Stimmen durch andere. Dabei wurde bei der Probandenakquivisition von den drei Kategorien Gesangsunerfahrene (im Folgenden GU), Laiensänger (LS) und professionelle Sänger (PS) ausgegangen. Bei der Sichtung

[16] = *Percentage of invoiced frames* > 0 im *Praat Voice Report*. Diese Maßnahme war nötig, da die Vokalisolierung auf Grundlage der automatischen WEBMAUS-Segmentierung erfolgte, die an Lautgrenzen, insbesondere bei Spontansprache häufig sehr ungenau ist. Die manuelle Korrektur aller Segmentgrenzen war zeitlich zu aufwändig.

der Angaben der Sprecher bezüglich ihres Stimmbildungsunterrichts und der Gesangserfahrung zeigt sich jedoch bereits, dass die Aufrechterhaltung der Kategorisierung kritisch ist. Man beachte Tabelle 1:

Sprecher	*Stimmbildung (h)*	*Gesang (h)*
GUm01GZ	0	0
GUm03BT	0	0
GUm04FT	0	124,5
GUw01TH	0	120
GUw02IG	0	200
GUw04PB	0	96
GUm02CS	10[17]	0
GUw03SH	22,5[18]	0
LSw06MD	216	1200
LSw04SD	240	4320
LSm04HS	240	2040
LSm03TT	248	1368
LSm02HB	356	3640
LSw01SI	636	2136
LSw05KP	720	3636
LSm01CH	880	6876
PSw03LS	905	2800
PSw01NY	934	4576
PSw02KR	1588	6840
PSm02MM	1884	7480
PSm01JR	1959	7440

Tabelle 1. Umfang von Stimmbildung und anderweitiger Gesangserfahrung in Stunden (h); die Sprecherkürzel setzen sich aus Sprecherkategorie, Geschlecht, fortlaufender Nummer und Initialen zusammen.

Die Tabelle ist nach dem Umfang der Stimmbildung geordnet, was der im Vorhinein getroffenen Kategorisierung nicht widerspricht. Der Sprung von GU zu LS ist dabei als ausreichend groß nachvollziehbar. Die Grenze zwischen LS und PS wirkt jedoch willkürlich. Der Unterschied im Stimmbildungsumfang zwischen LSm01CH und PSw03LS ist marginal, was darauf zurückzuführen ist, dass in die Kategorie LS, wie sie oben definiert ist, ebenso Personen mit abgebrochenem oder sogar abgeschlossenem Musikstudium fallen, welche das Singen nicht zum Beruf gemacht haben.[19] Das Problem löst sich, wenn der Stimmbildungsumfang in Stunden in der inferenzstatistischen Auswertung als verhältnisskalierte unabhängige Variable genutzt wird. Die Kategorisierung soll jedoch nicht verworfen, sondern als Ausgangspunkt der deskriptiven Statistik parallel überprüft werden.

[17] 10 h logopädische Behandlung im Kindesalter

[18] 22,5 h Schauspielunterricht; die tatsächliche Stimmbildung oder Sprecherziehung würde noch geringer ausfallen.

[19] Dies ist zum Beispiel bei LSw05KP der Fall, die als Musiklehrerin arbeitet.

Es werden sowohl die akustischen Messungen auf Grundlage von Stimulustyp 2 (im Folgenden nur noch ST2) als spontansprachliche Datenbasis und Stimulustyp 3 (ST3) als Beispiel für Lesesprache ausgewertet. Die geschlechterspezifische Darstellung der Parameter erfolgt nur bei relevanten Unterschieden in den Messwerten von weiblichen und männlichen Sprechern.

9.2 Zusammenhang zwischen Stimmbildung und akustischen Eigenschaften der Stimme

Die im Bezug auf das Perzeptionsexperiment ausgewählten stimmqualitativen und prosodischen Parameter sollen zunächst auf Zusammenhänge mit dem Stimmbildungsumfang untersucht werden.[20] Da die Wertebereiche der Parameter abhängig von den jeweiligen Messmethoden sind, soll für die folgende Auswertung die Darstellung der Verhältnisse zwischen den Messwerten anhand von Grafiken genügen. Im Anhang C können die konkreten Messwerte aller Sprecher nachvollzogen werden.

9.2.1 F0-bezogene Parameter

Für die durchschnittliche Grundfrequenz sind keine relevanten Unterschiede zwischen den Sprecherkategorien bzw. kein Zusammenhang zwischen Stimmbildungsumfang und durchschnittlicher Grundfrequenz zu erwarten, da die habituelle Stimmlage eine sprecherindividuelle Größe ist. Diese Annahme wird durch die Korrelationsmatrix in Tabelle 2 bestätigt, deren Korrelationskoeffizienten als stichprobenbedingt zufällig entstanden interpretiert werden müssen (die Irrtumswahrscheinlichkeit beträgt überall mindestens 10%).

	Korr(F0 Mean, SB (h))	
Sprecher	*ST2*	*ST3*
weiblich	r = -0,516	r = -0,331
	p > 0,1	p > 0,3
männlich	r = 0,455	r = 0,339
	p > 0,1	p > 0,3
alle	r = -0,109	r = -0,028
	p > 0,6	p > 0,9

Tabelle 2. Korrelationskoeffizienten für den Zusammenhang zwischen durchschnittlicher Grundfrequenz (F0 Mean) und Stimmbildungsumfang (SB (h)), df = 19.

Anders verhält es sich mit der F0-Variabilität und dem F0-Umfang, die insbesondere bei der Lesesprache bei ausgebildeten Sprechern größere Werte ausprägen sollten. Die Abbildungen 3a und 3b zeigen den Vergleich der drei Sprecherkategorien. Da es sich bei der F0-Standardabweichung und dem F0-Umfang um Intervallangaben handelt und

[20] Alle prüfstatistischen Auswertungen basieren auf einem α-Niveau von 5%.

diese später auch zu perzeptorischen Größen ins Verhältnis gesetzt werden, sind die entsprechenden Parameter F0 SD und F0 Range in Halbtönen (st) angegeben. Zusätzlich wurden Differenzmaße der Parameter zwischen Spontansprache und Lesesprache berechnet und ebenfalls für die drei Sprecherkategorien dargestellt (s. Abb. 3c).

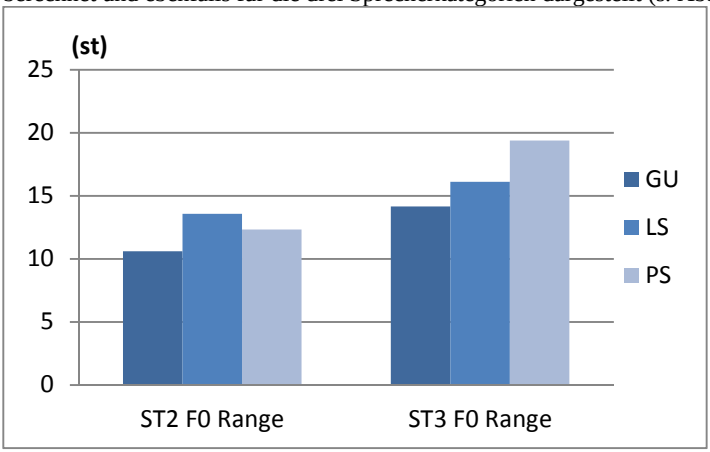


Abbildung 3a. F0-Umfang in Halbtönen nach Sprecherkategorien; links in ST2, rechts in ST3 gemessen.

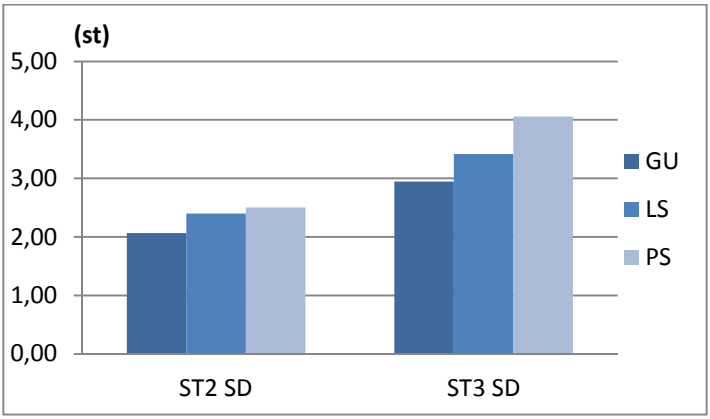


Abbildung 3b. F0-Standardabweichungen in Halbtönen nach Sprecherkategorien; links in ST2, rechts in ST3 gemessen.

Die F0-Umfang- und -Variabilitätsverhältnisse in der Spontansprache fallen zwar zugunsten der ausgebildeten Sprecher aus, sind jedoch statistisch nicht signifikant. Anders verhält es sich in der Lesesprache. Tabelle 3 zeigt die Korrelationskoeffizienten zwischen Stimmbildungsumfang und F0-Umfang bzw. F0-Standardabweichung für die

Spontansprache und die Lesesprache. In der Lesesprache ist dabei eine signifikante positive Korrelation von mittlerer Stärke zu beobachten.

	Korr(F0 Range, SB (h))		*Korr(F0 SD, SB (h))*	
Sprecher	*ST2*	*ST3*	*ST2*	*ST3*
alle	r = 0,313	***r* = 0,479**	r = 0,336	***r* = 0,547**
	$p > 0,1$	***p* = 0,028**	$p > 0,1$	***p* = 0,01**

Tabelle 3. Korrelationskoeffizienten für den Zusammenhang zwischen Stimmbildungsumfang (SB (h)) und durchschnittlichem F0-Umfang (F0 Range) in Halbtönen bzw. F0-Variabilität (F0 SD) in Halbtönen; df = 19.

Die Beobachtung der Differenzwerte der F0-Parameter (siehe Abb. 9c) zwischen Spontansprache und Lesesprache bestätigt ebenso den hypothetischen positiven Zusammenhang zwischen Stimmbildung und Stimmflexibilität bzw. expressiver Sprechweise bei Lesesprache.

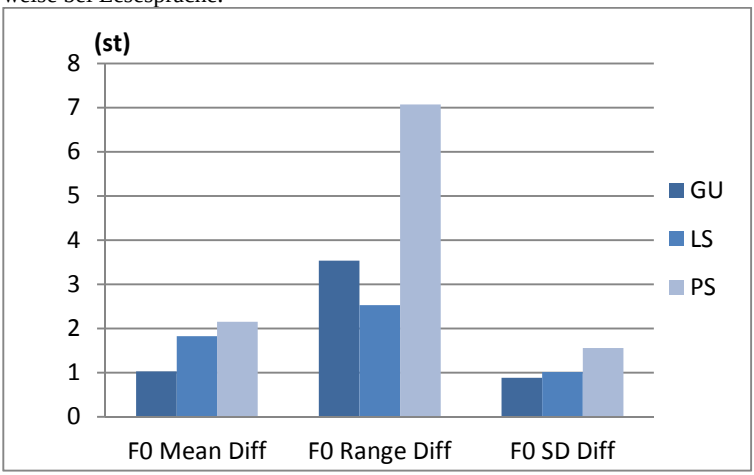


Abbildung 3c. Absolute Differenzen zwischen Spontansprache und Lesesprache in durchschnittlicher Grundfrequenz (F0 Mean Diff = |ST3 F0 Mean – ST2 F0 Mean|), F0-Umfang (F0 Range Diff = |ST3 F0 Range – ST2 F0 Range|) und F0-Variabilität (F0 SD Diff = |ST3 F0 SD – ST2 F0 SD|) in Halbtönen nach Sprecherkategorien.

Ein deutlicher Unterschied sowohl in Stimmlage als auch -umfang und -variabilität zeigt sich zwischen den unausgebildeten Sprechern (GU) und den professionell ausgebildeten Sprechern (PS), der allerdings aufgrund des Stichprobenumfangs nicht inferenzstatistisch belegt werden kann. Die Laiensänger reihen sich tendenziell auch an der hypothetisch richtigen Stelle ein, wenn man für die F0-Umfangsdifferenz zwischen Lese- und Spontansprache (F0 Range Diff) beachtet, dass die Sprechergruppe LS im Mittel bereits in der Spontansprache einen vergleichsweise hohen F0-Umfang aufweist

(siehe nochmals Abbildung 3a) und die Steigerung zur Lesesprache daher kleiner ausfällt.

9.2.2 Stimmqualitative Parameter

Wenn die in Abschnitt 4 umrissenen Trainingsziele für die Singstimme die Stimmlippenfunktion auch während des Sprechens beeinflussen, so sollten für die entsprechenden stimmqualitativen Parameter Unterschiede zwischen den ausgebildeten und nicht ausgebildeten Stimmen zu erkennen sein. Dabei sind für ausgebildete Sprecher kleinere *Jitter*-Werte und ebenso ein kleinerer HNR durch einerseits regelmäßige Stimmlippenschwingungen und andererseits ökonomische Stimmlippenadduktion (kein Luftverlust, der sich als Rauschen im Signal zeigt) während der Phonation zu erwarten. Letztere hat (neben anderen Muskelfunktionen) ebenso Einfluss auf den *Open Quotient* der Glottis, der bei ausgebildeten Sprechern nicht wesentlich vom Wertebereich modaler Stimmqualität abweichen sollte.

9.2.2.1 LTF2

In Kapitel 4 wurde die Tieflage des Kehlkopfes als Trainingsziel der Stimmbildung erwähnt, was zur Ausbildung des Sängerformanten beiträgt und sich zusätzlich im Absinken von F2 zeigen sollte, da sich das Ansatzrohr verlängert. Folgt man den Ausführungen Sundbergs (1987: 118), so ist dies jedoch auf das Singen beschränkt und würde daher in der akustischen Analyse der Sprechstimme keine Rolle spielen. Dies bestätigt sich zum einen im Vergleich der über die Sprecherkategorien gemittelten LTF2-Messwerte, deren Verhältnisse in Abbildung 4 zu erkennen sind.

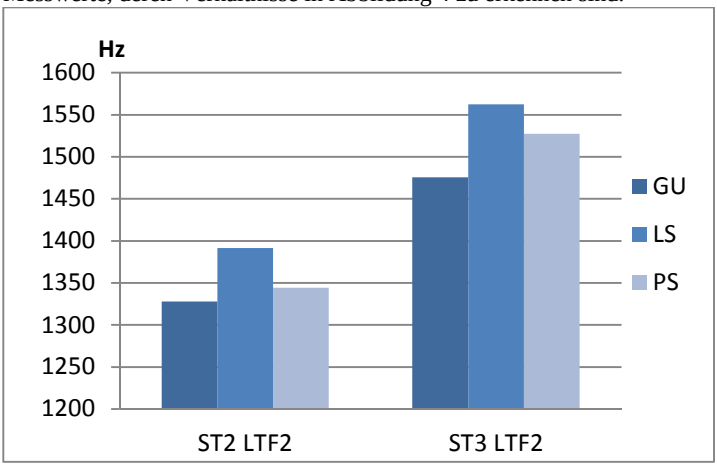


Abbildung 4. F2-Distribution nach Sprecherkategorie; links in ST2, rechts in ST3 gemessen.

Zum anderen zeigt die Korrelationsanalyse keinerlei Korrelation (ST2: r = 0,099; p = 0,573; df = 19; ST3: r = -0,13; p = 0,667; df= 19) zwischen dem Stimmbildungsumfang in Stunden (im Folgenden SB (h)) und dem LTF2-Wert eines Sprechers.

9.2.2.2 *Jitter* und HNR

Die in Kapitel 8 beschriebene Messmethode der Periodizitätsmaße *Jitter* und HNR stellte sich insbesondere für die Spontansprache erwartungsgemäß schwierig dar. Zum einen zeigt sich die HNR abhängig von der Vokalqualität, der Sprechervergleich muss also auf Basis derselben Vokalqualitäten erfolgen. Dies ist in ST2, der aufgrund des Einsatzes im Perzeptionsexperiment als Messgrundlage dienen sollte, nicht von Vornherein gegeben. Weiterhin sind die Vokale in kontinuierlicher und insbesondere spontaner Sprache stark von Koartikulation betroffen, die einerseits Rauschanteile durch die Friktion benachbarter Konsonanten erhöhen kann oder aber die stabilen Abschnitte in Vokalsegmenten für eine Analyse zu kurz werden lässt. Drittens unterliegt F0 starken Schwankungen in kontinuierlicher Sprache, was sich naturgemäß auf die *Jitter*-Werte auswirkt. Ein Vergleich der Messwerte für ST2 und ST3 mit einer zusätzlichen Messreihe aus den gehaltenen Vokalen der Intonationsübung (gekennzeichnet als INT) zeigt in der ST2-Messreihe eine hohe Anfälligkeit für widersprüchliche Ausreißer (siehe Anhang C). Daher wird die ST2 Messreihe aus der weiteren Analyse ausgeschlossen und durch die INT-Messreihe[21] ersetzt.

Abbildung 5zeigt die Verhältnisse der *Jitter*-Werte in den gehaltenen Vokalen (INT Jitter) und der Lesesprache (ST3 Jitter) zwischen den drei Sprecherkategorien.

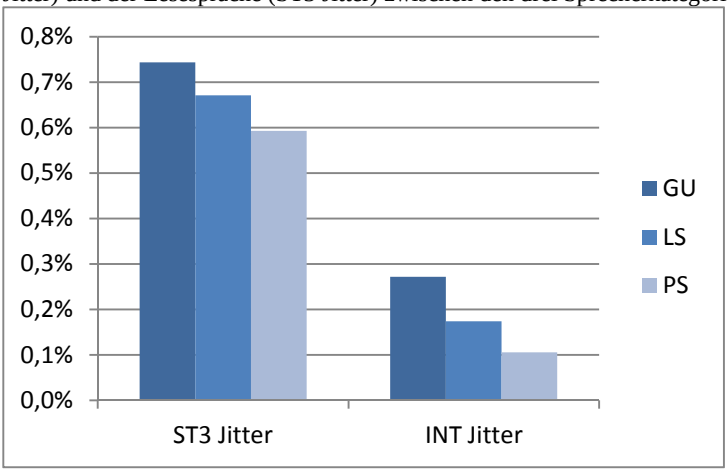


Abbildung 5. Jitterwerte nach Sprecherkategorien; links in ST3, rechts in gehaltenen Vokalen gemessen.

[21] Dies widerspricht dem eingangs erwähnten Prinzip, nur das Sprachmaterial des Perzeptionsexperiments für die akustische Analyse zu nutzen, ist aber methodisch notwendig, da sich die ST2-Messreihe als unzuverlässig darstellt.

Erwartungsgemäß nehmen die *Jitter*-Werte bei zunehmender Gesangsausbildung ab, die Unterschiede erweisen sich jedoch als nicht statistisch signifikant bei einem α-Niveau von 5% ($F_{(2)} = 3{,}19$; $p = 0{,}065$ für den Vergleich von INT Jitter der 3 Sprecherkategorien). Gleiches gilt für den hypothetischen Zusammenhang zwischen dem Stimmbildungsumfang in Stunden und den *Jitter*-Werten ($r = -0{,}422$; $p = 0{,}056$; df = 19 für INT Jitter).[22] Die Richtung des Zusammenhangs entspricht allerdings der Hypothese. Es ist also nicht auszuschließen, dass eine größere Stichprobe stabilere Ergebnisse bringen könnte.

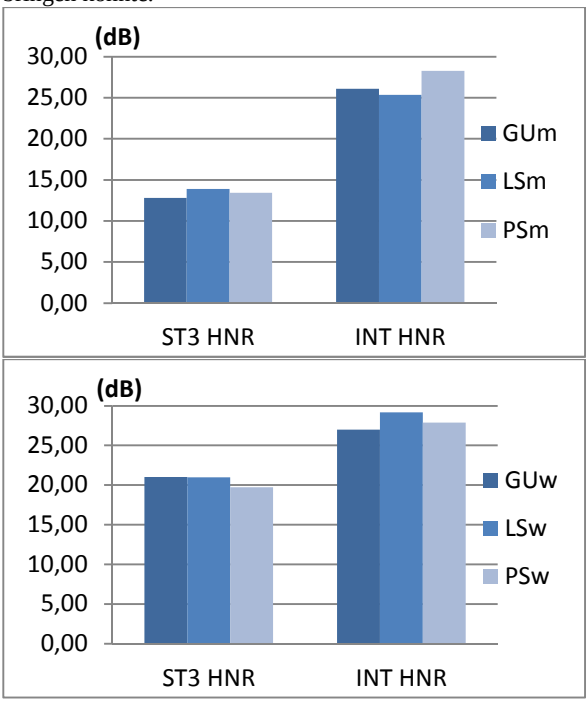


Abbildung 6. HNR-Werte für männliche (oben) und weibliche Sprecher (unten) nach Sprecherkategorie; links in ST3, rechts in gehaltenen Vokalen gemessen.

Die Analyse der HNR-Werte ergibt ebenfalls keine statistisch signifikanten Zusammenhänge (siehe Tabelle 4).[23] Bereits aus Abbildung 6 ist zu erkennen, dass sich die Werte nicht mit zunehmendem Stimmbildungsumfang erhöhen. In die vorliegenden

[22] Man beachte, dass die Irrtumswahrscheinlichkeiten nur knapp über 5% liegen. In den Messreihen aus ST3 ist keinerlei Tendenz zu erkennen: ANOVA(ST3 Jitter, SB (h)): $F(2) = 0{,}362$; $p = 0{,}701$; Korr(ST3 Jitter, SB(h)): $r = 0{,}082$; $p = 0{,}368$; df =19).

[23] Da männliche und weibliche Sprecher nicht in allen drei Kategorien gleich verteilt sind, sich jedoch die HNR-Werte für Männer und Frauen generell unterscheiden, müssen die Mittelwerte in den Kategorien hier geschlechtsspezifisch verglichen werden.

Messreihen sind daher keinerlei Tendenzen für einen positiven Zusammenhang von Stimmbildung und Harmonizitätswerten zu interpretieren.

Sprecher	*Korr(HNR, SB (h))*	
	ST3	*INT*
männlich	r = 0,082	r = 0,318
	p > 0,3	p > 0,1
weiblich	r = -0,225	r = -0,024
	p > 0,2	p > 0,3

Tabelle 4. Korrelationskoeffizienten für den Zusammenhang zwischen Stimmbildungsumfang (SB (h)) und HNR-Werten; df = 19.

9.2.2.3 OQ

In der Analyse des Open Quotients bietet sich ein komplexes Bild. Abbildung 7a stellt die Mittelwerte der Sprecherkategorien nach Geschlechtern getrennt dar. Bei den männlichen Sprecherkategorien lässt sich ein genereller Unterschied zwischen unausgebildeten und ausgebildeten Sprechern feststellen, die weiblichen Sprecherkategorien unterscheiden sich in den Werten kaum (OQ zwischen 0,57 und 0,64 für ST2 bzw. 0,54 und 0,59 für ST3). Die Auflösung nach einzelnen Sprechern (s. Abbildung 7b, hier anhand der spontansprachlichen Datenbasis (ST2)) gibt Aufschluss über das Zustandekommen der Mittelwerte.

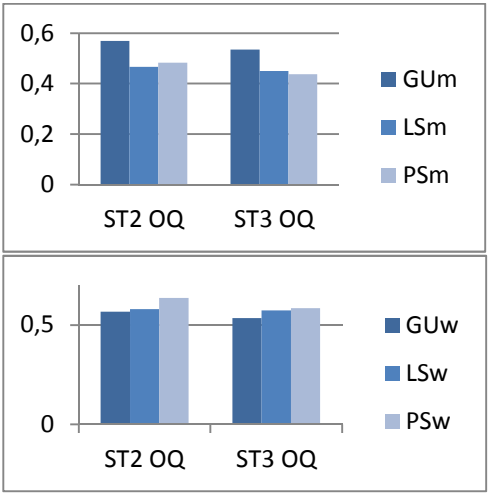


Abbildung 7a. Open Quotients für männliche (oben) und weibliche Sprecher (unten) nach Sprecherkategorie; links in ST2, rechts in ST3 gemessen.

Die teilweise große Streuung innerhalb der Kategorien zeigt, dass dieser Parameter als höchst sprecherindividuell einzuschätzen ist.[24] Einzig zu Gunsten der Hypothese B ließe sich das Fehlen von männlichen Sprechern mit hohem OQ (und damit der Tendenz zu behauchter Stimme bzw. zu schwacher Stimmlippenadduktion) bei vorhandener Stimmbildung interpretieren.

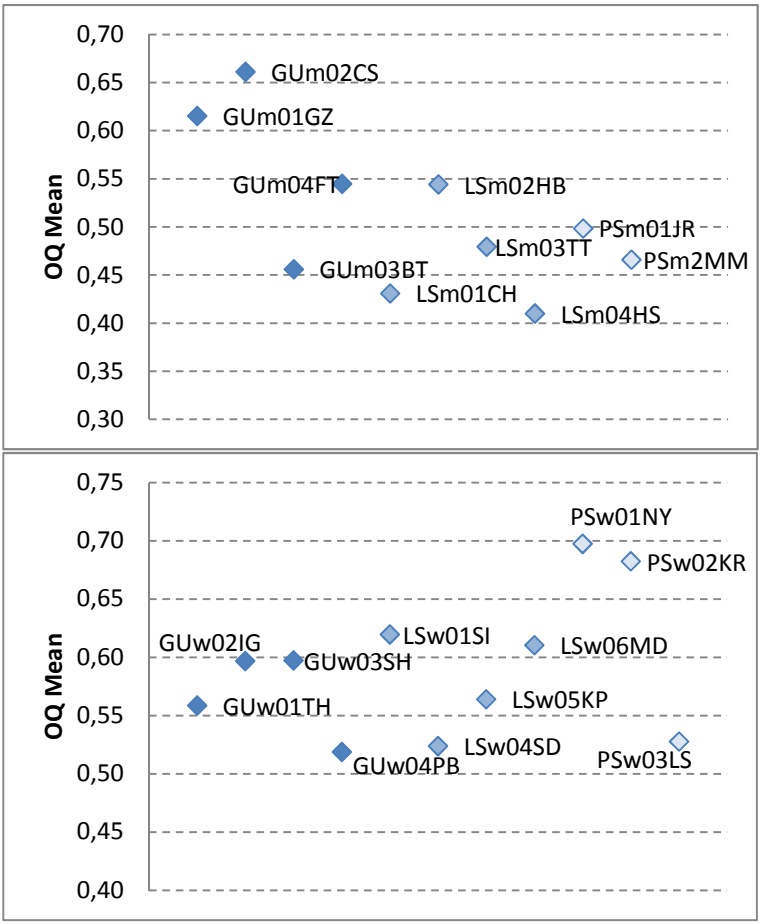


Abbildung 7b. Open Quotients für männliche (oben) und weibliche Sprecher (unten).

[24] Wenn die Werte der männlichen Sprecher GUm01GZ und GUm02CS und die der weiblichen Sprecherinnen PSw01NY und PSw02KR als Ausreißer angesehen werden, ist mit den verbleibenden Datenpunkten keinerlei Tendenz zu einem Zusammenhang zu erwarten. Daher wird hier von einer Korrelations- bzw. Regressionanalyse abgesehen.

9.2.3 Zusammenfassung und Vorhersagen zur Perzeption

Die deutlich größeren F0-Umfangs- und F0-Variabilitätswerte der professionellen Sänger und teilweise auch der Laiensänger in der Lesesprache bestätigen tendenziell die Vermutung eines positiven Einflusses von gesanglicher Ausbildung auf eine expressive Sprechweise in der Lesesprache.

Wenn die geschlechterspezifische Stimmattraktivität, wie sie z.B. bei Collins (2000) und Liu/Xu (2011) beschrieben wurde, auch im Hörerverhalten im vorliegenden Perzeptionsexperiment eine Rolle spielt, so spricht die LTF2-Verteilung gegen eine Bevorzugung ausgebildeter Stimmen, bzw. blieben die Stimmpräferenzen, wenn sie sich als abhängig von LTF2 zeigen, ebenso sprecherindividuell bedingt wie der LTF2-Wert selbst. Ähnliches gilt für den *Open Quotient*.

Die tendenziell kleineren *Jitter*-Werte für ausgebildete Stimmen sprechen zusammen mit der Hypothese, dass gleichmäßigere Stimmlippenschwingungen als schöner wahrgenommen werden, wiederum für eine Bevorzugung von ausgebildeten Stimmen.

9.3 Zusammenhänge zwischen Stimmbildung und Perzeption der Stimme

9.3.1 Stichprobenverteidigung

Vor dem Vergleich der Ergebnisse der Hörerbewertungen mit dem Stimmbildungsumfang und akustischen Eigenschaften der Sprecher wurde untersucht, ob die Hörerstichprobe aufgrund demographischer Variablen oder musikalischer Vorbildung unterteilt werden muss. Dabei wurde das Hörverhalten von Männern und Frauen, Personen mit und ohne musikalische Vorbildung und – nahegelegt durch die bimodale Altersverteilung – zwei Alterskategorien verglichen. Abbildung 8 verdeutlicht die Altersverteilung innerhalb der Stichprobe.

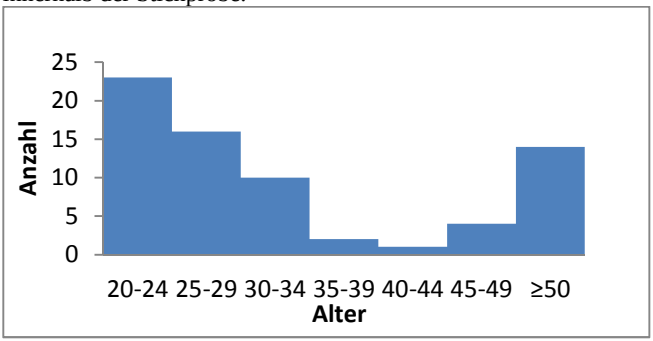


Abbildung 8. Altersverteilung in der Hörerstichprobe (N=70).

Die beiden Alterskategorien wurden auf 20 bis 34 (N_{A1} = 49) und 35 bis 60 (N_{A2} = 21) festgelegt. Die Überprüfung (zweiseitige t-Tests unabhängiger Stichproben) zeigte keine signifikanten Mittelwertunterschiede in den geforderten Bewertungen im Perzeptionsexperiment zwischen den oben genannten Stichprobenuntergruppen und damit keine Hinweise auf unterschiedliche Populationen. Die Hörerstichprobe kann damit

generell als Ganzes in den nachfolgenden Analysen überprüft werden. In einigen Fällen lohnt sich dennoch ein Blick auf Unterschiede zwischen männlichen und weiblichen Hörern.

9.3.2 Experimentteil 1 – Beurteilung der Stimmbildung der Sprecher

Im ersten Teil des Perzeptionsexperiments wurden die Probanden gebeten, das Stimmbildungslevel der Sprecher anhand von zerstückelter Spontansprache (ST1) und Lesesprache (ST3) zu schätzen. Zunächst sollen die Hörerurteile auf Basis dieser beiden Stimulustypen verglichen werden. Abbildung 9a zeigt die durchschnittlichen Werte für die einzelnen Sprecher.

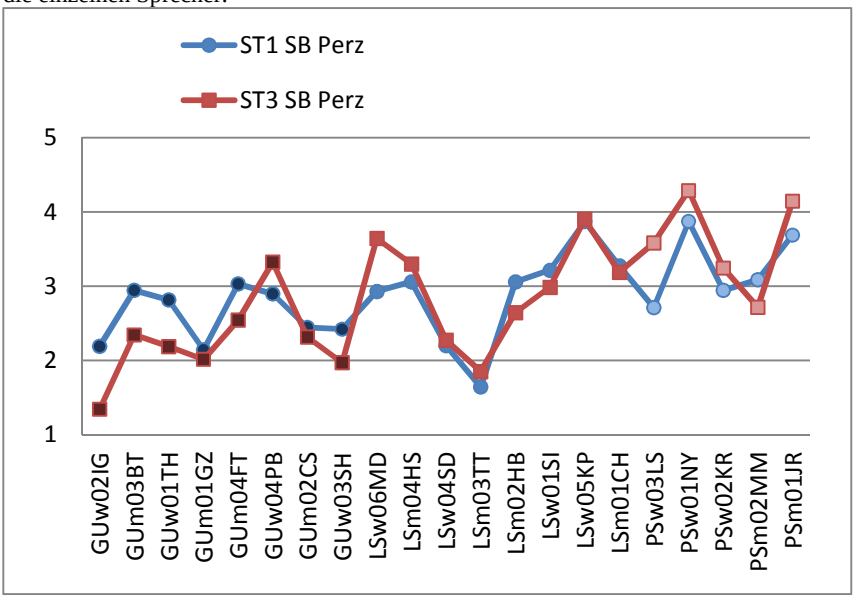


Abbildung 9a. Durchschnittliches geschätztes Stimmbildungslevel (SB Perz) der Sprecher für ST1 und ST3; Die Sprecher sind auf der x-Achse nach aufsteigendem Stimmbildungsumfang in Stunden angeordnet, die unterschiedliche Helligkeit der Punkte repräsentieren die drei Sprecherkategorien.

Die Ähnlichkeit der Hörerurteile für beide Stimulustypen ist an den annähernd gleich verlaufenden Kurvenverläufen abzulesen. Beide Kurven zeigen einen generellen Anstieg von links nach rechts. Die Kurve für ST3 ist durch zumeist extremere Abweichungen von der Mittelkategorie (3) gezeichnet. Es lässt sich hier bereits ein Unterschied zwischen unausgebildeten und professionellen Sprechern erkennen. Bei Ersteren sinkt das geschätzte Stimmbildungslevel von ST1 zu ST3 ab (mit einer Ausnahme), bei Letzteren steigt es an (wiederum mit einer Ausnahme). Die Laiensänger geben ein gemischteres Bild ab. Was in Abbildung 9a schon zu erkennen ist, verdeutlicht sich beim Vergleich der Mittelwerte für die drei Sprecherkategorien. Wie in Abbildung 9b deut-

lich wird, wurden Sprecher mit größerem Stimmbildungsumfang tendenziell erkannt. Die Überprüfung der Kategorienunterschiede in ANOVAs für die abhängige Variable SB Perz in ST1 und ST3 ergab zwar für die spontansprachlichen Stimuli keine Signifikanz ($F_{(2)}$ = 2,282; p = 0,131), für die Lesesprache konnten die Unterschiede jedoch auch inferenzstatistisch belegt werden ($F_{(2)}$ = 7,141; p = 0,0052).

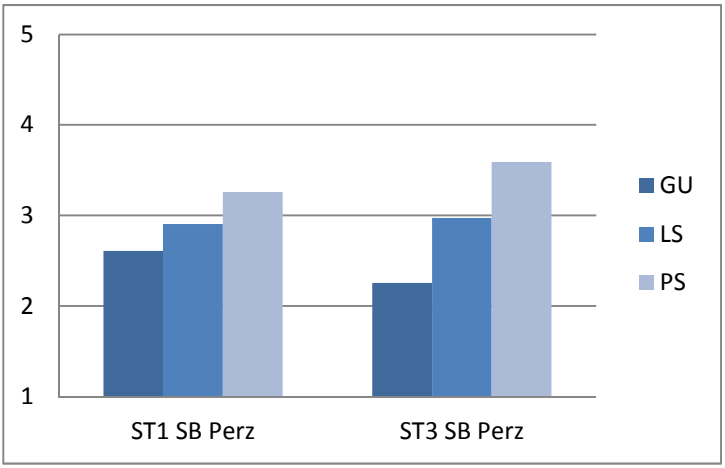


Abbildung 9b. Geschätztes Stimmbildungslevel (SB Perz) nach Sprecherkategorien; links Urteile für ST1, rechts für ST3.

Die Vorhersage der geschätzten Stimmbildung aus der tatsächlichen Stimmbildung, ausgedrückt durch die Variable SB (h), ergab sowohl für die Spontansprache als auch für die Lesesprache einen mittleren linearen Zusammenhang. Abbildung 10 zeigt die Ergebnisse der linearen Regression. Es scheint also bestätigt, dass die Hörer bei der Einschätzung des Stimmbildungslevels von bestimmten stimmqualitativen, prosodischen oder artikulatorischen *Cues*, die gehäuft bei Sprechern mit Stimmbildungshintergrund auftreten, geleitet wurden. Dabei ist allerdings zwischen realem und geschätztem Stimmbildungsumfang nur ein mittelbarer Zusammenhang zu unterstellen:

1. Der reale Stimmbildungsumfang zeigt einen Zusammenhang mit bestimmten Sprechereigenschaften bzw. einem Bündel A von akustischen Merkmalen einer Stimme, von denen einige im vorherigen Abschnitt gezeigt werden konnten (Aussagen über eine Kausalitätsrichtung werden hier vorsorglich nicht getätigt).
2. Die geschätzte Stimmbildung scheint mit einem Bündel B akustischer Merkmale von Stimmen zusammenzuhängen.
3. Es gibt eine Überschneidung zwischen den Merkmalsbündeln A und B, wodurch die geschätzte Stimmbildung teilweise aus der realen Stimmbildung vorhergesagt werden kann.

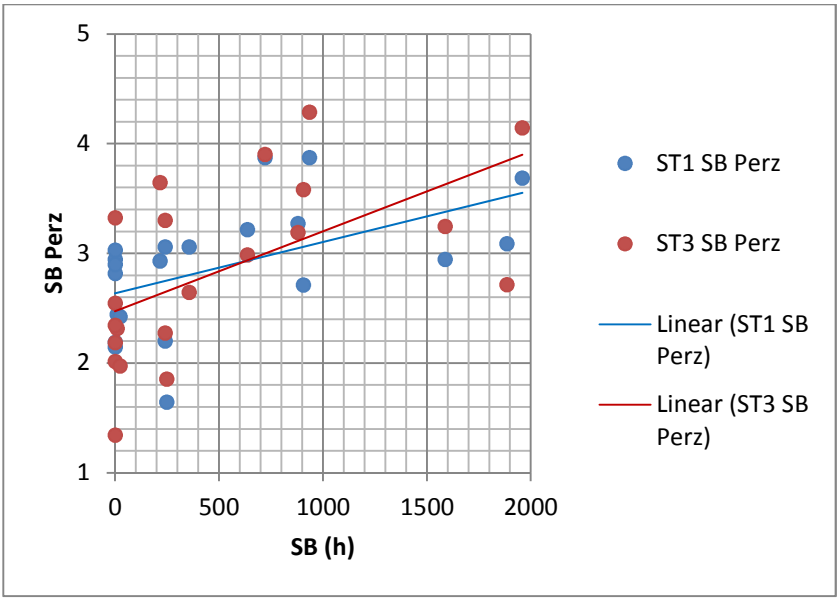


Abbildung 10. Lineare Regression mit Vorhersage von SB Perz aus SB (h): ST1: r^2_{adj} = 0,235; p = 0,015; df = 19; ST3: r^2_{adj} = 0,296; p = 0,006, df = 19.

Um welche Merkmale es sich dabei handelt, soll eine Korrelationsanalyse der akustischen Parameter mit dem geschätzten Stimmbildungslevel zeigen. Tabelle 5 und 6 enthalten die jeweiligen Korrelationskoeffizienten und deren Prüfwerte für männliche und weibliche Sprecher.

Wie erwartet, zeigen die Werte für die Lesesprache den Einfluss der prosodischen F0-Parameter auf die Hörerurteile. Ein größerer F0-Umfang und größere F0-Variabilität scheinen als Bewertungskriterien von den Hörern herangezogen worden zu sein.

Die stimmqualitativen Parameter bringen an dieser Stelle keine aussagekräftigen Ergebnisse. Vor allem für die Spontansprache scheinen für die Hörerentscheidung andere akustische Cues ausschlaggebend gewesen zu sein. In der Bewertung der Lesesprache ließen sich am ehesten die Korrelationswerte für Jitter der männlichen Sprecher als erwartungsgemäßen Einfluss auf die Beurteilung des Stimmbildungslevels interpretieren. Die einzige signifikante Korrelation innerhalb der stimmqualitativen Variablen zeigt sich hier im negativen Zusammenhang zwischen HNR und geschätztem Stimmbildungslevel bei den weiblichen Sprechern (sowohl für die Lesesprache als auch die Spontansprache). In Abschnitt 9.2 konnte allerdings kein Zusammenhang zwischen realem Stimmbildungslevel und HNR gezeigt werden. Auch die theoretische Fundierung würde allenfalls einen positiven Zusammenhang vorhersagen.

	Spontansprache		*Lesesprache*	
Variable	*Korr(Var x, ST1 SB Perz)*		*Korr(Var x, ST3 SB Perz)*	
ST2 F0 Mean	*r* = -0,08	*p* = 0,825		
ST3 F0 Mean			*r* = 0,068	*p* = 0,851
ST2 F0 SD (st)	*r* = 0,374	*p* = 0,286		
ST3 F0 SD (st)			***r* = 0,722**	***p* = 0,018**
ST2 F0 Range (st)	*r* = 0,133	*p* = 0,714		
ST3 F0 Range (st)			***r* = 0,665**	***p* = 0,036**
ST2 LTF2	*r* = -0,31	*p* = 0,259		
ST3 LTF2			*r* = 0,533	*p* = 0,113
INT HNR	*r* = 0,013	*p* = 0,393	*r* = 0,306	*p* = 0,147
ST3 HNR			*r* = 0,064	*p* = 0,378
INT Jitter	*r* = -0,233	*p* = 0,225	*r = -0,395*	*p = 0,072*
ST3 Jitter			*r = -0,334*	*p = 0,097*
ST2 OQ	*r* = -0,382	*p* = 0,257		
ST3 OQ			*r* = -0,485	*p* = 0,155

Tabelle 5. Ergebnisse der Korrelationsanalyse zwischen geschätzter Stimmbildung und akustischen Parametern für ST1 und ST3 der männlichen Sprecher (df =19). Da ST1 dasselbe akustische Material wie ST2 enthält, wurden die Messreihen aus ST2 für die Korrelationen genutzt.

	Spontansprache		*Lesesprache*	
Variable	*Korr(Var x, ST1 SB Perz)*		*Korr(Var x, ST3 SB Perz)*	
ST2 F0 Mean	*r* = -0,34	*p* = 0,306		
ST3 F0 Mean			*r* = -0,352	p = 0,288
ST2 F0 SD (st)	*r* = 0,299	*p* = 0,373		
ST3 F0 SD (st)			***r* = 0,792**	***p* = 0,0036**
ST2 F0 Range (st)	*r* = 0,309	*p* = 0,355		
ST3 F0 Range (st)			***r* = 0,884**	***p* = 0,0003**
ST2 LTF2	*r* = 0,509	*p* = 0,11		
ST3 LTF2			*r* = 0,23	*p* = 0,496
INT HNR	***r* = -0,598**	***p* = 0,005**	*r = -0,358*	*p = 0,099*
ST3 HNR			***r* = -0,631**	***p* = 0,002**
INT Jitter	*r* = 0,192	*p* = 0,27	*r* = -0,055	*p* = 0,382
ST3 Jitter			*r* = 0,038	*p* = 0,388
ST2 OQ	*r* = 0,421	*p* = 0,197		
ST3 OQ			*r* = 0,341	*p* = 0,304

Tabelle 6. Ergebnisse der Korrelationsanalyse zwischen geschätzter Stimmbildung und akustischen Parametern für ST1 und ST3 der weiblichen Sprecher (df = 19). Da ST1 dasselbe akustische Material wie ST2 enthält, wurden die Messreihen aus ST2 für die Korrelationen genutzt.

Zusätzlich zu diesen widersprüchlichen Ergebnissen zeigt die Korrelation zwischen dem geschätzten Stimmbildungsumfang aus ST1 und dem in ST3 gemessenen HNR einen noch stärkeren negativen Zusammenhang (r = -0,686; p = 0,0007). Natürlich kann aber die Harmonizität in der Lesesprache nicht bei der Beurteilung der vorher gehörten spontansprachlichen Stimuli als Cue fungiert haben. Die Werte können also nur darauf hindeuten, dass ein niedriger HNR-Wert als sekundärer Effekt einer (noch stärker bei Lesesprache) eingesetzten bestimmten Kehlkopf- und/oder Ansatzrohrkonfiguration aufgetreten ist, welche wiederum als Hinweis auf vorhandene Stimmbildung gedeutet wurde. Dafür spricht auch, dass dieser Zusammenhang nur bei den weiblichen Sprechern in Erscheinung tritt und die Werte der männlichen Sprecher eher auf einen umgekehrten (und damit erwartungsgemäß positiven) Zusammenhang hinweisen. Eine mögliche Erklärung für den Geschlechterunterschied wäre die für weibliche Stimmen typische längerer Öffnungsphase der Stimmlippen (höherer OQ), die mit behauchter Stimme und somit mit einer niedrigeren HNR einhergehen kann. (Vgl. Klatt 1990) Dass der *Open Quotient* von den Hörern geschlechtsspezifisch bewertet wird, zeigen die Zusammenhangstrends zwischen OQ und geschätzter Stimmbildung (s. Tab. 9.5 und Tab. 9.6). Während sich bei den männlichen Sprechern ein negativer Zusammenhang abzeichnet, also eine längere Öffnungsphase mit einem niedrigeren geschätzten Stimmbildungslevel zusammenfällt, tendieren die Korrelationskoeffizienten für die weiblichen Sprecher in die entgegengesetzte Richtung.

Der Vergleich von ST1 und ST3 legt die Vermutung nahe, dass bei den Probanden des Perzeptionsexperiments eine ähnliche Bewertungsstrategie genutzt wurde. Für die Lesesprache scheint diese Strategie noch durch die Nutzung prosodischer Cues ergänzt worden zu sein. Leider lieferten die hier ausgewählten stimmqualitativen Parameter und deren Messmethoden keine eindeutigen Hinweise auf Einflüsse auf das Hörverhalten. Die Korrelationsanalyse legt nahe, dass der Jitterwert mit in das Merkmalsbündel gehört, jedoch durch andere Cues überdeckt werden kann. Eine Annäherung daran, welches Merkmalsbündel tatsächlich das Hörverhalten beeinflusst hat, kann womöglich eine ausführliche akustische Analyse der Sprecher an den äußeren Rändern des Bewertungsspektrums liefern, was jedoch im Rahmen dieser Arbeit nicht geleistet werden kann. Die tendenziell richtige Beurteilung des Stimmbildungslevels auch in ST1 lässt eine weitergehende Untersuchung des Themas jedoch lohnenswert erscheinen.

9.3.3 Experimentteil 2 – Zuweisung von Klangeigenschaften und Wertung der Schönheit einer Stimme

In der Beschreibung von Klangeigenschaften und Bewertung der Schönheit einer Stimme im Experimentteil 2 legen die Daten nahe, die männlichen und weiblichen Sprecher zunächst getrennt zu betrachten. Für zwei der Adjektivpaare konnten durch t-Tests signifikante Mittelwertunterschiede zwischen männlichen und weiblichen Sprechern belegt werden. Es handelt sich um die Paare „dunkel-hell“ ($t_{(15)}$ = -4,711; p = 0,0003) und „dumpf-brillant“ ($t_{(19)}$ = -2,679; p = 0,015). Zwischen beiden Wortpaaren ist zudem eine gewisse semantische Ähnlichkeit vorhanden und tatsächlich zeigt die Korrelationsanalyse einen starken positiven Zusammenhang der Bewertungen auf den beiden Skalen (r = 0,824; p < 0,0001, df = 31). Da auch noch weitere Adjektivskalen miteinander korrelieren, wurde die Datenmenge für die weitere Analyse über eine Fak-

torenanalyse (Maximum-Likelihood-Faktorenanalyse mit Varimax-Rotation und Bestimmung der Faktorwerte nach der Bartlett-Methode), ebenfalls geschlechtsspezifisch, reduziert.

9.3.3.1 Männliche Sprecher

Die Faktorenanalyse weist auf die Reduzierung der sechs Indikatorvariablen auf drei latente Faktoren hin. Die Faktorenladungen sind in Tabelle 7 aufgelistet. Die Benennungen der Faktoren sind nur als Stellvertreter für die zugrundeliegenden Konzepte aufzufassen und erheben keinen Anspruch auf absolute begriffliche Präzision. Positive Werte im Faktor „Helligkeit" sind beispielsweise relativ zu den Hörerbewertungen auf der bipolaren Skala „dunkel-hell" zu verstehen, welche für alle männlichen Sprecher im Durchschnitt links von der Mittelkategorie lagen, d.h. die männlichen Stimmen wurden alle als mehr oder weniger dunkel bewertet.

	Faktor 1 ("Helligkeit")	*Faktor 2 ("Weichheit")*	*Faktor 3 ("Klangfülle")*
dunkel-hell	**0,996**	-	-
rauh-weich	-	**0,841**	-0,13
kräftig-schwach	0,235	0,149	**0,958**
dumpf-brillant	0,409	0,291	**0,862**
unebenmäßig-ebenmäßig	-0,5	**0,728**	-
angespannt-entspannt	0,129	**0,819**	0,177
Kommunalität	1,484	2,081	1,711

Tabelle 7. Faktorladungen der Adjektivskalen für 3 Faktoren (männliche Sprecher); Fettgedruckte Werte indizieren hohe Korrelationskoeffizienten einer Adjektivskala mit einem Faktor, ein Strich steht für einen Wert < 0,01.

Die weiteren statistischen Analysen geschehen auf Grundlage der ermittelten Faktoren. Tabelle 8 zeigt zunächst die Korrelationswerte der Faktorwerte mit den akustischen Parametern, die in ST2 (bzw. in den gehaltenen Vokalen (INT)) gemessen wurden.

Auch hier ermöglicht die Auswahl der akustischen Parameter noch keine Aussagen über eindeutige Zusammenhänge zwischen Klangeindrücken und akustischen Eigenschaften einer Stimme. Einzig der *Open Quotient* korreliert recht stark mit dem Faktor 3 („Klangfülle"), d.h. Stimmen mit hohem OQ wurden eher als schwach und dumpf eingeschätzt, Stimmen mit niedrigerem OQ als kräftig und brillant. Die mittlere positive Korrelation zwischen der F0-Standardabweichung und dem Faktor 2 („Weichheit"), der auch die Skala „angespannt – entspannt" mit einschließt, lässt sich womöglich noch

als entspannterer Höreindruck einer Stimme bei höherer F0-Variabilität interpretieren.[25] Weitere, annähernd signifikante Korrelationskoeffizienten lassen sich an dieser Stelle nicht sinnvoll deuten.

	Teil 2A					
Variable	*Korr(Var x, "Helligkeit")*		*Korr(Var x, "Weichheit")*		*Korr(Var x, "Klangfülle")*	
ST2 F0 Mean	*r* = 0,359	*p* = 0,531	*r* = 0,396	*p* = 0,442	*r* = -0,41	*p* = 0,568
ST2 F0 SD (st)	*r* = 0,422	*p* = 0,828	*r = 0,537*	*p = 0,064*	*r = 0,404*	*p = 0,073*
ST2 F0 Range (st)	*r* = 0,584	*p* = 0,707	*r* = 0,439	*p* = 0,137	*r = 0,235*	*p = 0,051*
ST2 LTF2	r = 0,469	*p* = 0,314	*r* = -0,406	*p* = 0,771	*r = 0,205*	*p = 0,087*
INT HNR	*r* = 0,297	*p* = 0,255	*r* = 0,3	*p* = 0,253	*r* = 0,147	*p* = 0,351
INT Jitter	*r* = -0,473	*p* = 0,124	*r* = -0,106	*p* = 0,368	*r* = -0,507	*p* = 0,101
ST2 OQ	*r* = 0,042	*p* = 0,517	*r* = 0,356	*p* = 0,644	***r* = -0,705**	***p* = 0,003**

Tabelle 8. Ergebnisse der Korrelationsanalyse zwischen den ermittelten Klangfaktoren und akustischen Parametern für ST2 der männlichen Sprecher (df = 19).

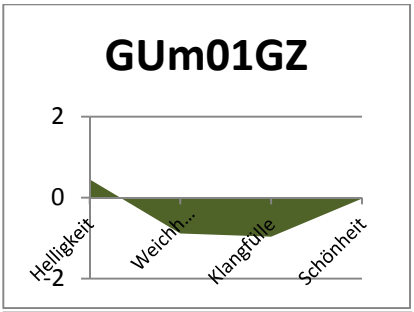
Interessant ist ein Blick auf die auf Grundlage der Hörerbewertungen entstandenen Sprecherprofile, die in Abbildung 11 durch Panoramen dargestellt sind. Mit eingebettet ist die Bewertung der Schönheit der Stimmen.[26] Im Vergleich der Sprecher lassen sich nur schwer Sprechercluster und eindeutige Sprecherprofile ausmachen. Allerdings sticht der Sprecher mit dem höchsten Score auf der Schönheitsskala auch in seinen Faktorenwerten heraus (PSm01JR).

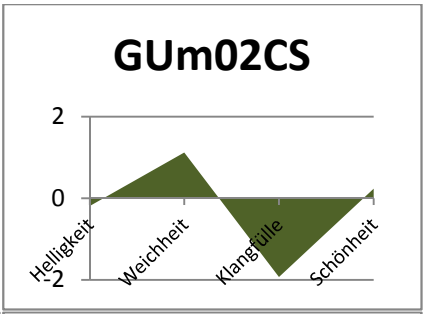
Besonders auffällig bei PSm01JR sind die hohen Werte für „Weichheit“ und „Klangfülle“. Wie bereits erwähnt sind positive Werte beim Faktor „Helligkeit“ für die männlichen Sprecher nicht gleichbedeutend mit einer hellen Stimme.[27] Wenn sich in der Korrelationsanalyse der Faktoren mit der Variable „Schönheit“ tatsächlich stabile Zusammenhänge zeigen, so kann PSm01JR innerhalb der Stichprobe als Sprecher mit prototypisch schöner Stimme und den damit verbundenen Klangeigenschaften verstanden werden - eine relativ dunkle, vor allem ebenmäßige, entspannte, weiche aber kräftige und brillante Stimme. Das Panorama des Sprecherprofils ohne Zusammenfassung der Adjektivskalen zu drei Faktoren kann zusammen mit denen der anderen Sprecher im Anhang C eingesehen werden.

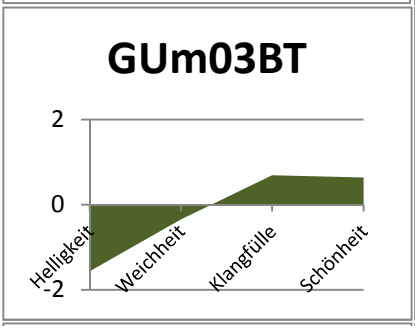
[25] Hier ist jedoch keinesfalls von einem monotonen Zusammenhang auszugehen, da eine hohe F0-Variabilität in Abhängigkeit von z.B. der Steilheit von F0-Anstiegen auch einen entgegen gesetzten, aufgeregten Höreindruck hervorrufen kann.

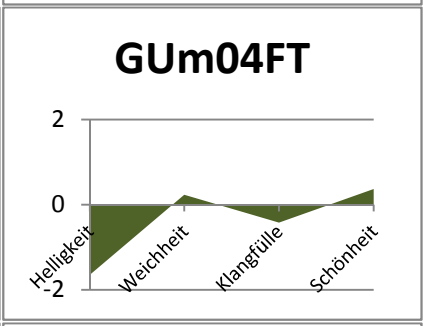
[26] Die ursprüngliche Skala mit den Werten 0 bis 100 („überhaupt nicht schön“ – „sehr schön“) wurde für den Zweck der Darstellung auf einen Wertebereich von -2 bis 2 transformiert, was in etwa dem Bereich der Faktorenwerte entspricht.

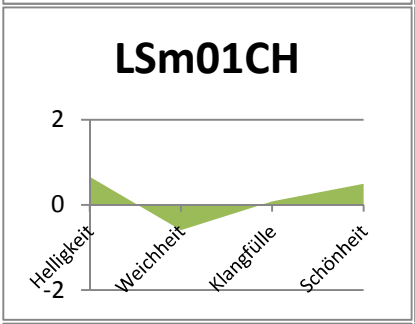
[27] Der durchschnittliche Score des Sprechers PSm01JR auf der „dunkel – hell“-Skala war 40,39.

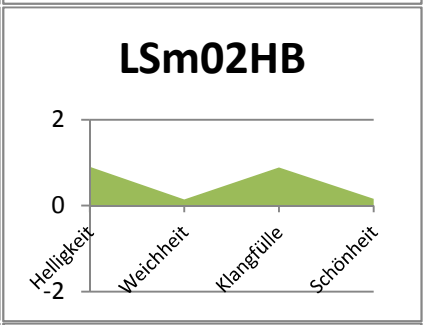

GUm01GZ
2
0
-2
Helligkeit
Weichh...
Klangfülle
Schönheit

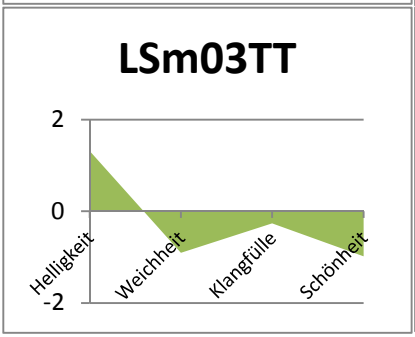

GUm02CS
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit

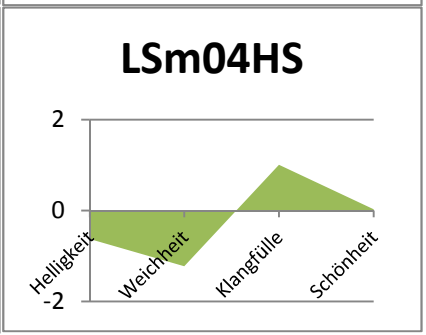

GUm03BT
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit


GUm04FT
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit


LSm01CH
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit


LSm02HB
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit


LSm03TT
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit


LSm04HS
2
0
-2
Helligkeit
Weichheit
Klangfülle
Schönheit

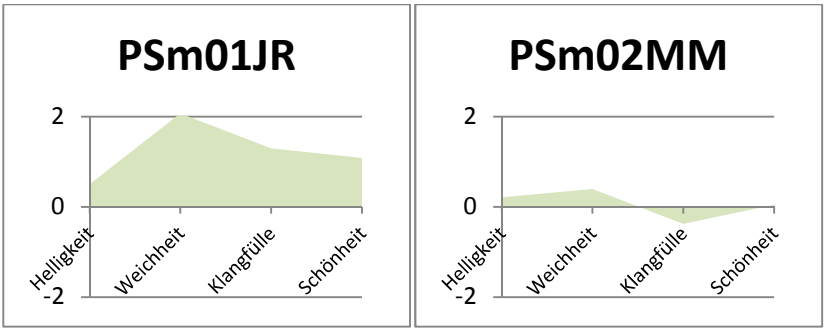


Abbildung 11. Panoramen zur Darstellung der männlichen Sprecherprofile nach wahrgenommenen Klangeigenschaften der Stimme inklusive der Bewertung der Schönheit der Stimme.

Über alle Hörer gemittelt scheint eine schöne Stimme tendenziell vor allem mit „Weichheit" einherzugehen – Faktor 2 zeigt eine mittlere positive Korrelation von r = 0,608 (p = 0,049; df = 19). Dem untergeordnet erscheinen die Eigenschaften „dunkel" (r = -0,402; p = 0,172; df = 19) und „brillant" bzw. „kräftig" (zusammengefasst im Faktor 3 mit r = 0,3523; p = 0,21; df = 19) begünstigend zu wirken. Bei der nochmaligen Betrachtung der Panoramen bestätigen sich diese Zusammenhänge ebenso beim Sprecher mit dem niedrigsten Score auf der „Schönheit"-Skala (LSm03TT), der sich durch einen hohen positiven Wert im Faktor 1 („Helligkeit") und einen negativen Wert sowohl in Faktor 2 („Weichheit") als auch in Faktor 3 („Klangfülle") auszeichnet.

Die Studien von Collins (2000) und Liu/Xu (2011) legen eine geschlechterspezifische Stimmbewertung insbesondere auf der Schönheitsskala nahe. Daher soll die Untersuchung der Zusammenhänge zwischen den Ergebnissen aus Teil 2B und 2A und den akustischen Parametern zusätzlich getrennt nach weiblichen und männlichen Hörern vorgenommen werden.

	männliche Hörer		*weibliche Hörer*	
	Korr(Var x, "Schönheit")		*Korr(Var x, ST1 "Schönheit")*	
ST2 F0 Mean	r = -0,029	p = 0,385	r = -0,041	p = 0383
ST2 F0 SD (st)	**r = 0,638**	**p = 0,036**	r = 0,317	p = 0,239
ST2 F0 Range (st)	r = 0,426	p = 0,157	r = 0,09	p = 0,372
ST2 LTF2	r = -0,226	p = 0,305	**r = -0,61**	**p = 0,048**
INT HNR	r = 0,284	p = 0,264	r = 0,144	p = 0,352
INT Jitter	r = -0,165	p = 0,341	r = -0,058	p = 0,381
ST2 OQ	r = -0,118	p = 0,362	r = 0,009	p = 0,386

Tabelle 9. Korrelationskoeffizienten für die Zusammenhänge zwischen der Variable „Schönheit" und den akustischen Parametern für männliche Sprecher getrennt nach männlichen und weiblichen Hörern (df = 19).

Aus den Korrelationskoeffizienten der Tabelle 9 lässt sich ein deutlicher Unterschied in der Reaktion auf akustische Cues zwischen männlichen und weiblichen Hörern ausmachen. Während die männlichen Hörer stärker auf den prosodischen Parameter der F0-Variabilität reagierten, zeigen weibliche Hörer eine Präferenz für männliche Stimmen mit niedrigem LTF2-Wert.

9.3.3.2 Weibliche Sprecher

In der Faktorenanalyse für die weiblichen Sprecher konnten ähnliche Variablengruppierungen wie bei den männlichen Sprechern gefunden werden. Jedoch sind die Skalen „dumpf – brillant“ und „rau – weich“ anders kombiniert (siehe Tabelle 10).

	Faktor 1 ("Brillanz")	*Faktor 2 ("Entspanntheit")*	*Faktor 3 ("Kraft")*
dunkel-hell	**0,973**	<0.01	0,212
rau-weich	**0,768**	*0,538*	0,209
kräftig-schwach	0,104	<0.01	**0,933**
dumpf-brillant	**0,933**	0,333	-0,121
unebenmäßig-ebenmäßig	0,19	**0,807**	<0.01
angespannt-entspannt	0,207	**0,973**	<0.01
Kommunalität	2,497	2,007	0,98

Tabelle 10. Faktorladungen der Adjektivskalen für 3 Faktoren (weibliche Sprecher); Fettgedruckte Werte indizieren hohe Korrelationskoeffizienten einer Adjektivskala mit einem Faktor.

Da die drei Faktoren hier etwas andere Faktorladungen der Adjektivskalen kombinieren als bei den männlichen Sprechern, wurden zur Unterscheidung neue Faktorenbezeichnungen gewählt. Beim Faktor 3 mit der Bezeichnung „Kraft“ ist zu beachten, dass hier kleinere Werte eine als kraftvoll wahrgenommene Stimme repräsentieren, da sie mit niedrigen Werten auf der bipolaren Skala „kräftig – schwach“ einhergehen.
In der Tabelle 11 sind die Ergebnisse der Korrelationsanalyse zwischen den drei Klangfaktoren und den akustischen Parametern aufgelistet.

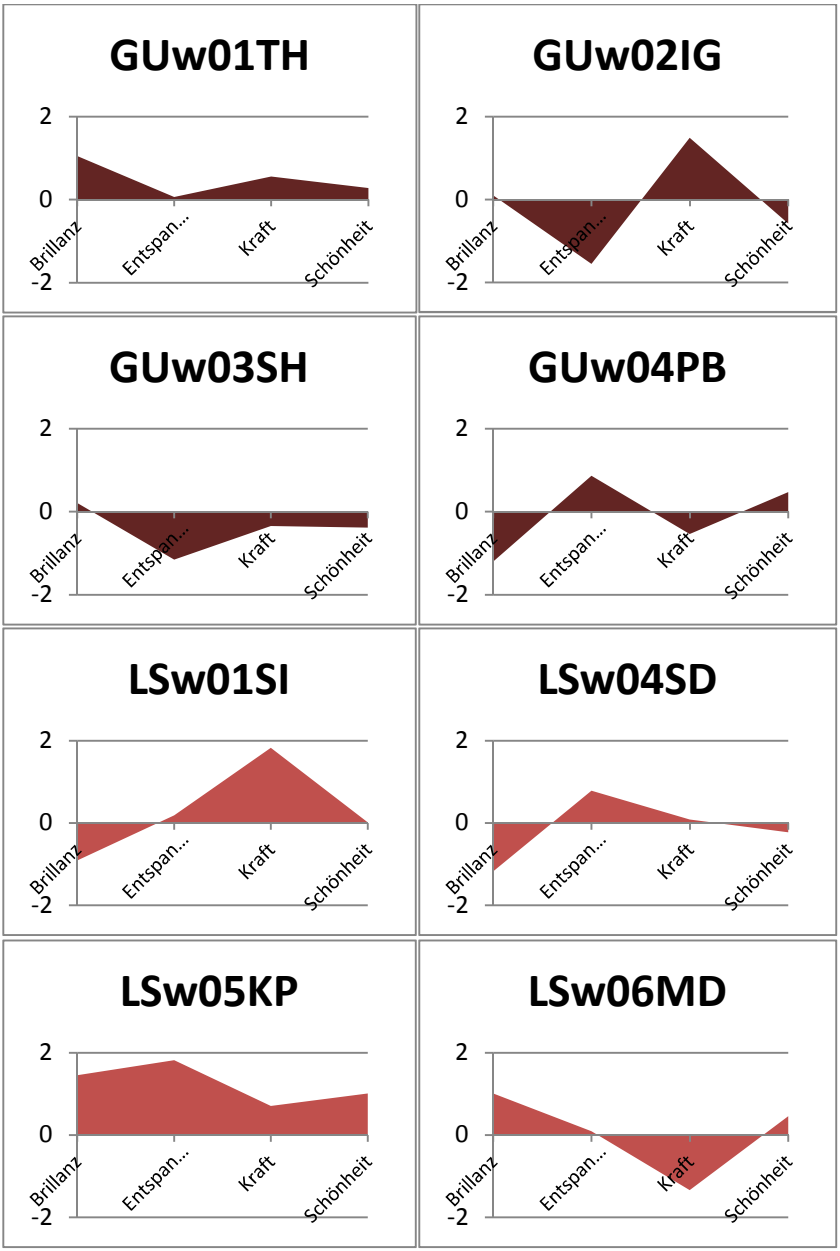
Bei den männlichen Sprechern konnte eine hohe Korrelation zwischen dem *Open Quotient* und der perzeptiven Klangfülle erkannt werden. Für die weiblichen Sprecher scheint der *Open Quotient* jedoch keinen Einfluss auf den Klangeindruck zu haben. Dies gilt zumindest für die im Perzeptionsexperiment gewählten Adjektivskalen. Der Faktor 1 („Brillanz“) korreliert stark positiv mit der durchschnittlichen Grundfrequenz, d.h. eine höhere Stimme wurde eher als hell und brillant eingestuft. Zudem besteht ein negativer Zusammenhang zwischen dem Faktor 3 („Kraft“) und der F0-Standardabweichung.

Variable	Teil 2A					
	Korr(Var x, "Brillanz")		Korr(Var x, "Entspanntheit")		Korr(Var x, "Kraft")	
ST2 F0 Mean	***r* = 0,729**	***p* = 0,011**	*r* = -0,429	*p* = 0,188	*r* = 0,239	*p* = 0,479
ST2 F0 SD (st)	*r* = 0,289	*p* = 0,387	*r* = 0,255	*p* = 0,449	***r* = -0,675**	***p* = 0,023**
ST2 F0 Range (st)	*r* = 0,232	*p* = 0,492	*r* = 0,348	*p* = 0,294	*r* = -0,269	*p* = 0,422
ST2 LTF2	*r* = 0,336	*p* = 0,312	*r* = 0,332	*p* = 0,318	*r* = 0,406	*p* = 0,215
INT HNR	*r* = 0,019	*p* = 0,387	*r* = 0,33	*p* = 0,218	*r* = 0,059	*p* = 0,381
INT Jitter	*r* = -0,018	*p* = 0,387	*r* = 0,187	*p* = 0,324	*r* = 0,197	*p* = 0,318
ST2 OQ	*r* = -0,069	*p* = 0,839	*r* = -0,142	*p* = 0,676	*r* = -0,15	*p* = 0,659

Tabelle 11. Ergebnisse der Korrelationsanalyse zwischen den ermittelten Klangfaktoren und akustischen Parametern für ST2 der weiblichen Sprecher (df = 19).

Abbildung 12 zeigt die Panoramen der Sprecherprofile für die weiblichen Sprecher. Hier sticht ebenfalls die Sprecherin mit dem höchsten Score auf der „Schönheit"-Skala (LSw05KP) durch die hohen positiven Werte in den drei Klangfaktoren heraus. Bei der Betrachtung der anderen Sprecherinnen, die auf der „Schönheit"-Skala am besten abgeschnitten haben (GUw01TH, GUw04PB, LSw06MD und PSw01NY), liegt die Vermutung nahe, dass vor allem positive Werte im Faktor 2 („Entspanntheit") aber auch im Faktor 1 („Brillanz") und untergeordnet ein negativer Wert im Faktor 3 („Kraft")[28] für die Beurteilung einer Stimme als „schön" begünstigend wirken. Die Korrelationsanalyse zwischen den Faktoren und den Wertungen auf der „Schönheit"-Skala können diese Vermutung bestätigen (s. Tabelle 12).

[28] Entspricht einem niedrigen Wert auf der Skala „kräftig – schwach".


GUw01TH
GUw02IG
GUw03SH
GUw04PB
LSw01SI
LSw04SD
LSw05KP
LSw06MD
2
0
-2
Brillanz
Entspan...
Kraft
Schönheit

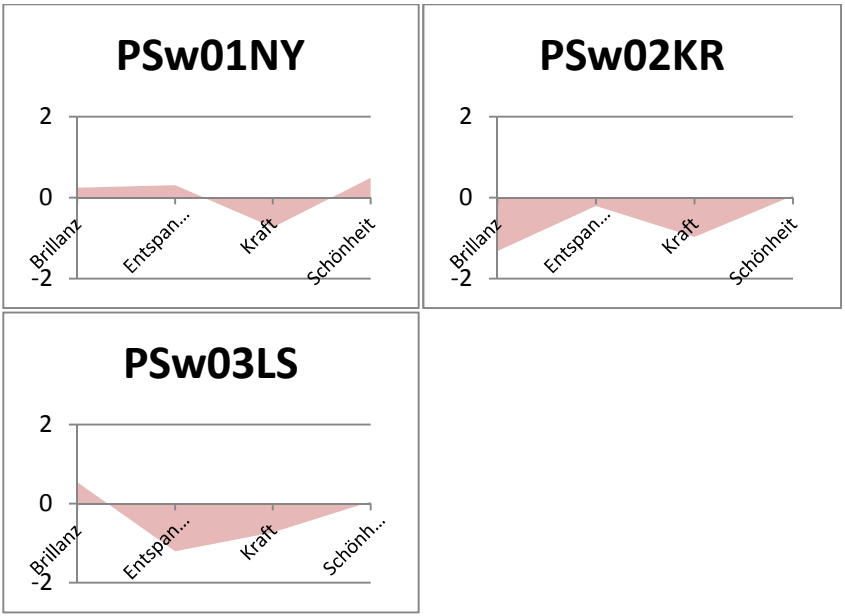


Abbildung 12. Panoramen zur Darstellung der weiblichen Sprecherprofile nach wahrgenommenen Klangeigenschaften der Stimme inklusive der Bewertung der Schönheit der Stimme.

Faktor	*Korr(Faktor x, "Schönheit")*	
"Brillanz"	*r* = 0,408	*p* = 0,155
"Entspanntheit"	**r = 0,78**	***p* = 0,0035**
"Kraft"	*r* = -0,248	*p* = 0,282

Tabelle 12. Korrelationskoeffizienten für weibliche Sprecher zwischen der Variable „Schönheit" und den Klangfaktoren (df = 19).

Beim Vergleich männlicher und weiblicher Hörer in der Beurteilung weiblicher Stimmen zeigt sich ein zu den männlichen Sprechern reziprokes Bild. In der obigen Tabelle 9 und der folgenden Tabelle 13 zeigen sich die Parameter F0-Standardabweichung oder LTF2 mit „Schönheit" korreliert, wobei LTF2 für die Kombination weibliche Sprecher/männliche Hörer nicht ganz das hier angesetzte Signifikanzniveau erreicht (in der Tabelle kursiv gedruckt). Nichtsdestotrotz ist zum einen interessant, dass der prosodische Parameter F0 SD größere Korrelationswerte bei Sprechern und Hörern gleichen Geschlechts und der stimmqualitative Parameter LTF2 größere Korrelationswerte bei Sprechern und Hörern unterschiedlichen Geschlechts auf-

weist und zum anderen, dass männliche Stimmen mit niedrigerem LTF2 von weiblichen Hörern als schöner bewertet wurden, weibliche Stimmen jedoch tendenziell einen höheren LTF2 aufweisen, wenn sie von männlichen Hörern als schön angesehen wurden.

	männliche Hörer		*weibliche Hörer*	
	Korr(Var x, "Schönheit")		*Korr(Var x, ST1 "Schönheit")*	
ST2 F0 Mean	r = -0,026	p = 0,386	r = -0,328	p = 0,231
ST2 F0 SD (st)	r = 0,159	p = 0,341	**r = 0,675**	**p = 0,024**
ST2 F0 Range (st)	r = 0,071	p = 0,378	r = 0,481	p = 0,118
ST2 LTF2	*r = 0,532*	*p = 0,074*	r = 0,2	p = 0,321
INT HNR	r = -0,423	p = 0,144	r = -0,474	p = 0,122
INT Jitter	r = 0,34	p = 0,209	r = 0,144	p = 0,351
ST2 OQ	r = -0,172	p = 0,333	r = 0,086	p = 0,373

Tabelle 13. Korrelationskoeffizienten für die Zusammenhänge zwischen der Variable „Schönheit" und den akustischen Parameter für weibliche Sprecher getrennt nach männlichen und weiblichen Hörern (df = 19).

9.3.4 Zusammenführung: Reale und geschätzte Stimmbildung im Verhältnis zu Klangeigenschaften und Schönheit der Stimme

Der Score auf der „Schönheit"-Skala zeigt einen starken positiven Zusammenhang zur geschätzten Stimmbildung (r = 0,831; p < 0.001). Da geschätzte Stimmbildung und reales Stimmbildungslevel ebenfalls leicht korrelieren, wie in Abschnitt 9.3.2 gezeigt, wäre hier ebenfalls eine Tendenz zum positiven Zusammenhang mit der „Schönheit"-Skala zu erwarten. Die Regressionsanalyse (analog zur Regression zwischen geschätzter Stimmbildung und realer Stimmbildung aus Abschnitt 9.3.2) zur Vorhersage der Schönheit einer Stimme aus dem realen Stimmbildungslevel (SB (h)) ergibt allerdings einen Determinationskoeffizienten von r^2_{adj} = 0,049 (p = 0,17). Hier ist demnach inferenzstatistisch kein Hinweis auf einen linearen Zusammenhang zu erkennen.

Die Verhältnisse zwischen geschätztem (und realem) Stimmbildungslevel und den Klangfaktoren sind wiederum über eine Korrelationsanalyse in Tabelle 14 dargestellt (hier aufgrund der Faktorenanalysen nochmals getrennt für weibliche und männliche Sprecher). Für die männlichen Sprecher zeigt sich hier ein signifikanter Zusammenhang zwischen dem geschätzten Stimmbildungslevel und großer „Klangfülle" (Faktor 3). Bei weiblichen Sprechern korreliert „Entspanntheit" (Faktor 2) mit dem geschätzten Stimmbildungslevel. Weitere, statistisch nicht signifikante, Korrelationsergebnisse können dennoch einige Tendenzen aufzeigen, die im folgenden Kapitel diskutiert werden sollen.

Sprecher	*Faktor*	*Korr(Faktor x, SB (h))*	*Korr(Faktor x, ST1 SB Perz)*
männlich	Faktor 1 "Helligkeit"	r = 0,354 p = 0,198	r = -0,271 p = 0,265
	Faktor 2 "Weichheit"	r = 0,542 p = 0,068	r = 0,482 p = 0,103
	Faktor 3 "Klangfülle"	r = 0,343 p = 0,207	**r = 0,598 p = 0,043**
weiblich	Faktor "Brillanz"	r = -0,187 p = 0,198	r = 0,314 p = 0,231
	Faktor 2 "Entspanntheit"	r = 0,091 p = 0,068	**r = 0,602 p = 0,041**
	Faktor 3 "Kraft"	r = -0,281 p = 0,207	r = -0,069 p = 0,379

Tabelle 14. Korrelationskoeffizienten zwischen Klangfaktoren und geschätztem bzw. realem Stimmbildungslevel.

10. Diskussion

10.1 Zusammenfassende Interpretation der Ergebnisse und Hypothesenprüfung

Die in Kapitel 5 aufgestellten Hypothesen lauteten wir folgt:

A) Eine wohlklingende Stimme ist mit bestimmten stimmqualitativen Eigenschaften verbunden.
B) Bevorzugt (sängerisch) ausgebildete Sprecher haben Stimmen mit solchen Eigenschaften.
C) Ihre Stimmen werden als wohlklingender empfunden als nicht ausgebildete Stimmen.
D) Es gibt hörerseitige Stimmstereotype für ausgebildete Stimmen.
E) Dieser Stereotyp überschneidet sich mit dem Stereotypen für wohlklingende Stimmen.

Für die Bestätigung von Hypothese A zeigen sich erste Hinweise in der Betrachtung der subjektiven Maße für die Beschreibung des Stimmklangs, die in Teil 2A des Perzeptionsexperiments untersucht wurden. Die Korrelationsanalysen zwischen Stimmattraktivität und den zugeschriebenen Klangeigenschaften konnten geschlechterspezifische Einflussfaktoren bestimmen. Eine männliche Stimme wird dabei bevorzugt als schön beurteilt, wenn sie dunkel und weich, aber dennoch kräftig erscheint. Eine schö-

ne weibliche Stimme ist vor allem brillant, ebenmäßig und entspannt. Auf akustischer Seite konnten hörerseitig geschlechtsspezifische Unterschiede festgestellt werden. Wenn man den LTF2-Wert als Korrelat zur Vokaltraktlänge annimmt, wobei ein tieferer Wert mit einem längeren Vokaltrakt einhergeht, so kann man aus den beschriebenen Zusammenhängen zwischen den LTF2-Werten der Sprecher und der Stimmattraktivität ableiten, das weibliche Hörer männlichen Sprechern mit längerem Vokaltrakt eine schöne Stimme zuweisen und andersherum männliche Hörer eher die weiblichen Sprecher mit kürzerem Vokaltrakt bevorzugten. Diese Ergebnisse stehen im Einklang mit den Studien von Collins (2000) und Liu/Xu (2011), die in Abschnitt 3.2 beschrieben wurden.

Bezüglich des Verhältnisses von Stimmbildung zur Stimmattraktivität können die formulierten Hypothesen sowohl sprecherindivuell als auch kategoriell (anhand der drei Sprecherkategorien GU, LS und PS) und anhand der kontinuierlichen Skala des realen Stimmbildungsumfangs überprüft werden. Für die sprecherindividuelle Perspektive lohnt sich nochmals ein genauerer Blick auf die Sprecher mit den extremsten Urteilen. PSm01JR und LSw01KP, die in der Stichprobe als Repräsentanten eines schönen Stimmstereotyps angesehen werden können, verfügen beide über einen hohen Stimmbildungsumfang (1959 Stunden für PSm01JR und 720 h für LSw05KP). Auch KP, die als Laiensängerin kategorisiert wurde, befindet sich damit in der oberen Hälfte des Kontinuums. Die weibliche Sprecherin mit der deutlich schlechtesten Bewertung verfügt hingegen über gar keine Stimmbildung (GUw02IG – 0 h), der männliche Sprecher (LSm03TT – 248 h) ist in seiner Sprecherkategorie am unteren Ende des Kontinuums von Stimmbildungsstunden (zwischen 216 und 880 h für die Kategorie LS). Interessant ist hier wieder der Zusammenhang zum wahrgenommenen Stimmbildungsumfang, da sowohl für PSm01JR als auch für LSw05KP sehr hohe Werte[29] erreicht wurden. Andersherum sind LSm03TT und GUw02IG auch die beiden Sprecher mit den geringsten Werten beim geschätzten Stimmbildungslevel (wiederum innerhalb der entsprechenden Geschlechter-kategorie). Daraus lässt sich der Gedanke ableiten, dass im Stimmstereotyp für eine ausgebildete Sängerstimme eine hohe Stimmattraktivität enthalten ist. Der Vergleich der vier Sprecher scheint die Bestätigung der Hypothesen B, C und E zu begünstigen.

Im Vergleich der Sprecherkategorien muss jedoch die Hypothese C bereits in Frage gestellt werden, da hier zwar die professionell ausgebildeten Sprecher tendenziell als schöner bewertet wurden, zwischen den Gesangsunerfahrenen und den Laiensängern jedoch kein Kategorienunterschied mehr feststellbar ist. Gleiches bestätigt die Unterstellung eines linearen Zusammenhangs zwischen Stimmbildungsumfang und Stimmattraktivität, der sich als quasi nicht vorhanden herausstellte (r^2_{adj} = 0,049). Daher liegt die Vermutung nahe, dass der Kategorienunterschied zwischen professionellen Sängern und den anderen Sprechern der Stichprobe als nicht repräsentativ zu betrachten ist.
Nach der akustischen Analyse der Sprecher und deren Betrachtung im Zusammenhang mit der Stimmbildung wurde hypothesengerecht davon ausgegangen, dass Sprecher mit

[29] PSm01JR wurde auf der SB Perz Skala sowohl für die spontansprachlichen Stimuli als auch für Lesesprache innerhalb der männlichen Sprecher am höchsten bewertet. LSw05KP erhielt innerhalb der weiblichen Sprecher für die Lesesprache die zweithöchsten, für die Spontansprache die höchsten Werte (zusammen mit PSw01NY).

höherem Stimmbildungsumfang als schöner bewertet werden, da sie tendenziell niedrigere *Jitter*-Werte und auf die männlichen Sprecher bezogen keine erhöhten *Open Quotients* zeigten. Da jedoch weder *Jitter* noch OQ im Perzeptionsexperiment für die Stimmattraktivität eine Rolle spielten, kann auch die Hypothese B zumindest im methodischen Rahmen dieser Untersuchung nicht bestätigt werden.

Anders sieht es für die Hypothesen D und E aus. Die hohe Korrelation zwischen geschätzter Stimmbildung und Stimmattraktivität lässt darauf schließen, dass für die Bewertung von Schönheit und für die Beurteilung des Stimmbildungslevels ähnliche Urteilsstrategien angewendet wurden. Dies wiederum legt die Vermutung nahe, dass hier entweder ähnliche Stimmstereotype oder sogar der gleiche Stimmstereotyp zugrunde gelegt wird, da möglicherweise davon ausgegangen wird, dass Personen, die gut singen können, auch eine schöne Sprechstimme haben.

Insofern scheint sich die Idee der gemeinsamen Stimmstereotype für attraktive Stimmen und ausgebildete Sprecher zu bewahrheiten. Die Hypothesen B und C, die das reale Stimmbildungslevel der Sprecher mit einbeziehen, lassen sich hingegen nicht bestätigen. Es zeigen sich allerdings auch keine gegenteiligen Trends, sodass die ursprüngliche Idee noch nicht verworfen werden muss, sondern mithilfe fokussierterer Experimentsettings neu untersucht werden sollte.

10.2 Methodenkritik

Im Perzeptionsexperiment war die größte Herausforderung, die Probanden auf den Stimmklang zu konzentrieren und für den Erfolg dieses Vorhabens kann hier nicht garantiert werden. Es ist beispielsweise nicht auszuschließen, dass die Adjektivskalen unbewusst oder bewusst nicht immer nur mit dem Stimmklang sondern mit vermuteten emotionalen Zuständen oder physischen Eigenschaften eines Sprechers in Verbindung gebracht wurden. Dies könnte z.B. bei der Skala „angespannt-entspannt“ oder „kräftig-schwach“ der Fall zu sein. Hier wäre für zukünftige Untersuchung eine Orientierung und damit Schulung der Hörer mit Ankerstimuli möglich. Weiterhin kann der Einfluss von Experimentteil 1 auf den Teil 2 nicht gänzlich ausgeschlossen werden, weswegen auch die Korrelationen zwischen geschätztem Stimmbildungslevel und vor allem Schönheit der Stimme mit großer Aufmerksamkeit zu interpretieren sind und eine nochmalige, real unabhängige Prüfung nötig ist.

Die spontansprachlichen Daten als Stimulusgrundlage für das Perzeptionsexperiment und auch für die akustischen Analysen zu nutzen, ist zwar in der Theorie sinnvoll, da der Untersuchungsgegenstand nicht zuletzt für die Erforschung von Stimmstereotypen die natürliche Sprache sein sollte. Auf praktischer Seite birgt dies jedoch einige Risiken, da die Anzahl zu kontrollierender Faktoren ungleich höher ist als beispielsweise bei isolierten Vokalen oder synthetisierter Sprache. Dies schlägt sich vor allem in den Schwierigkeiten der Messmethoden nieder. Die aktuelle phonetische Forschung zeichnet sich jedoch durch ständige methodische Reflektion und Innovation aus, sodass der Untersuchung von kontinuierlicher oder sogar spontaner Sprache alle Türen offen stehen und an dem zugrunde liegenden Grundsatz festgehalten werden sollte. Für weiterführende Untersuchungen können die hier verwendeten Messmethoden in jedem Fall hinsichtlich des Arbeitsaufwands verbessert werden.

10.3 Ausblick

Eine noch tiefergehende Beschäftigung mit dem Thema erscheint nach den ersten Ergebnissen lohnenswert. Das in Abschnitt 6 beschriebene Sprachkorpus bietet entsprechend viel Sprachmaterial für eine ausführlichere akustische Analyse oder Perzeptionsexperimente mit dem Fokus auf prosodische Faktoren. Allerdings ist die Erweiterung der Stichprobe in der Anzahl der Subjekte bzw. Kategorien (z.B. durch Ergänzung einer Kategorie „Laiensänger ohne Gesangsunterricht“) empfehlenswert, um statistisch besser gesicherte Ergebnisse zu erhalten bzw. mehr Hinweise auf Kausalität zu ermöglichen. Eine Aufhebung der Kategorien und größer angelegte Bedienung der kontinuierlichen Stimmbildungsskala wäre dafür ebenso von Nutzen. Bezüglich des kausalen Aspekts der Frage zur Erlernbarkeit bleibt natürlich immer noch die Langzeitstudie als beste Methode bestehen. Diese gestaltet sich jedoch nicht nur aufgrund des zeitlichen Rahmens, sondern auch hinsichtlich der Stichprobenzusammenstellung als schwierig, da beispielsweise Gesangsstudenten zum Beginn des Studiums, der ein kontrollierter Zeitpunkt einer Längschnittstudie sein könnte, bereits unterschiedliche Stimmbildungsvoraussetzungen mitbringen.

Weiterhin könnte das Parameterset für potentielle akustische Korrelate von Stimmklang erweitert werden, z.B. in Anlehnung an die klinische Phonetik und das in Abschnitt 3.1 erwähnte MDVP, wobei dieses Analysetool für Messungen pathologischer Stimmen und nicht für gesunde Stimmen ausgerichtet ist und daher womöglich andere Phänomenbereiche fokussiert.

Eine erste Sichtung der Texturassoziationen des Experimentteils 3 bringt vielversprechende Einblicke in dieses spannende psychoakustische Forschungsgebiet. Betrachtet man wiederum die Sprecher mit den positivsten und negativsten Werten auf der „Schönheit“-Skala bezüglich der assoziierten Texturen, so ergibt sich ein nicht überraschendes Bild. Abbildung 13 zeigt eine Darstellung des Antwortverhaltens der Hörer für die 4 Sprecher. Dabei wurden die drei Texturbilder mit den höchsten Häufigkeiten[30] ausgewählt und die Größe ihrer Darstellung den Häufigkeitsverhältnissen angepasst.

Folgende Dinge lassen sich hier direkt beobachten: Vergleicht man die beiden männlichen Sprecher, so zeichnen sich die gewählten Texturen für PSm01JR durch eine weiche Oberfläche oder eine gewisse Regelmäßigkeit im Muster aus, während die Texturen für LSm03TT hauptsächlich durch eine raue Oberfläche oder eine unregelmäßige Struktur gekennzeichnet sind. Diese Eigenschaften scheinen sich mit der in Abschnitt 9 beschriebenen Korrelation zwischen Schönheit und dem perzeptiven Faktor „Weichheit“ zu decken, der sowohl die Adjektivskala „rau-weich“ als auch „ebenmäßig-unebenmäßig“ mit einschloss. Für die weiblichen Sprecher zeigt sich auf den ersten Blick kein so eindeutiges Bild. Die gewählten Texturen für GUw02IG lassen allerdings eher auf den Eindruck eines hauchigen, dünnen Stimmklangs schließen, während sich das prominente Merkmal der Texturen für LSw05KP wohl auch mit „Weichheit“ beschreiben lässt. In jedem Fall lässt dieser erste Einblick für die noch ausstehende quantitative Analyse auf interessante Ergebnisse hoffen.

[30] Die Übereinstimmung in der Stichprobe wurde noch nicht getestet, aber die hier dargestellten Texturen verfügen alle über eine Häufigkeit, die über dem Zufallsniveau von 4 (70 Hörer/16 Texturen) liegt.

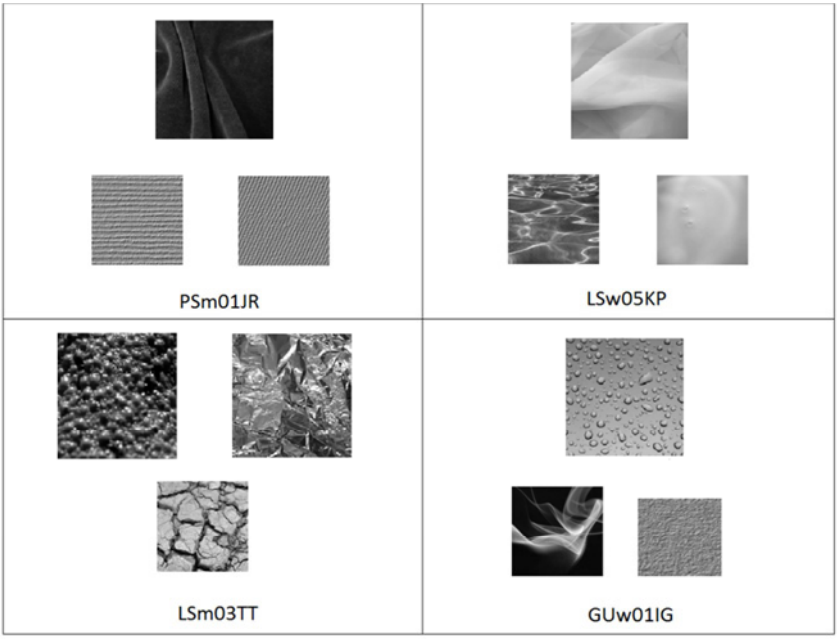


Abbildung 13. Übersicht zu den am häufigsten zugewiesenen Texturen für die Sprecher PSm01JR, LSw05KP, LSm03TT und GUw02IG.

11. Referenzen

http://www.dwds.de

http://clarin.phonetik.uni-muenchen.de/BASWebServices/

ABERCROMBIE, D. 1967. Elements of General Phonetics. Edinburgh: EUP.

ALKU, P. / T. BÄCKSTRÖM / E. VILKMAN. 2002. Normalized amplitude quotient for parametrization of the glottal flow. Journal of the Acoustical Society of America 112, 701-710.

BAER, T. / A. LÖFQVIST / N. S. MCGARR. 1983. Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Journal of the Acoustical Society of America 73, 1304-1308.

BANSE, R. / K. R. SCHERER. 1996. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70, 614-636.

BERRY, D. S. 1992. Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. Journal of Nonverbal Behavior 16, 41-54.

BHUTA, T. / L. PATRICK / J. D. GARNETT. 2003. Perceptual evaluation of voice quality and its correlation with acoustic measurements. Journal of Voice 18, 299-304.

BOERSMA, P. 2004. Stemmen meten met Praat. Stem-, Spraak-, en Taalpathologie 12, 237-251.

BOERSMA, P. / D. WEENINK. 2014. Praat: Doing Phonetics by Computer. [freie Software], http://www.praat.org/.

BRÜCKL, M. / W. SENDLMEIER. 2003. Aging female voices: An acoustic and perceptive analysis. VOQUAL'03, Genf.

CATFORD, J. C. 1988. A Practical Introduction to Phonetics. Oxford: Clarendon.

CHILDERS, D. G. / C.K. LEE. 1991. Vocal quality factors: Analysis, synthesis, and perception. Journal of the Acoustical Society of America 90, 2394-2410.

COLLINS, S. A. 2000. Men's voice and women's choice. Animal Behaviour 60, 773-780.

FANT, G. 1960. Acoustic Theory of Speech Production. The Hague: Mouton.

FEINBERG, D. R. / B. C. JONES / A. LITTLE / D. M. BURT / D. I. PERRETT. 2005. Manipulations of fundamental frequencies influence the attractiveness of human male voices. Animal Behaviour 69, 561-563.

FERRAND, C. T. 2002. Harmonics-to-noise ratio: An index of vocal aging. Journal of Voice 16, 480-487.

FISCHER, P.-M. 1993. Die Stimme des Sängers. Analyse ihrer Funktion und Leistung – Geschichte und Methodik der Stimmbildung. Stuttgart: Metzler.

GORDON, M. / P. LADEFOGED. 2001. Phonation types: A cross-linguistic overview. Journal of Phonetics 29, 383-406.

GÖRS, K. 2011. Von früh bis spät - Phonetische Veränderungen der Sprechstimme im Tagesverlauf. Bachelorarbeit, ISFAS, CAU Kiel.

KLATT, D. H.. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87, 820-857.

KÖSTER, O. / J.-P. KÖSTER. 1996. Acoustic of the Vocal Tract: Synchronization of Digital High Speed Imaging, EGG, and the Speech Wave. In: Braun, A. (Hrsg.), Untersuchungen zu Stimme und Sprache (S. 1-11). Stuttgart: Franz Steiner.

KREIMAN, J. / D. VANLANCKER-SIDTIS / B. GERRATT. 2005. Perception of voice quality. In: Pisoni, D., Remez, R. (Hrsg.), The Handbook of Speech Perception (S. 338-362). Oxford: Blackwell.

DE KROM, G. 1994. Acoustic Correlates of Breathiness and Roughness. Experiments on Voice Quality. Utrecht: LEd.

LAVER, J. 1980. The Phonetic Description of Voice Quality. Cambridge: CUP.

Liu, X. / Y. Xu. 2011. What makes a female voice attractive? ICPhS XVII, Hongkong, 1274-1277.

MIYAKE, K. / M. ZUCKERMAN. 1993. Beyond personality impressions: Effects of the physical and vocal attractiveness on false consensus, social comparison, affiliation, and assumed and perceived similarity. Journal of Personality,61, 411-437.

MOOS, A. 2010. Long-term formant distribution as a measure of speaker characterictics in read and spontaneous speech. Phonetician 101/102, 7-24.

MOOS, A. / SIMMONS, D. / J. SIMNER /R. SMITH. 2013. Color and texture associations in voice-induced synesthesia. Frontiers in Psychology 4(568), 1-12.

MÜLLER-BRUNOW. 1890. Tonbildung oder Gesangsunterricht? Beiträge zur Aufklärung über das Geheimnis der schönen Stimme. Leipzig: Friedrich Hofmeister.

NOLAN, F. / C. GRIGORAS. 2005. A case for formant analysis in forensic speaker identification. Journal of Speech, Language and the Law 12, 143-173.

ROTHENBERG, M. / J. J. MAHSHIE. 1988. Monitoring vocal fold abduction through vocal fold contact area. Journal of Speech and Hearing Research 31, 338-351.

SCHARRER, L. / U. CHRISTMANN / M. KNOLL. 2011. Voice modulations in German ironic speech. Language and Speech 5(44), 435-465.

SCHERER, K. R. 1972. Judging personality from voice: A cross-cultural approach to an old issue in inter-personal perception. Journal of personality 40, 191-210.

SCHERER, K. R. / R. BANSE / H. G. WALLBOTT / T. GOLDBECK. 1991. Vocal cues in emotion encoding and decoding. Motivation and Emotion 15, 123-148.

SCHÖTZ, S. / C. MÜLLER. 2007 A study of acoustic correlates of speaker age. In: Müller, C. (Hrsg), Speaker Classification II (S. 1-9). Heidelberg: Springer.

SEIDNER, W. / J. WENDLER. 1997. Die Sängerstimme. phoniatrische Grundlagen der Gesangsausbildung. Berlin: Henschel.

SHERER R. C. / D. G. DRUKER / I. R. TITZE. 1988. Electroglottography and direct measurement of vocal fold contact area. In: Fujimura, O. (Hrsg), Vocal Physiology: Voice Production, Mechanisms and Functions. New York: Raven Press Ltd.

SUNDBERG, J. 1987. The Science of the Singing Voice. DeKalb, IL: Northern Illinois University Press.

SJÖLANDER, K. / J. BESKOW. 2013. WaveSurfer. [freie Software], http://www.speech.kth.se/wavesurfer/.

TREDE, D. 2011. Ist Ironie nur Prosodie? Zu lautlichen Reduktionen ironischer und nicht-ironischer Äußerungen. Bachelorarbeit, ISFAS, CAU Kiel.

WEIKL, B. 1998. Vom Singen und anderen Dingen. Ein Ratgeber für alle, die beruflich oder privat mit einer klangvollen Stimme erfolgreicher sein wollen. Wien: Kremayr & Scheriau.

XUE, S. A. / D. DELIYSKI. 2001. Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. Educational Gerontology 21, 139-168.

ZUCKERMAN M. / R. DRIVER. 1989. What sounds beautiful is good: the vocal attractiveness stereotype. Journal of Nonverbal Behavior 13, 67-82.

ZUCKERMAN, M. / K. MIYAKE. 1993. The attractive voice: what makes it so? Journal of Nonverbal Behavior 17(2), 119-135.

12. Anhang

ANHANG A

Fragebogen zu Personendaten – Seite 1

Probandendaten **Datum:** ..... **Probandenkürzel:** .....

Studie: „Einfluss von Stimmbildung und Gesang auf die Sprechstimme"

Name: .....

Geschlecht: weiblich ☐ männlich ☐ **Herkunft:** .....

Alter: .....

Stimmlage der Gesangsstimme: Sopran ☐ Mezzosopran ☐ Alt ☐
Tenor ☐ Bariton ☐ Bass ☐
k. A. ☐

Beruf: .....

gegenwärtige oder vergangene Erkrankungen des Stimmapparats: ja ☐ nein ☐
(wenn ja, bitte Rücksprache mit dem Experimentator halten)

Angaben zu vergangenen und aktuellen stimmbildnerischen Maßnahmen:

Art	Zeitraum	Intensität

ANHANG A

Fragebogen zu Personendaten – Seite 2

Angaben zu anderweitiger Gesangstätigkeit:

Art	Zeitraum	Intensität

Ich bin mit der Weiterverarbeitung der von mir angegebenen Daten im Rahmen der Studie, d.h. nur durch die mitwirkenden wissenschaftlichen Mitarbeiter, einverstanden. Bei Veröffentlichung der Ergebnisse dürfen keine Informationen, durch die ich identifiziert werden könnte, weitergegeben werden.

.....

Unterschrift

Kürzel	Geschlecht	Alter	Herkunft	Stimmlage	Beruf	Stimmbildung (h)	Gesang (h)	Anmerkung
GUm01GZ	m	30	S-H	k.A.	Student	0	0	
GUm02CS	m	36	S-H	k.A.	NH Lehrer	10	0	Logopädie
GUm03BT	m	24	HH	k.A.	Student	0	0	
GUm04FT	m	30	BB	k.A.	Student	0	124,5	
GUw01TH	w	29		k.A.	Doktorandin	0	120	
GUw02IG	w	23	S-H	k.A.	Studentin	0	200	
GUw03SH	w	30	Bayern	k.A.	Studentin	22,5	0	Theater AG
GUw04PB	w	45	S-H	k.A.	Sekretärin	0	96	
LSm01CH	m	46	S-H	Bariton	Rektor	880	6876	
LSm02HB	m	45	S-H	Bariton	Lehrer	356	3640	Musiklehrer
LSm03TT	m	26	S-H	Bariton	Student	248	1368	
LSm04HS	m	34	S-H	Bass	Kaufm. Ang.	240	2040	
LSw01SI	w	33	S-H	Mezzo	Dozentin	636	2136	
LSw04SD	w	41	NRW	Sopran	Ärztin	240	4320	144h Logopädie
LSw05KP	w	41	S-H	Sopran	Lehrerin	720	3636	Musiklehrerin, HF Gesang
LSw06MD	w	41	Hessen/HH	Sopran	Lehrerin	216	1200	Musiklehrerin, HF Flöte
PSm01JR	m	38	Niedersachsen	Bass	Sänger	1959	7440	
PSm02MM	m	32	Saarland	Tenor	Opersänger	1884	7480	
PSw01NY	w	32	Schwaben	Sopran	Opernsängerin	934	4576	
PSw02KR	w	43	NRW	Mezzo	Sängerin	1588	6840	
PSw03LS	w	33	Niedersachsen	Sopran	Sängerin	905	2800	

Tabelle A1. Probandendaten der Sprachaufnahmen.

ANHANG B

Transkripte der ausgewählten Phrasen für Stimulustyp 2

GUm01GZ: also es is ja nich selbstverständlich wenn ich in ja zum beispiel in amerika leben würde ja also in den usa dann könnt ich mir das gar nich leisten / ähm hab viel erlebt auch viele sachen so gesehen / was mich stört oder was ich aber auch nicht ändern kann ist natürlich der alterunterschied / das war so mit das prägendste äh erlebnis

GUm02CS: und er sagte er sei am donnerstag abend um halb 8 nach hause gekommen / auch wenn das jetzt so dinge sind die ich dann im beruf gar nicht mehr brauche das interessiert mich einfach trotzdem dann noch ne / is auch ein namensvetter von mir /es geht ja auch nicht nur ums geld sondern es geht auch darum was einem das gibt was man macht

GUm03BT: da ging dann 40 jahre lang im prinzip der kampf zwischen den behörden und diesen mittlerweile althippies hin und her / und da hab ich dann festgestellt dass ich ähm was ändern muss / ein sehr guter freund von ihm hat dort ne ganze insel / das hat schon spaß gemacht dort der unterricht auf jeden fall

GUm04FT: ähm ich bin schon gespannt auf was es später hinausläuft / und hab irgendwie aber mit der zeit gemerkt dass es so ganz interessant is aber es is eben halt doch nich äm meine leidenschaft / was es war wirklich ähm musst ich auch erst mit der zeit rausfinden / ja das war so der anstoß ich bin dann äh ich hab mich dann immer mehr informiert

GUw01TH: dass ich hierhin gekommen bin nach kiel war eher zufall / im gegensatz zu kiel wirkt es einfach heimlicher / also ich bin relativ spontan hierher gekommen / ähm ja und ich bin da mmh bin da relativ glücklich

GUw02IG: ja nach dem abitur hab ich damit auch gleich angefangen / und um in beiden perfekt zu werden also is sehr schwierig / ja das hat dann auch alles geklappt / und deswegen hab ich mich entschieden dadurch dass man irgendwie keine vorkenntnisse haben musste dann anzufangen

GUw03SH: also der studienort der den fand ich nich so prickelnd / aber so in die richtung ja mehr so lustig mehr so nich das leben nicht allzu ernst nehmend / ich weiß dass is jetzt schon ne zeit lang her / berlinerisch find ich ganz interessant und natürlich alles was in und um köln rum gesprochen wird

GUw04PB: ja und dann hab ich einige verschiedene jobs gehabt / und dann eben auch mal an sowas wie wie so ner bitte nachzukommen oder so dass man einfach mal für sowas ruhe hat / da hätt ich gar nich mit gerechnet vorher / wir haben telefoniert uns irgendwie auch so ein bisschen angefreundet am telefon

LSm01CH: also dann erzähl ich vielleicht zwei besondere szenen eine szene da war ich acht jahre alt / so das war für mich als kind schon sehr beeindruckend / da könnte man jetzt sehr lange drüber reden / ja trotzdem ist das komfortable daran sich auszusuchen was man grade machen möchte

LSm02HB: ja müssen sie sich vorstellen äh da sitzt man da in der neuen rabenstraße in am am dammtorbahnhof in hamburg / und das ist großartig ein großartiger spaß den kleinen nick zu lesen / das hat auch durchaus seine schattenseiten / zum beispiel hör ich den radiotatort total gerne

LSm03TT: natürlich merkt man dann dass das nich geht / allerdings hat man sich da nach zwei jahren irgendwann dran gewöhnt und am ende haben wir uns super verstanden / gut aber das gehört dazu und im nachhinein ist das sicherlich ne witzige und gute erfahrung gewesen / und das war absolut toll

LSm04HS: und damals fand ich das ziemlich grauenhaft ich denke die meisten in der klasse fanden das ganz nett / und insofern blieb ich da viele jahre lang bis ich irgenwann mal wegzog und äh das nich mehr beibehalten konnte / also vielleicht hab ich durchaus was von ihm geerbt in der hinsicht also was das interesse angeht / und das gefiel mir tatsächlich relativ gut

LSw01SI: für ungefähr drei jahre zweieinhalb jahre bis zum schulabschluss / ja und wie so das alltägliche geschäft aussieht wie lange sie ausbleiben / also hier in kiel auch wieder so für ein anderthalb jahre / zwischendurch ein halbes jahr pause aber ansonsten durchgängig

LSw04SD: ich weiß nich so richtig warum die leute passten mir eigentlich immer gar nich so richtig fand die eigentlich eher so ein bisschen brav und langweilig und bisschen bieder so / also seit seit zwei jahren glaub ich is das fusioniert worden dieses ganze konstrukt / das hab ich glaub ich so vom sechsten bis zum vierzehnten lebensjahr ungefähr gemacht / auch wenn die räumlichkeiten alt sind aber die ganzen abläufe da sind total professionell

LSw05KP: mit 16 ging das los / und da ist die entscheidung irgendwann von alleine gefallen weil das da so schön war an dem gymnasium / ja und das hab ich dann gemacht nochmal drei jahre / hat also diesen ganzen skandinavischen die ganze skandinavische literatur mit nach deutschland genommen und hat da den schwerpunkt gesetzt

LSw06MD: stimmt das ich muss mal kurz nachrechnen zwei vier sechs die dritte schwester ja alle zwei jahre ungefähr / manchmal behaupt ich einfach die stadt heißt seitdem so / aber es war auch nich so richtig ich wurde nie so richtig warm damit / das schaffen wir noch wir fahren nach darmstadt damit das nachher im perso steht

PSm01JR: joa das is so das ganz normale geschäft eigentlich / und die hab ich mir dann tatsächlich ausgeschnitten aus ner zeitung und an die zimmertür geklebt / das endete zweitausendzwölf / streng war er eigentlich nich er war lieb bestimmt es war einfach ne schöne atmosphäre

PSm02MM: und ich sagte ja da möchte ich gerne mitmachen / schnell hab ich aber gemerkt dass ich dort auch die chance habe mich selbst zu perfektionieren beziehungsweise dieses handwerk zu perfektionieren / äh das war zweitausend und fünf / so kam es dass ich durch eine bewerbung äh ein stipendium bekommen habe

PSw01NY: also würd ich jetzt vom jetztigen standpunkt aus so bezeichnen / die koordination ist jetzt endlich mal so weit dass ich damit arbeiten will davor war ich immer sehr unzufrieden / da hab ich dann so n kleinen trick angewendet / also alles was man an funktionen irgendwie umschalten können muss konnt ich ausprobieren

PSw02KR: aber mein bruder war in die dritte klasse gekommen und dann haben wir das verheimlicht / das waren dann drei jahre wo ich ähm alle sechs wochen in den odenwald gefahren bin / und dann äh musst ich das dann noch irgendwie erledigen / das is mein ding das is einfach total spannend und danach arbeite ich halt jetzt auch

PSw03LS: ich hatte aber gehört dass das nun gar nich gut sei war ja aber auch noch sehr jung wusste nich wirklich bescheid / und dem war dann leider gar nich so / das hab ich aber erst hinterher erfahren / der hat da nie was gesagt ich hab auch nie was gesagt aber es war immer so unten drunter zu spüren

ANHANG C

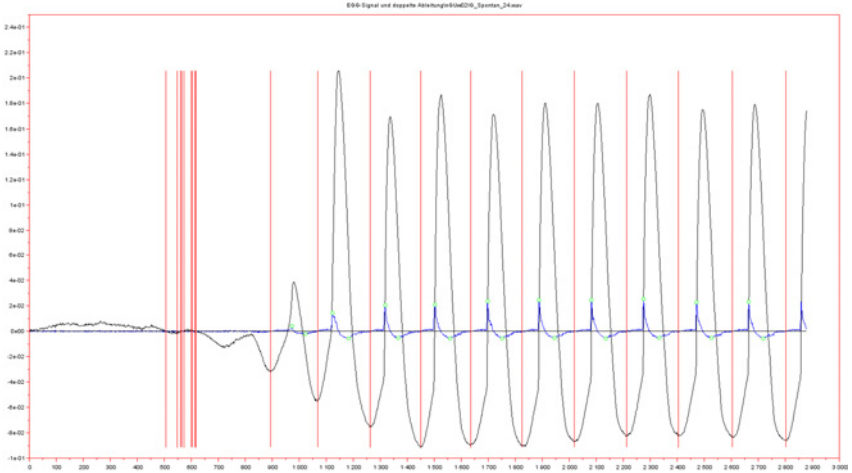


Abbildung C.1. Beispiel für die Analysemethode des OQ: EGG-Signal (schwarz) und differenziertes Signal (blau). Die roten Linien markieren die berechneten Periodengrenzen, die grünen Punkte markieren die berechneten Maxima und Minima des differenzierten Signals.

Sprecher	Ge-schlecht	Stimu-lustyp	f0 mean	f0 SD (Hz)	f0 SD (st)	f0 min	f0 max	f0 range (Hz)	f0 range (st)	LTF2 mean	LTF2 SD	OQ mean	HNR mean	HNR mean (INT)	jitter mean	jitter mean (INT)
GUm01GZ.mp3	**m**	ST2	121,07	12,41	1,70	95	170	75	10,07	1298,55	166,41	0,62	15,21	29,87	0,58%	0,20%
GUm02CS.mp3	**m**	ST2	118,11	20,10	2,84	80	200	120	15,86	1249,85	305,37	0,66	16,91	26,16	0,55%	0,43%
GUm03BT.mp3	**m**	ST2	97,23	11,42	1,95	75	135	60	10,18	1224,04	160,32	0,46	12,88	25,32	0,88%	0,25%
GUm04FT.mp3	**m**	ST2	99,16	8,74	1,47	80	130	50	8,41	1174,32	138,13	0,54	15,89	23,05	1,12%	0,50%
GUw01TH.mp3	**w**	ST2	226,07	30,88	2,32	180	295	115	8,55	1385,04	232,38	0,56	23,53	25,37	0,62%	0,38%
GUw02IG.mp3	**w**	ST2	231,57	20,70	1,45	200	330	130	8,67	1459,37	207,74	0,60	24,48	30,37	0,57%	0,07%
GUw03SH.mp3	**w**	ST2	234,43	33,78	2,34	180	390	210	13,39	1409,97	160,35	0,60	21,59	28,34	0,48%	0,08%
GUw04PB.mp3	**w**	ST2	168,53	24,43	2,45	130	230	100	9,88	1423,23	175,46	0,52	17,32	23,86	0,89%	0,25%
LSm01CH.mp3	**m**	ST2	99,50	18,58	3,16	60	160	100	16,98	1229,63	165,48	0,43	13,40	23,52	1,12%	0,30%
LSm02HB.mp3	**m**	ST2	108,14	17,44	2,70	75	175	100	14,67	1261,23	140,68	0,54	12,32	25,41	0,92%	0,24%
LSm03TT.mp3	**m**	ST2	103,11	16,88	2,59	75	195	120	16,54	1336,75	119,90	0,48	14,42	25,50	0,58%	0,26%
LSm04HS.mp3	**m**	ST2	102,94	15,41	2,42	80	155	75	11,45	1362,07	135,21	0,41	8,18	26,90	0,58%	0,24%
LSw01SI.mp3	**w**	ST2	193,99	19,91	1,68	165	295	130	10,06	1531,19	184,70	0,62	18,62	26,97	1,36%	0,19%
LSw04SD.mp3	**w**	ST2	197,35	18,80	1,61	160	290	130	10,30	1422,33	148,97	0,52	21,51	32,79	0,69%	0,05%
LSw05KP.mp3	**w**	ST2	215,88	30,96	2,26	165	405	240	15,55	1519,75	160,15	0,56	22,61	28,10	0,64%	0,06%
LSw06MD.mp3	**w**	ST2	227,23	38,32	2,80	175	375	200	13,19	1469,45	177,45	0,61	20,94	28,90	0,35%	0,05%
PSm01JR.mp3	**m**	ST2	114,48	34,11	4,77	70	240	170	21,33	1275,17	166,43	0,50	12,74	30,72	1,83%	0,08%
PSm02MM.mp3	**m**	ST2	127,02	10,18	1,37	100	160	60	8,14	1235,51	167,12	0,47	17,65	25,82	0,72%	0,11%
PSw01NY.mp3	**w**	ST2	182,25	22,30	2,04	140	240	100	9,33	1537,29	96,87	0,70	18,48	23,68	0,66%	0,19%
PSw02KR.mp3	**w**	ST2	170,12	25,45	2,40	135	290	155	13,24	1278,05	161,00	0,68	19,83	28,31	0,58%	0,10%
PSw03LS.mp3	**w**	ST2	216,63	25,09	1,93	175	305	130	9,62	1396,19	189,87	0,53	22,61	31,64	0,46%	0,05%

Tabelle C.1. Messwerte der akustischen Parameter für alle Sprecher in ST2.

Sprecher	Ge-schlecht	Stimu-lustyp	f0 mean	f0 SD (Hz)	f0 SD (st)	f0 min	f0 max	f0 range	f0 range (st)	LTF2 mean	LTF2 SD	OQ mean	HNR mean	HNR mean (INT)	jitter mean	jitter mean (INT)
GUm01GZ.mp3	**m**	ST3	129,01	19,32	2,47	95	210	115	13,73	1395,71	306,89	0,53	14,17	29,87	1,23%	0,20%
GUm02CS.mp3	**m**	ST3	125,12	29,94	3,86	85	230	145	17,23	1367,15	370,60	0,62	11,91	26,16	1,12%	0,43%
GUm03BT.mp3	**m**	ST3	126,42	35,31	4,40	70	250	180	22,04	1346,99	341,10	0,49	12,08	25,32	0,77%	0,25%
GUm04FT.mp3	**m**	ST3	104,29	15,61	2,51	75	170	95	14,17	1426,92	312,11	0,50	13,00	23,05	0,93%	0,50%
GUw01TH.mp3	**w**	ST3	214,41	40,32	3,09	160	330	170	12,53	1560,62	421,97	0,53	19,69	25,37	0,60%	0,38%
GUw02IG.mp3	**w**	ST3	235,92	27,04	1,92	200	330	130	8,67	1582,07	382,89	0,57	23,73	30,37	0,55%	0,07%
GUw03SH.mp3	**w**	ST3	234,82	25,86	1,84	200	310	110	7,59	1588,74	413,51	0,58	22,25	28,34	0,34%	0,08%
GUw04PB.mp3	**w**	ST3	180,91	37,95	3,50	125	340	215	17,32	1538,28	387,69	0,46	18,46	23,86	0,40%	0,25%
LSm01CH.mp3	**m**	ST3	101,14	23,29	3,61	70	195	125	17,74	1440,64	352,18	0,46	12,89	23,52	1,14%	0,30%
LSm02HB.mp3	**m**	ST3	169,98	44,11	4,59	95	300	205	19,91	1481,97	395,57	0,45	15,94	25,41	0,41%	0,24%
LSm03TT.mp3	**m**	ST3	106,66	18,17	2,76	80	170	90	13,05	1480,35	344,34	0,45	12,57	25,50	1,01%	0,26%
LSm04HS.mp3	**m**	ST3	123,57	30,56	4,20	70	205	135	18,60	1564,68	439,33	0,43	14,22	26,90	0,95%	0,24%
LSw01SI.mp3	**w**	ST3	199,15	38,39	3,05	140	350	210	15,86	1657,45	400,99	0,59	18,64	26,97	0,77%	0,19%
LSw04SD.mp3	**w**	ST3	211,41	29,78	2,33	165	300	135	10,35	1566,61	398,90	0,52	22,66	32,79	0,33%	0,05%
LSw05KP.mp3	**w**	ST3	225,87	43,62	3,10	170	380	210	13,93	1692,03	395,76	0,60	20,47	28,10	0,34%	0,06%
LSw06MD.mp3	**w**	ST3	230,91	56,63	3,68	165	510	345	19,54	1617,16	446,31	0,59	22,12	28,90	0,41%	0,05%
PSm01JR.mp3	**m**	ST3	131,15	47,69	5,86	70	265	195	23,05	1497,14	327,45	0,43	12,77	30,72	0,69%	0,08%
PSm02MM.mp3	**m**	ST3	159,95	46,52	4,74	100	310	210	19,59	1477,29	400,05	0,44	14,09	25,82	0,84%	0,11%
PSw01NY.mp3	**w**	ST3	195,59	38,91	3,13	135	440	305	20,45	1587,35	439,96	0,68	17,52	23,68	0,74%	0,19%
PSw02KR.mp3	**w**	ST3	190,37	37,53	3,25	140	360	220	16,35	1517,14	424,91	0,52	21,89	28,31	0,37%	0,10%
PSw03LS.mp3	**w**	ST3	232,76	45,33	3,30	145	400	255	17,57	1558,46	433,33	0,55	19,73	31,64	0,32%	0,05%

Tabelle C.2. Messwerte der akustischen Parameter für alle Sprecher in ST3.

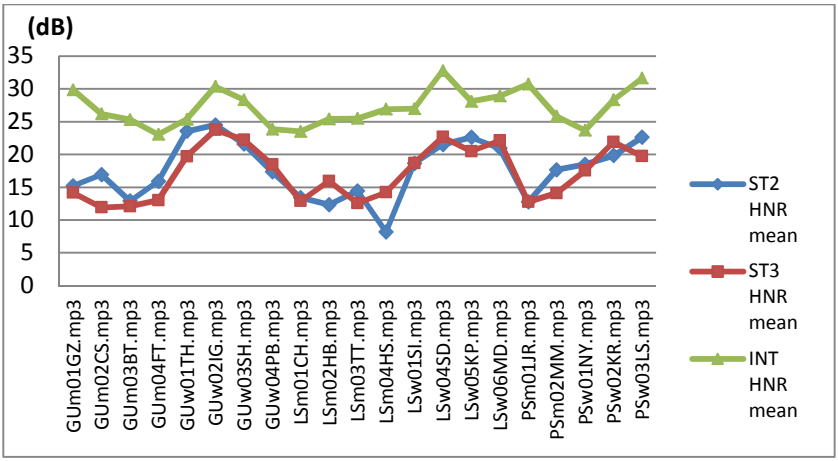


Abbildung C2. HNR-Messreihen aus Spontansprache (ST2), Lesesprache (ST3) und gehaltenen Vokalen (INT).

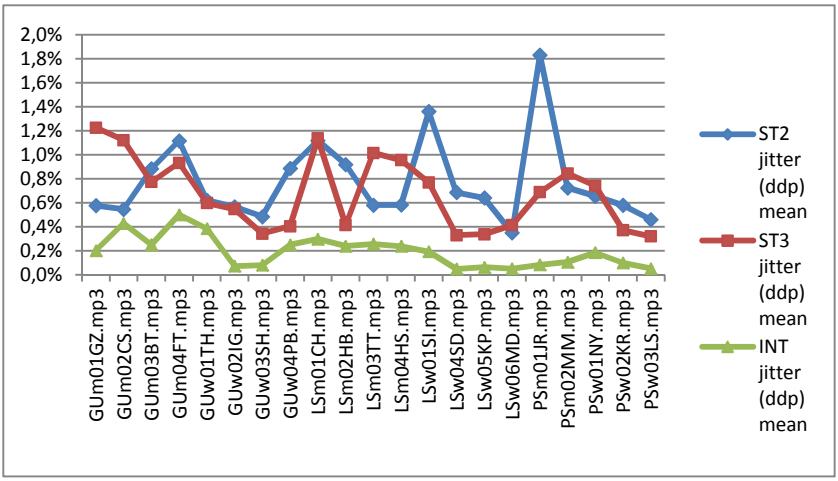
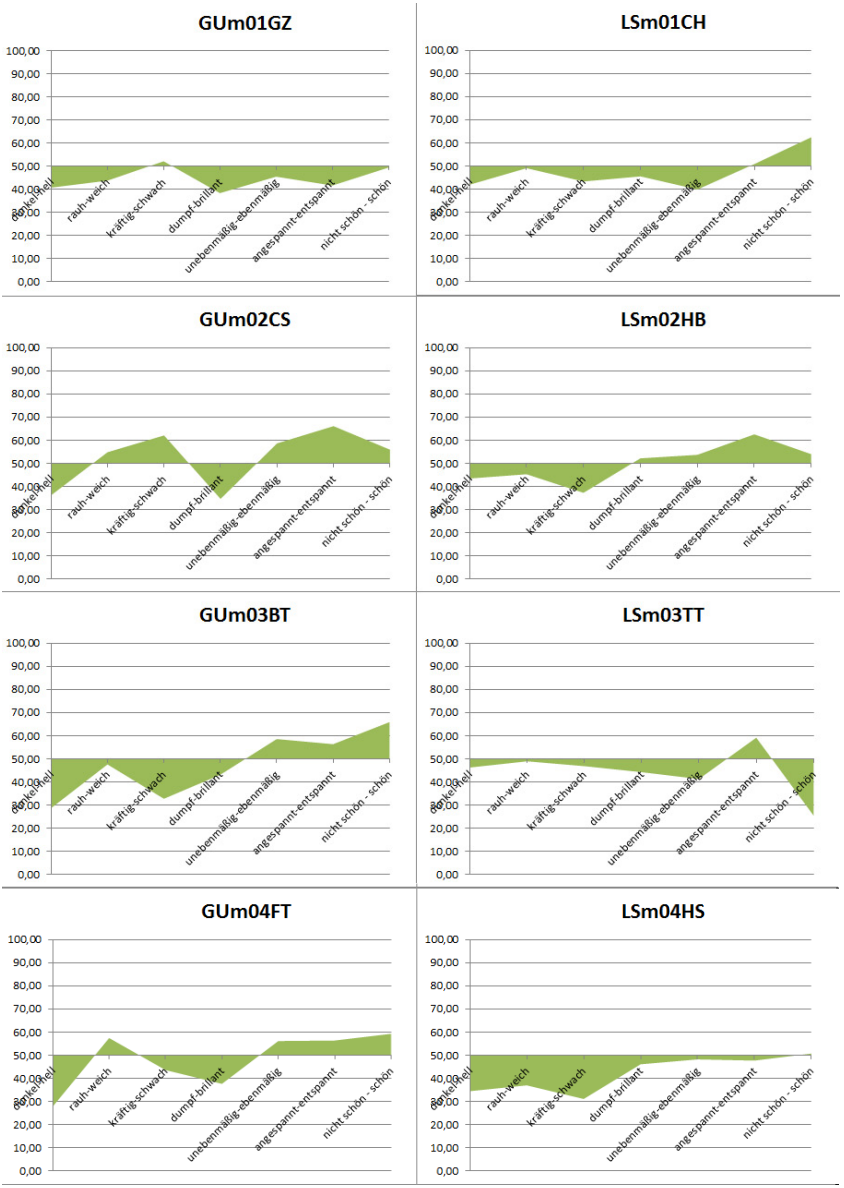
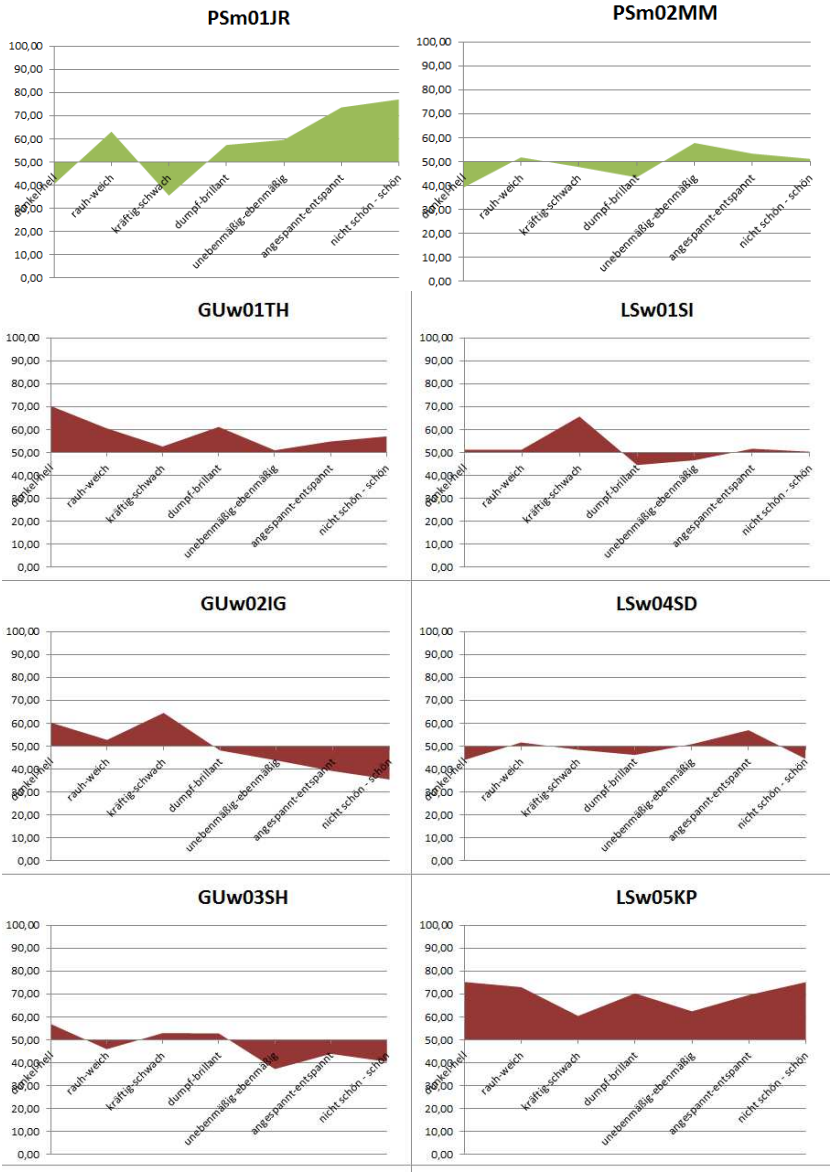
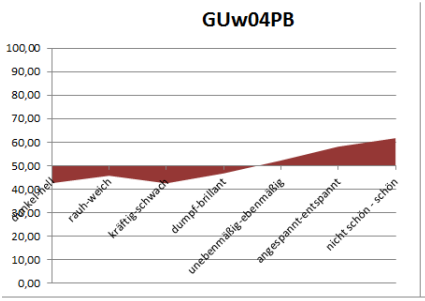


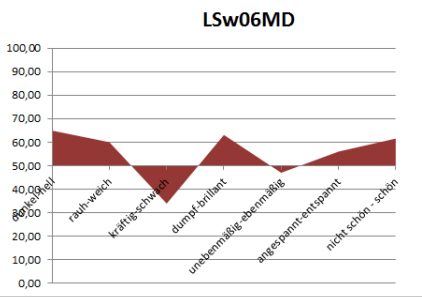
Abbildung C3. Jitter-Messreihen aus Spontansprache (ST2), Lesesprache (ST3) und gehaltenen Vokalen (INT).

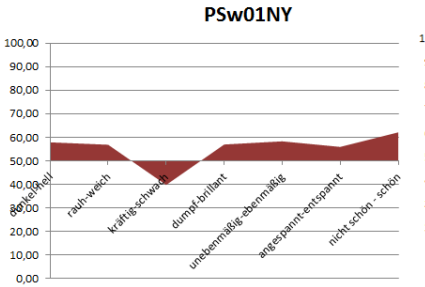
Panoramen der männlichen (grün) und weiblichen Sprecher (rot) anhand der Adjektivskalen aus Teil 2A und 2B

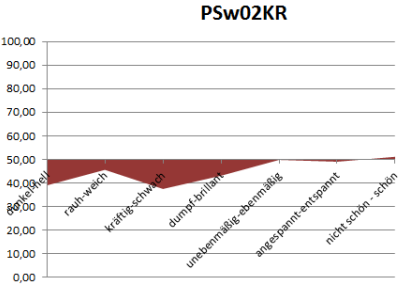


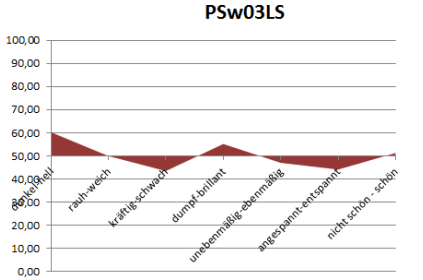

PSm01JR
PSm02MM
GUw01TH
LSw01SI
GUw02IG
LSw04SD
GUw03SH
LSw05KP
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön


GUw04PB
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön


LSw06MD
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön


PSw01NY
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön


PSw02KR
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön


PSw03LS
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
dunkel-hell
rauh-weich
kräftig-schwach
dumpf-brillant
unebenmäßig-ebenmäßig
angespannt-entspannt
nicht schön - schön

Kieler Arbeiten in Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

1 Speech Data Acquisition: The Underestimated Challenge

Oliver Niebuhr
Analyse gesprochener Sprache
Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
niebuhr@isfas.uni-kiel.de

Alexis Michaud
International Research Institute MICA
Hanoi University of Science and Technology, CNRS, Grenoble INP, Vietnam
Langues et Civilisations à Tradition Orale, CNRS/Sorbonne Nouvelle, France
alexis.michaud@mica.edu.vn

The second half of the 20th century was the dawn of information technology; and we now live in the digital age. Experimental studies of prosody develop at a fast pace, in the context of an "explosion of evidence" (Janet Pierrehumbert, Speech Prosody 2010, Chicago). The ease with which anyone can now do recordings should not veil the complexity of the data collection process, however. This article aims at sensitizing students and scientists from the various fields of speech and language research to the fact that speech-data acquisition is an underestimated challenge. Eliciting data that reflect the communicative processes at play in language requires special precautions in devising experimental procedures and a fundamental understanding of both ends of the elicitation process: speaker and recording facilities. The article compiles basic information on each of these requirements and recapitulates some pieces of practical advice, drawing many examples from prosody studies, a field where the thoughtful conception of experimental protocols is especially crucial.

1. Introduction: Speech Data Acquisition as an Underestimated Challenge

The second half of the 20th century was the dawn of information technology; and we now live in the digital age. This results in an *"explosion of evidence"* (Janet Pierrehumbert, Speech Prosody 2010, Chicago), offering tremendous chances for the

analysis of spoken language. Phoneticians, linguists, speech therapists, speech technology specialists, anthropologists, and other researchers routinely record speech data the world over. There remains no technological obstacle to collecting speech data on all languages and dialects, and to sharing these data over the Internet. The ease and the speed with which recordings can now be conducted and shared should not veil the complexity of the data collection process, however.

Phonetics "calls on the methods of physiology, for speech is the product of mechanisms which are basically there to ensure survival of the human being; on the methods of physics, since the means by which speech is transmitted is acoustic in nature; on methods of psychology, as the acoustic speech-stream is received and processed by the auditory and neural systems; and on methods of linguistics, because the vocal message is made up of signs which belong to the codes of language" (Marchal 2009:ix). In addition to developing at least basic skills in physiology, physics, linguistics, and psychology, each of which has complexities of its own, people conducting phonetic research are expected to have a good understanding of statistical data treatment, combined with a command of one or more specific exploratory techniques, such as endoscopy, ultrasonography, palatography, aerodynamic measurements, motion tracking, electromagnetic articulography, or electroencephalography (for a description of the many components of a multisensor platform see Vaissière et al. 2010). As a result, it tends to be difficult to maintain a link between the phonetic sciences and fields of the humanities that are highly relevant for phonetic studies, and in particular for the study of prosody. Phoneticians' training does not necessarily include disciplines that would develop their awareness of the complexity and versatility of language, such as translation studies, languages, literature and stylistics, historical phonology, and sociolinguistics/ethnolinguistics. Moreover, the increasing use of digital and instrumental techniques in phonetic research is, taken by itself, a welcome development. But more and more phoneticians neglect explicit and intensive ear training, forgetting that an attentive, trained ear is the key to observations and hypotheses and hence the prerequisite for any analysis by digital and instrumental techniques. For example, we do not think that successful research on prosody can be done without the ability to produce and identify the prosodic patterns that one would like to analyse. As Barbosa (2012:33) puts it: *"The observation of a prosodic fact is never naïve, because formal instruction is necessary to see and to select what is relevant"*.

In summary, advances in phonetic technologies impose many challenges on modern phoneticians, and they can tend to replace rather than complement traditional skills. This has a direct bearing on data collection procedures. To a philologist studying written documents, it is clear that every detail potentially affects interpretation and analysis (the complexities of Greek and Latin texts are perfect examples; see, e.g., Probert 2009; Burkard 2014). Carrying the same standards into the field of speech data collection, it goes without saying that every speaker is unique, that no two recording situations are fully identical, and that human subjects participating in the experiments are no "vending machines" that produce the desired speech signals by paying and pressing a button. An experience of linguistic fieldwork, or of immersion learning of a foreign language, entails similar benefits in terms of awareness of the central importance of communicative intention (see in particular Bühler 1934, passim; Culioli 1995:15; Barnlund 2008), and of the wealth of expressive possibilities and redundant

encoding strategies open to the speaker at every instant (as emphasized, e.g., by Fónagy 2001). Researchers working on language and speech are no "signal hunters", but hunt for functions and meanings[1] as reflected in the speech signal, which itself is only one of the dimensions of expression, together with gestures and facial expressions. The definition of tasks, their contextualization, and the selection of speakers are at the heart of the research process.

The diversification of the phonetic sciences is likely to continue, together with technological advances; the literature within each subfield is set to become more and more extensive, making it increasingly impractical for an individual to develop all the skills that would be useful as part of a phonetician's background. This results in modular approaches, as against a holistic approach to communication. What is at stake is no less than a cumulative approach to research. The quality of data collection is inseparable from the validity and depth of research results; and data sharing is indispensible to allow the community to evaluate the research results and build on them for further studies.

Against this background, the present article is primarily intended for an audience of advanced students of phonetics. However, it is hoped that it can also serve as a source of information for phonetic experts and researchers who have a basic understanding of phonetics but work in other linguistic disciplines, including speech technology. The present article summarizes some basic facts, methods, and problems concerning the three pillars of speech data acquisition: the speaker (§2), the task (§3), and the recording (§4). Discussion on these central topics builds on our own experiences in the field and in the lab. Together, the chapters aim to convey to the reader in what sense data acquisition is an underestimated challenge. Readers who are pressed for time may want to jump straight to the Summary in section 5, which provides tips and recommendations on how to meet the demands of specific research questions and achieve results of lasting value for the scientific community.

Given its aim, our article is both more comprehensive and introductory than other methodologically oriented papers such as those by Mosel (2006), Himmelmann (2006), Ito and Speer (2006), Xu (2011), Barbosa (2012), and Sun and Fletcher (2014), which are all highly recommended as further reading. Most readers are likely to know much if not most of what will be said. Different readers obviously have different degrees of prior familiarity with experimental phonetics; apologies are offered to any reader for whom nothing here is new.

[1] The two terms 'meaning' and 'function' tend not to be clearly separated in the literature – including in the present article, in which we simply use both terms in combination. In the long run, a thorough methodological discussion should address the issue of the detailed characterization of 'meaning' and 'function'. To venture a working definition, meanings refer to concrete or abstract entities or pieces of information that exist independently of the communication process and are encoded into phonetic signs. Functions, on the other hand, are conveyed by phonetic patterns that are attached to these phonetic signs; they refer to the rules and procedures of speech communication. If meanings are the driving force of speech communication, then functions are the control force of speech communication.

2. The speaker

2.1 Physiological, social, and cognitive factors

Individual voices differ from one another. Physiological differences are part of what Laver (1994, 27–28) refers to as the "organic level"; they are extralinguistic, but are nevertheless of great importance to analyzing and interpreting speech data. Age and body size are perfect examples for this (cf. Schötz 2006), affecting, among others, F0, speaking rate (or duration) and spectral characteristics such as formant frequencies. Physiological variables are intertwined with social variables. For instance, there are physiological and anatomical differences between the male and female speech production apparatus, which lends female speakers a higher and breathier voice as well as higher formant values and basically allows them to conduct more distinct articulatory movements than their male counterparts within the same time window (Sundberg 1979; Titze 1989; Simpson 2009, 2012). So, *"if we randomly pick out a group of male and female speakers of a language, we can expect to find several differences in their speech"* (Simpson 2009:637).

However, Simpson (2009) also stresses in his summarizing paper that gender differences in speech do not merely have a biophysical origin. Some differences are also due to learned, i.e. socially evoked behaviour, and the dividing line between these two sources of gender-related variation cannot always be easily determined. The social phenomenon of "doing gender" is well documented; it is an object of attention on the part of speakers themselves, and 'metalinguistic' awareness of gender differences in speech is widespread, particularly with respect to grammar and lexicon (cf. Anderwald 2014). Gender-related phonetic differences are less well documented. The frequent cross-linguistic finding that women speak slower and more clearly than men is probably at least to some degree attributable do "doing gender" (cf. Simpson 2009). Further, more well-defined differences between the speech of men and women are documented by Haas (1944) for Koasati, a Native American language. Sometimes women have exclusive mastery of certain speaking styles: mastering whispered speech, including the realization of tonal contrasts without voicing, used to be part of Thai women's traditional education (Abramson 1972). In languages where the differences are less codified, they are nonetheless present: Ambrazaitis (2005) found gender differences in the realization of terminal F0 falls at the ends of utterances in German and – more recently – also in English and Swedish (see also Peters 1999:63). Compared with male speakers, female speakers prefer pseudo-terminal falls that end in a deceleration and a slight, short rise at a relatively low intensity level (Ambrazaitis 2005). This pseudo terminal fall reduces the assertiveness/finality of the statement, as compared with a terminal fall. In extreme cases, this pattern might be mistaken for an actual falling-rising utterance-final intonation pattern, which has a different communicative function. Phonetically, the difference is not considerable: a rise on the order of 2 to 4 semitones for the pseudoterminal fall, of 6 semitones for a falling-rising utterance-final pattern.

Another socially-related phenomenon is the so-called 'phonetic entrainment' or 'phonetic accommodation'. That is, when two speakers are engaged in a dialogue, they become phonetically more similar to each other, particularly when the interaction is cooperative and/or when the two dialogue partners are congenial with each other (cf.

Lee et al. 2010). Phonetic entrainment can include levels and ranges of intensity and F0, voice quality (e.g., shimmer), and speaking rate (cf. Pardo 2006; Levitan and Hirschberg 2011; Heldner et al. 2010; Hirschberg 2011; Manson et al. 2013), as well as VOT patterns, vowel qualities, and speech reduction (Giles and Coupland 1991; for a summary: Kim 2012:14-29). Delvaux and Soquet (2007) provide evidence that a speaker tends to approximate the phonetic patterns of another speaker even when the latter is not present as a dialogue partner but just heard indirectly from a distance. The affected phonetic parameters are language-specific and differ, for example, between languages with and without lexical tone (Xia et al. 2014). Moroever, entrainment is not restricted to the phonetic domain. It can equally affect syntax and wording of utterances as well as body and face gestures (Nenkova et al. 2008; Reitter and Moore 2007; Ward and Litman 2007). Entrainment emerges quickly at the beginning of a dialogue, but can also increase further during a dialogue, which is why it is often conceptualized as a combination of lower-level cognitive and higher-level social skills (cf. Pickering and Garrod 2004).

A similar combination of cognitive and social factors probably accounts for the effects of musical training on linguistic habits. It is well documented that musical training affects the way the brain works, and hence constitutes an important source of cross speaker variation. Musically trained subjects outperform untrained subjects in experiments on the perception of prosody and intonation (cf. Schön et al. 2004) and speech comprehension in noise (Parbery-Clark et al. 2009), cf. Federman (2011) for an overview. Compared with the comprehensive research on perception, relatively little is known about effects of musical training on speech production. However, there is sufficient evidence to assume that musical training does affect speech production. For example, Stegemöller et al. (2008) found global spectral differences between speakers with and without musical training, and Graupe (2014) found her musically trained subjects to have a more distinct pronunciation, including larger F0 and intensity ranges and a lower speaking rate.

There are many more sources of cross-speaker variation that we cannot list here in full detail, including the individual adaptation of speakers to adverse speaking (e.g., Lombard) conditions (cf. Mixdorff et al. 2007).

2.2 Linguistic experience and linguistic skills

An individual's experience of different languages, dialects and sociolects also exerts a deep influence on the way s/he speaks. Among other spectacular experimental findings, it has been shown that one minute of exposure is enough for the ear to attune to a foreign accent (Clarke and Garrett 2004). Dialogue partners accommodate to one another in conversation; depending on their strategy to bring out or tone down social distance, dialogue partners will tend towards either convergence or divergence. This phonetic-accommodation effect described earlier in 2.1 can have far-reaching consequences in the long run. It is reflected in amplified and entrenched forms in the speech of bilingual or multilingual speakers. A review *entitled* "The leakiness of bilinguals' sound systems" concludes that, *"although functional separation of sound systems may be both the aim and (actually quite frequent) achievement of bilinguals they are unable*

to avoid long term interference" (Watson 2002:245). Prosody is especially susceptible to the effects of language contact (Hualde 2003).

Moreover, some persons are more susceptible to influences of language contact than others. Subjects who have a good ear and a love of languages are potentially brilliant collaborators in speech data collection, able to understand tasks quickly and to apply instructions in a sensible and sensitive way. But their interest in language can adversely affect their performance when living in a language environment other than their mother tongue: speakers with high hearing sensitivity, good short-term and working memory, high attentional abilities and extensive vocabulary knowledge attune fastest to different accents (Janse and Adank 2012).

Finding out about language consultants' language abilities requires paying attention to the cultural issue of their perception of the languages they speak. Diglossia is an extremely common situation worldwide, but bilingualism is often non-egalitarian, with a prestigious national standard on the one hand and a local variety debased as 'dialect' or 'patois' on the other (on the distinction between egalitarian and nonegalitarian bilingualism: Haudricourt 1961; François 2012). In countries that enforce the adoption of a national standard, varieties other than the norm are deprecated, and speakers may consider it inappropriate to mention their mother tongue – considered as coarse, ridiculous or useless – in a curriculum vitae or a questionnaire. In China, students' résumés typically indicate Chinese as mother tongue, plus a degree of proficiency in English graded along the TOEFL scale, and sometimes other foreign languages. A native speaker of Wu or Min Chinese – branches of Sinitic that are not intelligible to speakers of Mandarin Chinese – may avoid mention of this competence in a language variety that is referred to in China as a 'dialect' and has no status as a language of culture and education. For research purposes, it may obviously be misleading to pool results from speakers whose native language has six to nine tones (see, e.g., Shen 2013) with those of native speakers of Mandarin, which has four. This is an extreme example, but issues of dialectal differences and dialect contact are well worth scrutinizing even when studying the national language of a country that has less language diversity, for instance in Europe and North America.

The second author of this paper can report first-hand on a case of code-switching in the course of data elicitation. He participated in a study of nasalization involving fiberscopic imaging of the velopharyngeal port. He unwittingly switched to Vietnamese mode when reading the logatoms /ap/, /at/ and /ak/. That day, he was accompanied by a Vietnamese speaker who was going to record a list of words designed for the study of syllables with final consonants, as an extension of studies based on electroglottographic data (Michaud 2004a) and on airflow measurements (Michaud et al. 2006). This example illustrates the fact that even in seemingly simple tasks, which in principle do not involve a high cognitive load or create particular fatigue, there can be interference between one's native language and other speech varieties, even those in which the speaker is not bilingual. It appears highly advisable to record native speakers in their home country, but even then, a sensitive inquiry into the speakers' language experience is highly advisable.

2.3 Individual strategies and preferences

To a present-day audience, it may not be necessary to emphasize the diversity of individual strategies in speech production, which is now increasingly recognized and documented. Divergences extend on a continuum from transient through idiosyncratic to cross-dialectal. They can yield decisive insights into language structures and their evolution: for instance, analyses of speaker-specific strategies shed light on how different articulatory configurations are able to generate similar acoustic outputs for vowels (Johnson et al. 1993; Hoole 1999) and consonants (Guenther et al. 1999). Investigations into connected speech processes show that some speakers do produce complete assimilations or elisions in a given context while others do not (Nolan 1992; Ellis and Hardcastle 2002; Kühnert and Hoole 2004; Niebuhr et al. 2011b). Such findings on diversity depending on speakers and styles make a major contribution to shifting connected speech processes from a cognitive level of categorical feature-based operations to a level of basically gradual articulatory interactions. In turn, this offered a fresh perspective on such fundamental issues as the role of the phoneme in speech production and perception (Gow 2003; Niebuhr and Meunier 2011; Kohler and Niebuhr 2011). Speaker-specific differences sometimes stand in a close relationship of correspondence with trading relations as brought out by perception experiments. The perception of a well-established intonational contrast in German involves an interplay of peak alignment and peak shape (Niebuhr 2007a); production data confirm that peak alignment and peak shape are both used by speakers to signal the two contrasting intonation patterns (Niebuhr et al. 2011a). However, the 34 analyzed speakers differ in the extent to which they make use of the alignment and shape cues. While the majority of speakers use both F0-peak parameters to different degrees, a small group of speakers (15%) – the "shapers" – prefer to signal the two contrastive intonation patterns by means of peak-shape differences alone. Pure "aligners" are about twice as numerous.

Similarly, making syllables perceptually salient and indicating broad, narrow, and contrastive-focus pitch accents both involve a number of articulatory and phonatory cues; and besides the fact that these cues are used in language-specific ways (cf. Andreeva and Barry 2012), there are also differences between speakers of the same language. For example, some speakers make more extensive use of changes in F0 peak range and timing, whereas others prefer using local or global changes in the duration structure or variation in articulatory dynamics and precision (cf. Hermes et al. 2008; Cangemi 2013). Although some of these differences may actually be an artefact of the elicitation task, reflecting the speaker-specific degree to which the signalling of contrastive focus is coloured by emphatic accentuation (cf. Görs and Niebuhr 2012), there is no doubt that prominence, rhythm, and focus all involve individual differences. Recent analyses of Northern Frisian prosody showed that some speakers vary the perceptual prominence of syllables by lifting the F0 maximum of the associated pitch-accent peak, whereas other flatten and hence extend the F0-peak maximum. So, in addition to distinguishing "aligners" and "shapers", it may also be necessary to search and separate "lifters" and "flatteners" (cf. Niebuhr and Hoekstra 2014). Both lifting and flattening F0 peak maxima are suitable means to make high pitch stand out in the listeners' ears, and high pitch is well known to be an attention-attracting signal. Thus, this prominence-related example shows that we generally need a better understanding

of how acoustic parameters merge into the decisive perceptual parameters in order to explain and anticipate individual strategies and preferences.

A textbook example of a discovery based on the observation of cross-speaker differences is the study of Khmer by Eugénie Henderson. She worked mainly with one speaker, also checking the results with a second speaker; both were students at the School of Oriental and African Studies in London. Differences between the pronunciations of these two subjects helped her identify a major process of the historical phonology of Khmer: registrogenesis – the transphonologization of laryngeal oppositions on onsets.

> "The differences in usage lay chiefly in (1) the realization of the registers, and (2) the use in rapid speech of alternative forms such as those described on p. 172. Mr. Keng, as a philosophy student with literary and dramatic leanings, was aware of and interested in language from both the philosophic and aesthetic standpoints. His style of utterance was in general more deliberate and controlled than that of Mr. Mongkry, who as a student of economics was less concerned with language for its own sake. The two styles complemented each other well. Mr. Keng's style was helpful in that the different voice quality and manner of utterance of the two registers were clearly, sometimes startlingly, recognizable, even in fairly rapid speech, whereas Mr. Mongkry appeared often to make no distinction other than that of vowel quality. On the other hand, Mr. Mongkry's style of utterance was valuable for the ease and naturalness with which the alternative pronunciations proper to rapid speech were forthcoming" (Henderson 1952:149).

E. Henderson can be said to have been extremely fortunate to have come across two speakers who exemplified widely different pronunciations, one of whom was a strongly conservative speaker. However, in Pasteur's often-cited phrase (1915 [1939:131]), *"chance favours only the prepared mind"*: it is much to Henderson's credit that, rather than choosing one of the two speakers as her reference – which would have saved trouble –, she noted the differences and was able to identify the direction of change. Her findings set the key for a series of studies of register which proved fundamental to an understanding of the historical phonology of a great number of languages in East and Southeast Asia (Huffman 1976; Ferlus 1979; Edmondson and Gregerson 1993; Brunelle 2012; for a review see Michaud 2012).

2.4 The relationship between the consultants and the researcher

For field workers, the paramount importance of the relationship established with language consultants is well-recognized. *"In order to avoid disappointment and frustration, some time needs to be allocated for identifying [the consultants'] strengths and weaknesses, and most important, they themselves need some time to overcome shyness and insecurity and discover their own talents and interests"* (Mosel 2006:72; see also Mithun 2001). The example of Henderson's study of Khmer illustrates the fact that this process of mutual understanding does not necessarily take a very long time: rather, it is an issue of the investigator's attention to human subjects' personality, and to the stylistic preferences that contribute to shaping their pronunciation. Some phonetici-

ans choose to have the least possible contact with the experimental subjects taking part in their experiments; to us, this appears as a misguided interpretation of the notion of scientific objectivity. Scientific objectivity by no means requires the investigator to overlook such important parameters.

Good communication between the investigator and the participants in an experiment is essential to assessing to what extent differences found across the different participants' data sets reflect ingrained speaker-specific strategies, and to what extent they reflect different understandings of the tasks to be performed. The default hypothesis is that the experimental condition is the same for all speakers, and that differences in the recorded data therefore reflect cross-speaker differences. But different subjects may have interpreted the instructions differently, so that the differences in the data reflect in part the stylistic choices that they adopted: an experiment providing more precise guidance, such as a more explicit contextualization of the communicative setting that the experiment aims to simulate, may bring out greater closeness between speakers. In order to spot and interpret speaker-specific strategies, and to adjust the data collection procedure accordingly, the experimenter requires a trained ear, as well as a good command of the investigated language.

A sensitive definition of tasks also benefits greatly from exchanges with the subjects, especially when adapting a setup originally devised for another language. For instance, a study of Vietnamese prosody (Dô et al. 1998) initially calqued studies of Germanic or Romance languages, and attempted to contrast segmentally identical declarative and interrogative sentence pairs such as "Bao đi Việt Nam" ('Bao goes to Vietnam') and "Bao đi Việt Nam?" ('Is Bao going to Vietnam?'). This setup neglected the central role played by particles in conveying sentence mode in Vietnamese. Recordings were carried out in France. Bilingual speakers who had been living in France for many years had no difficulty in producing and differentiating the sentence pairs, as in French, whereas speakers who had just arrived and had little command of French practically refused to read such interrogative sentences. *"Either they spontaneously added a final particle, à, or they pronounced them in a very emphatic, exclamatory way, or on the contrary like the declarative counterparts"* (Dô et al. 1998:401). In less extreme cases, speakers may simply comply with the instructions, silencing any misgivings they may have – unless the investigator takes care to discuss the experimental setup with them.

Applying an experimental setting with new speakers, and on a new language, requires thorough re-examination of the method (a highly recommended reading on this topic is Vaissière 2004). An case in point is that of an experiment on tone identification, performed under fieldwork conditions: the investigators calqued a procedure that had been used for a national language, playing a signal from which segmental information had been masked and requiring listeners to select one of several real words (presented in written form) constituting minimal sets or quasiminimal sets. Speakers of the language under investigation had some command of the national language in which the words were presented to them, but the task of recognizing the written words, translating them mentally, and matching the tone pattern of the heard stimulus with their tonal representation of the minimal sets in their native language proved highly challenging, so that the participants' performance was poor; data for some of them had to be discarded altogether.

A side advantage of experimental setups based on a sensitive cultural and sociolinguistic contextualization is that they stand up to the highest ethical standards, and answer the concerns embodied in the guidelines for human subject research of the investigators' home institutions and funding agencies. Distress caused by culturally inappropriate recording tasks could be considered as a form of abuse of human subjects. Beyond formal guidelines, which can hardly anticipate the range of actual situations, the responsibility for relating with language consultants in the best possible way is ultimately the researcher's own; and ethical concerns coincide with scientific concerns. *"Ethical issues are embedded in a host of other '- ical' issues, such as methodological and technological ones"* (Grinevald 2006:347): adopting a culturally appropriate behaviour, valuing the knowledge that the consultants share with the investigator, giving a fair compensation for their time and effort, explaining the process of data collection and research, and preserving the collected data, are both ethical imperatives and important aspects of successful data collection.

3. The task

3.1 Recording settings and the issue of "laboratory speech"

The out-of-the-way setting of a recording booth can be conducive to out-of-the-way linguistic behaviour, in cases where the speaker lacks a real addressee or a real communicative task to perform. We were keenly reminded of this when participating as subjects in an experiment on foreign-accented English: the task consisted in telling the story of "Little Red Riding Hood" under two conditions, once with a child of age 10 present in the booth to serve as audience, and once alone in the recording booth, without an audience. Being familiar with recording studios, we did not expect to be deeply influenced by these different settings, but the difference proved considerable. It was excruciatingly difficult to flesh out a narrative without an audience; this was reflected in numerous disfluencies. Phonetic studies confirm that speakers behave differently when reading isolated sentences or monologues than when reading turns of dialogue together with another speaker. A comparison of read monologues and dialogues by Niebuhr et al. (2010) shows that, even though the style of the script was the same in both cases – an informal style, intended as a close approximation of conversational speech –, the read dialogues were prosodically closer to spontaneous dialogues (as recorded and analyzed by Mixdorff and Pfitzinger 2005) in terms of F0 level, declination and variability, speaking rate and phonation mode.

Researchers in phonetics are often aware of the potential distance between linguistic behaviour in the lab and outside, witness this reflection found in the introduction to "Intonation systems: A survey of twenty languages":

> "The majority of the work reported in this volume is based on the analysis of a form of speech which has come to be known, sometimes rather disparagingly, as "laboratory speech", consisting of isolated sentences pronounced out of context, usually read rather than produced spontaneously (…). An obvious question which needs to be answered is how far does variability in

> the situations in which speech is produced influence the results obtained under these conditions? To what degree do generalisations obtained from isolated sentences apply to more spontaneous situations of communication? (…) There is obviously still a great deal of work to be done in this area before we can even begin to answer these questions." (Hirst and Di Cristo 1998:43)

While it is often difficult to assess in detail the influence exerted by laboratory conditions, it is clear that the settings of a recording exert a decisive influence on a speaker's performance. The subjects taking part in an experiment have some expectations about the experiment, and some representations about what a phonetics laboratory may look like. They may feel called upon to adopt a specific style, in unpredictable ways. When confronting a microphone, some speakers may adopt a more formal, deliberate style of speech than the investigator aims to capture; visual details such as the distance to the microphone, and the presence of a pop shield hiding the microphone from view, all contribute to shaping the subject's experience, relating to highly personal factors such as their fondness or dislike for public addresses, and their degree of self-confidence in oral expression.

People maintaining online databases and language archives often find it difficult to elicit reasonably detailed metadata from the researchers: information about the speakers, the recording tasks, the time of recording… Researchers have a lot on their plate, and the task of documenting their data sets may appear to them as a distraction from research – no matter how interested they are in these data, whose importance to research they acknowledge in principle. For instance, an archivist asking researchers to indicate the time of day when each recording was made is likely to be considered too fussy. Yet there is evidence that this parameter exerts an influence on speech. Görs (2011) created a corpus of more than 30 German speakers, who read texts (i) early in the morning; (ii) at noon; and (iii) late in the evening. She found systematic prosodic differences as a function of the time of day. In the morning, speakers show a slower speaking rate and a lower average F0, as well as stronger glottalization at prosodic boundaries. Speaking rate and average F0 increase at noon; the same applies to the level of speech reduction. In the evening, average F0 is lower again; the speaking rate remains high, but with fewer speech reductions; and voice quality is overall breathier, among other differences.

The personal experiences and findings summarized in this chapter boil down to the self-evidence that communication is context-sensitive. 'Spontaneous speech' is not a homogeneous category; and 'naturalness' is not a straightforward criterion to assess speech data, since speech can be said to constitute a natural response to a particular setting, even in the case of "unnatural" settings. Ultimately, this means that speech recordings from one setting cannot be more natural than speech recordings from another setting. Speech data of any kind can be described as a natural response to the settings under which they were recorded; in this sense, 'naturalness' is not a relevant criterion to evaluate speech recordings (Wagener 1986). What matters is the investigator's in-depth understanding of the communication setting. Speech data recorded at the linguist's initiative under highly controlled laboratory conditions can offer an appropriate basis for research, provided the researcher ensures that the communication setting is well-defined, and is clarified to the speakers' satisfaction.

3.2 The range of recording tasks and cross-task differences

Salient differences are observed across different types of elicited speech material. The investigator should be aware of the implications of the choice of materials. For classifying and comparing basic types of speech material, it appears convenient to start from a six-way typology. It relies on the fact that recordings can be made with and without a dialogue partner and on a read or spontaneous (i.e. unscripted) basis[2]. In addition to these two binary parameters, isolated words/logatoms – typically monosyllabic nonce words like [bab], [pap] and [pip] – and isolated sentences should be regarded as two separate subtypes of read monologues. In this light, one may distinguish six (4+2) types of speech materials: isolated logatoms or words; isolated sentences; read monologues; read dialogues; unscripted monologues; and unscripted dialogues.

The methods behind these six types can be rated along various dimensions, among which we will discuss five: (i) degree of control over experimental variables (i.e. dependent and independent variables) as well as other variables (control variables); (ii) event density: the number of analyzable tokens per time unit; (iii) expressiveness; (iv) communicative intention: the speaker's concern to actually convey a message; and (v) homogeneity of behaviour: the probability that the elicitation condition is defined in such a way that it leads speakers to behave in a comparable way. The diagram in Figure 1 is an attempt to represent how the six types of methods perform in terms of these five dimensions. The performance is given as a simple relative ranking from 1 (worst) to 6 (best), based on notes and findings in the literature and on our own experience.

First of all, Figure 1 brings out the evident fact that there is no ideal method/material that performs best on all dimensions. Methods based on read speech allow for a high degree of control and yield a relatively high event density. However, they tend to dampen speaker involvement and hence do not perform well in terms of expressivity and communicative intention. This is particularly true for read monologues – such as newspaper texts (Amdal and Svendsen 2006) or prose (Zellers and Post 2012) – as well as for readings of logatoms and isolated sentences. Isolated logatoms allow for an even greater control in terms of prosody than isolated sentences (cf. Cooke and Scharenborg 2008); and as they are shorter, the event density is also higher. Dialogues increase expressiveness, informality and communicative intention (see, e.g., the evidence from Dutch presented by Ernestus 2000), and the presence of a dialogue partner stabilizes the speech behaviour of the recorded subject (Fitzpatrick et al. 2011). A richer semantic-pragmatic context has the same stabilizing effect (this is referred to as the "richness principle" in Xu 2012). This is another reason why one speaker's homogeneity of

[2] 'Unscripted speech' is in our opinion a more precise term than 'spontaneous speech', because all it means is that speakers do not produce predetermined utterances. The attribute 'spontaneous' can easily be misinterpreted as 'impulsive', 'instinctive', or 'automatic' and hence associated with a particularly emotional or agitated way of speaking, which can, but need not be applicable. However, 'spontaneous speech' is the more established term, which is why we use the two terms interchangeably, focussing on 'unscripted' speech in the context of the present section (3.2) because of its focus on the nature of the tasks entrusted to the speakers.

behaviour paradoxically increases from isolated logatoms and sentences through read monologues and dialogues to unscripted dialogues.

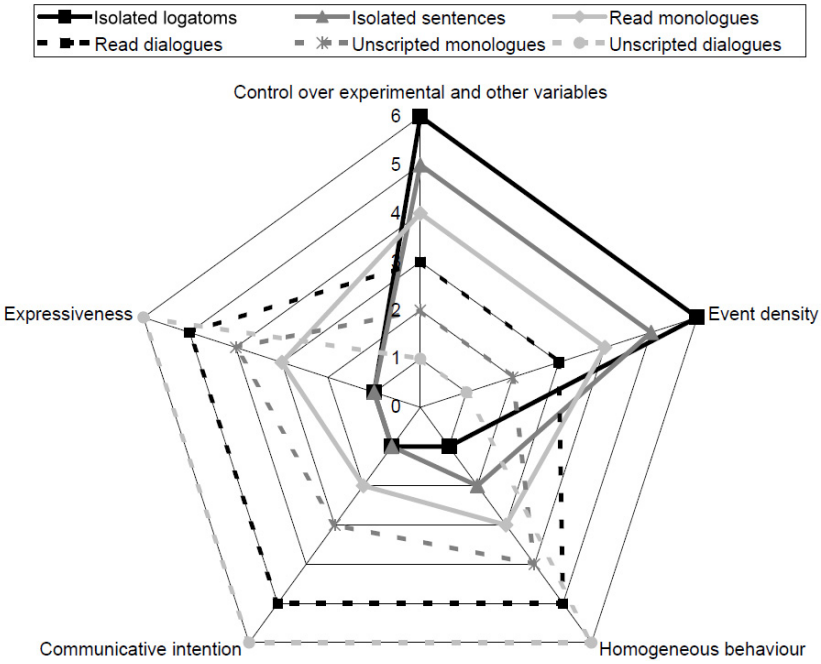


Figure 1. Ranking of six basic types of recording tasks on five dimensions that represent characteristics of experimental designs and speech communication.

A homogeneous speech behaviour is not only important for cross-speaker comparisons, but also with regard to replicability. Some research questions require the same speech material to be recorded twice with different recording devices. For instance, to study nasality, fiberscopy and airflow provide complementary perspectives, but recording both at the same time is impractical. To overcome this difficulty, it is possible to record nasal airflow and images of the velopharyngeal port separately, in two sessions, and to time-align the two data sets on the basis of landmarks on the acoustic signals recorded under both setups. For these post-aligned data to yield valuable results, the subjects' performance under both setups must be as close as possible to complete identity.

The disadvantages of dialogues are their lower event density (i.e. conducting a study becomes much more time-consuming) and the lack of control over experimental and other variables, particularly in the case of unscripted dialogues. Prototypical examples of unscripted dialogues are free conversations without any guidelines or topic specifications (cf. CID, Bertrand et al. 2008) or recordings of TV or radio broadcasts (cf. RUNDKAST, Amdal et al. 2008). Broadcast speech often constitutes the backbone of the large databases used in speech processing; in-depth statistical treatment of these

databases sheds new light on sound systems (see, e.g., Gendrot and Adda-Decker 2007), but a limitation for prosody research is that broadcast speech is tilted towards a relatively narrow range of styles.

On the whole, read dialogues (like the KIESEL corpus described in Niebuhr 2010) seem to strike a reasonable compromise between the conflicting demands of symmetry, on the one hand, and ecological validity, on the other. Read dialogues combine an informal, expressive speaking style – which can be enhanced by using a corresponding orthography and font type – with relatively high degrees of communicative intention, homogeneous behaviour, event density, and a relatively high degree of control over experimental and other variables. The control of the experimenter in read dialogues extends to the semantic-pragmatic context, which is why read dialogues are ranked higher than unscripted monologues in terms of the homogeneous behaviour of speakers. Moreover, the control over the semantic-pragmatic context can also be exploited to elicit specific melodic signs on key words. Finally, by selecting and combining appropriate dialogue partners, experimenters can make use of phonetic entrainment, in order to direct the speech behaviour of the two speakers into a certain direction. For example, if a privy dialogue partner is instructed to speak in a highly-reduced or very expressive fashion, then this speech behaviour is likely to rub off to some degree on the naïve dialogue partner. This entrainment strategy is of course also applicable to unscripted dialogues. Even if there is no privy dialogue partner, it is possible for the investigator to elicit different speech behaviour from his/her subject simply by matching the same speaker with different dialogue partners.

Figure 1 only constitutes a convenient means to represent several parameters, in order to compare widely different types of methods and materials. In detail, there is of course much more to say about each of these methods, and about strategies to improve each type of elicited speech material along several dimensions.

For instance, in unscripted dialogues, the event density can be increased, and controlling elements added: this is the famous case of 'Map tasks' (Anderson et al. 1991), which make it possible to introduce key words. A privy dialogue partner in a Map task recording can furthermore trigger certain communicative actions of the speaker. For example, in the study of Görs and Niebuhr (2012), a privy dialogue partner repeatedly pretended misundertandings, allowing for the elicitation of key words with narrow-focus intonation patterns. A privy dialogue partner can also help a speaker overcome the intimidating effects of recording-booth settings, before or during the actual recordings (see, e.g., Torreira et al. 2010). Most map task data have been elicited in Indo-European languages like American and Australian English, German, and Italian. However, given its success, its transfer to a wide range of languages appears feasible and promising.

Tasks similar to the Map task are the 'Shape-Display task' (cf. Fon 2006), or the 'Appointment-Making task' and the 'Videotask', which were used in the collection of German data by Simpson et al. (1997), Peters (2005), and Landgraf (2014). While the appointment-making scenario generates a high number of day, time, and place expressions, the Videotask scenario, in which the speakers first see two slightly divergent video clips of their favourite TV series and then confer to find the differences, additionally exploits an emotionally charged common ground between the speakers in order to enhance their expressiveness. The Videotask idea can be imple-

mented with very different types of broadcasts, including cartoons, and it has been shown for the latter type of broadcasts that the Videotask is basically able to trigger phonetic entrainment, just as real everyday dialogues (cf. Mixdorff et al. 2014).

The appointment-making scenario and the Videotask scenario are representatives of two different strategies that have been used to elicit unscripted dialogues: role-play tasks and quiz tasks. The appointment-making scenario is a typical role-play task. Other role-play tasks are a 'Sales Conversation' (cf. Ernestus 2000) and a 'Design-Team Project Meeting' (cf. http://corpus.amiproject.org/). Further typical quiz tasks are, for example, the 'Picture-Difference Task' (cf. Turco et al. 2011) and the 'Joint Crossword Puzzle Solving' used by Crawford et al. (1994). The Map task belongs to yet another type of task that may be called 'Instruction-Giving task'. Other examples are the 'Tree-Decoration task' (cf. Ito and Speer 2006), the 'Picture-Drawing task' of Spilková et al. (2010), the 'Card Task' (Maffia et al. 2014), and the 'Toy Game'. Unlike the Map task, the Toy Game has already been successfully applied for eliciting natural conversation and prosody in the field. It is a simple, portable set up developed in conjunction with the 'Dene Speech Atlas' (http://ling.rochester.edu/people/mcdonough/dnld/JMcDonough/dene-speechatlas.html)

In the Toy Game, two players sit on opposite sides of a table with an occlusion between them. On a table in front of each player is a sheet of paper and some small objects (toy animals, cups, fences, etc.). The sheets of paper have three shapes drawn on them: circle, square and triangle. Both sheets are the same. The goal of the game is for both players to have the same arrangements on their sheets. How players proceed can be given some leeway, but in general players accomplish the task by taking turns asking questions, starting with one player, then the second player gets a turn. Recording begins well before the game begins, because the interaction that takes place in agreeing on names for objects is extremely useful for later analysis. The Toy Game is typically played three times with increasing complexity. The first game is a short warm up game. In the second game, the two players still have the same small number of items but in different arrangements. In the third game, there is an increase in both the number of types of toys and the number of tokens of each toy.

It may be assumed that role-play tasks perform better in terms of event density and experimental control (richness of semantic-pragmatic context). But they are outperformed by quiz and instruction-giving tasks with respect to expressiveness and communicative intention. In instruction-giving tasks, it also seems easier to foist the speaker on a privy dialogue partner, who then takes the role of the instruction receiver.[3]

[3] It should not be forgotten in this context that the experimenter's creativity to control the behavioural responses of subjects needs to be channelled by ethical considerations. These are formalized as ethical guidelines at some research institutions, but in practice the bulk of the real responsibility rests with the investigator. Foisting a privy dialogue partner on speakers is a type of deception; instructions that lack crucial information or deliberately provide misinformation in order to distract the subjects from the actual aim of the experiment are also problematic. Such strategies are common practice in many fields of research, most prominently in psychology, and tend to be considered as ethically acceptable, so long as the behavioural responses are interpreted in view of the distractor strategy. Other elicitation scenarios can create serious conflicts by implicitly

Quiz tasks are basically applicable for speakers of very different cultures in the field and in the lab. Moreover, when the different stimuli are shown (simultaneously or subsequently) to the same speaker, quiz tasks can also be used to elicit monologues, which is not possible with the role-play or instruction-giving tasks.

In a similar way as for unscripted dialogues, unscripted monologues can be based on retelling picture stories (cf. Iwashita et al. 2001; Mosel 2006, 2011) in order to include key words and/or a semantic-pragmatic context frame that ensures homogeneity of the speech behaviour. Alternatively, speakers can be asked to recite lyrics, poems or traditional texts that they know; this can make them feel more comfortable compared with previously unknown picture stories, but recitation constitutes a highly specific activity, often associated with specific styles. These tasks and similar other tasks as well as suitable elicitation material like "The pear story" or "Frog, where are you?" are explained in more detail in the book "Questionnaire on Information Structure (QUIS): Reference Manual" by Skopeteas et al. (2006). When eliciting monologues, it is useful for the speaker to have an addressee. Even if s/he does not say anything, subjects feel more comfortable and produce speech in a different way when the act of speaking is a social activity (cf. 3.1 and Fitzpatrick et al. 2011).

Another way to elicit expressive unscripted monologues is to record speakers during or after computer games (cf. Mixdorff 2004). This creates a fairly specific semantic-pragmatic context frame and allows for the elicitation of key words. Johnstone et al. (2005) even controlled the outcome of the games (win or failure) to stimulate positive and negative emotions. Similar manipulations of the environmental conditions were used by Maffia et al. (2014) for eliciting expressiveness and emotions during Card-Task dialogues.

Finally, in accordance with the conclusion in 2.1, Figure 1 suggests that investigating a research question on speech production should involve several recordings tasks, starting from isolated logatoms or sentences, through read monologues or dialogues to unscripted dialogues. The 'ShATR Corpus' (Crawford et al. 1994) and the 'Nijmegen Corpus of Casual French' (Torreira et al. 2010) are good representatives of such a multiple-recordings strategy. Pilar Prieto's "Grup d'Estudias de Prosodia" (GrEP) has developed various innovative methods for the contextual elicitation of prosodic and gestural patterns. Some of these methods are summarized in Prieto (2012) and are worth considering for those who are interested in studying the many interrelations between prosody and gestures, since most traditional tasks are not suitable for this purpose.

3.3 Within-task differences

In addition to cross-task differences, there exist multifarious within-task differences. The present survey focuses on artefactual within-task differences. The examples below show the usefulness of developing an awareness of these pitfalls, in order to devise strategies to overcome them.

---

forcing speakers to choose between violating linguistic or cultural norms and questioning the authority of the experimenter. This point is further detailed in 3.1.

A typical example of an intentional within-task difference is when speakers are asked to perform the same reading task with the same material at different speaking rates. Typically, speakers are asked to read a set of isolated sentences at their normal rate and then additionally very slowly and/or as fast as they can. The sentences are either directly repeated at different rates, or the rate differences are produced blockwise. Such a within-task difference is used among others as a means to get a dynamic view of the realization and timing of intonation patterns, of speech rhythm, and of connected-speech processes such as assimilation. A frequent (and not always explicitly noted) by-product is that the F0 level is raised when speakers produce the presented sentences as fast as they can (cf. Kohler 1983; Steppling and Montgomery 2002; Schwab 2011). Interestingly, a raised F0 level is also typical of speech produced under high cognitive load or (physical or mental) stress (cf. Scherer et al. 2002; Johannes et al. 2007; Godin and Hansen 2008). It is reasonable to assume that the instruction to read and produce sentences at the fastest possible rate requires a higher cognitive load and puts speakers under stress. This requires great caution when examining data collected through deliberate speaking-rate variation: part of the phonetic differences between the speaking-rate categories set up by the experimenter may reflect differences in cognitive load and stress level. This illustrates the introductory statement that speakers are no "vending machines": the implications of instructions for within-task differences must be carefully considered.

Furthermore, fatigue and boredom can soon seep in when going through an experimental task. This can lead to unintentional within-task differences. Repetition detracts greatly from illocutionary force. This point is brought out by the thesis of Kohtz (2012). Her aim was to investigate if and how subtypes of the common sentence-list elicitation task affect the production of nuclear accent patterns. To this end, nine speakers read two individually randomized lists of 200 sentences presented separately. The sentences in both lists ended in disyllabic sonorous target nouns produced with rising nuclear accents on the initial (lexically stressed) syllable. The target nouns in list A were embedded in the classic carrier sentence "The next word is ___". The sentences of list B had a lexically more variable NP-VP-PP structure (such as "The cat sleeps on the sofa"), in which the target nouns occurred at the end of the PP. Kohtz found consistently higher F0 variability and intensity level for list B than for list A. The speaking rate on the other hand was higher and consistently increased for the list-A as compared to the list-B sentences. However, the most crucial within-task difference applied to both lists and is displayed in Figure 2. The more sentences the speakers produced, the earlier and more stably aligned were the rising nuclear accents in the sentence-final target words. Only 50 sentences were already sufficient to halve the standard deviations for the alignment of rise onset and peak maximum relative to their respective segmental landmarks, i.e. the beginning of the accented syllable or its vowel. The standard deviations of the final 20 sentences were up to 85% smaller than those of the initial 20 sentences.

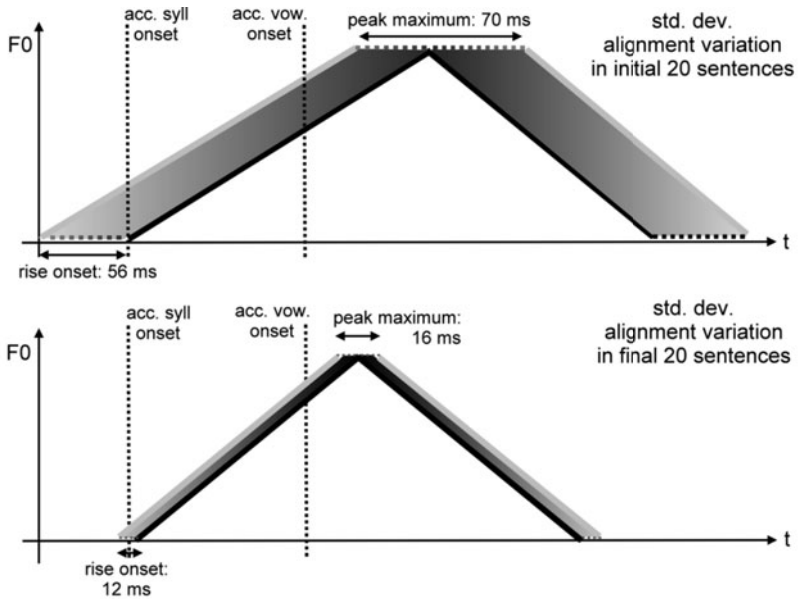


Figure 2. Schematic illustration of the fading alignment variation of F0 rise onset and peak maximum from the initial 20 to the final 20 sentences in the study of Kohtz (2012) on German intonation patterns.

Speech production is essentially a muscular task; and as for every muscular task, repetitive training in a controlled, undisturbed environment reduces variability and increases precision, efficiency, and speed. The successive changes in the nuclear accent patterns of Kohtz must be seen in this light. Lists of 200 isolated similar sentences are an ideal training ground. It gradually reduces speech production to a mere muscular exercise and allows speakers to (unconsciously) train production and timing of their rising nuclear accents, undisturbed by syntactic and prosodic variation, communicative purposes, and interference of voiceless segments so that they can achieve an extraordinarily high level of precision and constancy.

So far, studies like that of Kohtz (2012), which critically evaluate and compare elicitation methods, are still rare. However, such studies are highly relevant to assess research results, and even to look back on the development of strands of research within phonetics/phonology. Among other implications, Kohtz's findings raise the thought-provoking issue of the extent to which the great amount of attention attracted by 'segmental anchoring' (cf. Ladd 2003) is due to the nature of the data sets under examination: segmental anchoring is primarily (and even somewhat exclusively) investigated on sets of read sentences. It is possible that the essence of segmental anchoring, i.e. an extraordinarily high level of precision and constancy in accent-contour alignment, does not show up for other kinds of elicitation tasks or for shorter lists of sentences. Initial evidence in favour of this possibility comes from the study of Welby and Loevenbruck (2006) on the segmental anchoring of rising accent contours in

French. Welby and Loevenbruck elicited a small corpus of less than 50 isolated sentences, as well as a paragraph corpus with a similarly small number of target sentences that were framed by syntactically and phonologically diverse context sentences. Although the target sentences in this procedure are equally carefully controlled as in all other studies on segmental anchoring, it is obvious that the procedure does not allow speakers to intensively train the production of their accent productions. Accordingly, the alignment patterns found by Welby and Loevenbruck showed "*a fair amount of variability […] within and across speakers […], in contrast to the very stable 'segmental anchors' found for other languages"*. Moreover, *"comparisons between the two corpora also reveal intra-speaker variability […]. There were almost no significant results for a given speaker that held across the two corpora"* (Welby and Loevenbruck 2006:110). This led to the assumption that tonal alignment in French is guided by wider anchorage areas in the segmental string. However, the actual reason why Welby and Loevenbruck found anchorage areas rather than specific segmental anchor points may be a methodological one.

In conclusion, the advantages of sentence list elicitations are undeniable. They yield a high number of relevant tokens from any number of speakers in a short amount of time and with a high degree of segmental and prosodic control (cf. Figure 1). Yet, it seems fairly evident that the "instrument of spoken language" can become blunt if your list contains too many sentences. Sentences that share a similar morphosyntactic structure may further accelerate the erosive process.

4. The recording

4.1 The necessity to use professional recording equipment

On a technical note, one must emphasize the necessity of using professional recording equipment and of making and sharing sustainable recordings, as opposed to the widespread practice of recording "disposable data". It is useful to think in terms of future uses of the data, beyond one's immediate research purposes. Sampling at 16,000 Hz may seem fine for drawing spectrograms, since a display from 0 to 5,000 or 8,000 Hz is sufficient for the study of vowels, and for spectrogram reading. But if at some later date the original team of researchers (or other colleagues to whom they kindly communicate their data) wish to look at fine phonetic details, for instance allophonic/intonational variation in the realization of fricatives such as [s] and [ʃ], the sampling rate will prove too low: when separating [s] and [ʃ] on the basis of center of gravity measurements, it was found that the best results obtained when frequencies up to 15,000 Hz were included (Niebuhr et al. 2011b). Acoustic data may also be used at some point to conduct perception tests, and a sampling rate of 44,100 Hz is designed to capture all the frequencies that the human ear can perceive. Adopting such a rate (or a higher one) therefore seems advisable, especially since digital storage of files in this format is now technically easy.

The same "the more the merrier" principle applies to data other than audio as well. Electroglottography (Fabre 1957; Abberton and Fourcin 1984) allows for high-precision measurements of the duration of glottal cycles, and for obtaining other infor-

mation about glottal behaviour, such as the glottal open quotient/closed quotient. In view of the most common uses of the electroglottographic signal, a sampling rate of 16,000 Hz would seem to be more than enough. Rather unexpectedly, this turns out to be too low for some research purposes. Research on the derivative of the electroglottographic signal brings out the significance of peaks on this signal: a strong peak at the glottis-closure instant, marking the beginning of the closed phase, and a weaker one at opening, marking the beginning of the open phase (Henrich et al. 2004). The closing peak is referred to as DECPA, for Derivative-Electroglottographic Closure Peak Amplitude (Michaud 2004b), or as PIC, for Peak Increase in Contact (Keating et al. 2010). When measuring the amplitude of this peak, signals with a sampling frequency of 16,000 Hz do not provide highly accurate information. The peak is abrupt; at a sampling rate of 16,000 Hz, widely different values are obtained depending on the points where the samples are taken at digitization. Technically, a high-precision measurement of DECPA can be obtained with a signal sampled at 44,100 Hz, with interpolation of the signal in the area of the closure peak (following a recommendation by Wolfgang Hess, p.c. 2004).

Likewise, concerning bit-depth, 24-bit may seem way too much by present-day standards, but it can still reasonably be considered, since it gives a great margin of comfort for digital amplification of portions of the signal that have extremely low volume. One bit more improves the signal-to-noise ratio by 6 dB. So, especially for recordings outside the laboratory and/or if speakers are a little further away from the microphone, as in the case of Figure 3, 24-bit should be the rule rather than the exception.

Figure 3. Recordings of narratives elicited at Mr. Vi Khăm Mun's home in Tuong Duong, Nghe An, Vietnam.

The choice of microphone is particularly critical for your recording. Compared with normal, omnidirectional microphones, head-mounted microphones or microphones with

cardioid or shotgun characteristics are usually more expensive. However, the investment pays off, as the directionality of these microphones helps keep the two channels of dialogue partners distinct, even if the dialogue partners are not physically separated, but sit face-to-face one or two meters away from each other, as is shown in Figure 4. Moreover, irrespective of whether dialogues or monologues are recorded, directional microphones help reduce environmental noise in your recordings, especially under fieldwork conditions. Head-mounted microphones have the further advantage that the distance between speaker/mouth and the microphone remains constant. Hence the intensity level is independent of the speaker's head or body movements. Microphones should always be equipped with a pop filter and pointed to the larynx rather than the mouth of the speaker, cf. Figure 4. The orientation towards the larynx has no negative effect on the recorded speech signals, but further contributes to dampen popping sounds. It is also possible to use a windshield.

Figure 4. Recording of scripted dialogues conducted in the sound-treated room of the General Linguistics Dept, Kiel University, May 2004.

Some institutions do not have high-fidelity portable devices for field workers; others only have short supplies of them, sometimes without technical staff to manage these fragile equipments, and so they do not lend them to students. Equipment is expensive. However, whenever possible, students should consider investing in their own equipment, which they will know well, and which will be available to them at any time. The purchase of recording equipment could be considered on a par with the purchase of a personal computer: while it may appear as an unreasonable demand on a student budget (especially in countries with low living standards), having one's own equipment typically increases the quality of one's data, and of the research based on them. For fieldwork, the cost of the equipment should be weighed against the overall expenses of the

field trip(s) and of the months or years of research time spent annotating and analyzing the data. In principle, speech data have an endless life expectancy, and endless reusability. Seen in this light, the common practice of recording MP3 files – with lossy compression – from a flimsy microphone (such as the internal microphone of a laptop or telephone) hardly appears as an appropriate choice.

Monitor your recording carefully, especially if you are not yet thoroughly familiar with the equipment. Prevention is better than cure; in the case of audio recordings, there is simply no way to 'de-saturate' a clipped signal in a satisfactory way, or to remove reverberation as one would brush off a layer of dust. Sound engineers have to make choices and compromises in the complex process of tidying historic music recordings; for the acquisition of new data, you should get a good signal from the start, and limit its processing to volume amplification, without special effects.

Needless to say, these remarks about speech signals can be extended to other types of data, such as video recordings.

4.2 Selection of subjects

"It is a truism but worth repeating that different informants have different talents. Some are truly excellent at explaining semantic subtleties, while others have deep intuitions about the sound structure of their language" (Dimmendaal 2001:63). For some experi-mental purposes, subjects with an awareness of linguistic structures or even of linguis-tic theory may be appropriate; for other purposes, subjects with such an awareness are best avoided. Producing spontaneous speech in the lab and producing spontaneous-sounding read speech both require a certain extroversion, fluency, language compe-tence, and self-confidence; speakers should be pre-selected accordingly.

When dialogues are to be elicited, a deliberate selection and pairing of speakers is also important with respect to phonetic entrainment. Additionally, if the dialogue partner is a good friend, this greatly helps creating a relaxed, informal atmosphere for the recording.

Concerning the speaker sample, while four or five speakers constitute a good begin-ing for a reasonable sample of a well-defined social group, it should be kept in mind that they do not represent the full complexity of the language at issue (in particular, its sociolectal complexity). During analysis, compare within-subject means, and – if ne-cessary – create sub-samples before you calculate overall means for each measurement.

Let your speakers/informants fill out questionnaires that collect as detailed informa-tion (metadata) as possible – not just the three usual suspects: age, gender, and home town. Rather, type and amount of musical experience, level of education as well as smoking habits and the like should also be asked for. You could even include the question "How do you feel today?", stating explicitly that answering this and all other questions is voluntary, and that the investigator will be responsible for ensuring that these pieces of information are not made public.

For practical reasons, phoneticians often record speakers living away from the area where the target language (or dialect) is spoken. The consequences of language contact can be reduced by selecting people who have recently arrived from their homeplace, but the investigator should remain on the watchout for effects of language contact none-

theless and be additionally very specific about linguistic life course in the questionnaire. It seems safer to select subjects who experienced the smallest possible amount of language/dialect contact. There is little hope of factoring out interferences between languages when employing bilingual or multilingual speakers in phonetic studies, since the type and extent of interference varies according to numerous parameters that include age and the place of residence (Watson 2002:243).

In order to assess the expressiveness or extroversion of your speakers, we suggest making use of existing scales and questionnaires from the field of psychology (e.g., Gross and John 1997). Concerning recording settings, provide copious detail, including pieces of information that may seem anecdotal or irrelevant, such as the time of day.

4.3 Special precautions when using written prompts

Written prompts are a major source of artefacts. To linguists working on languages without a written tradition, it is obvious *that "speaking and writing are conceptually different activities, and so is a language in its spoken and written form"* (Mosel 2006:70). Linguists working on national languages may have less awareness of this central point. A study of colloquial Khmer reports that, *"in a pilot experiment, it was determined that participants had a difficult time producing colloquial variants when presented with visual primes – imagine being presented with the written sentence <I am not going to...> but being instructed to produce 'I ain't gonna...' – so instead a system was devised where the experimenter prompted the participant orally with the Standard Khmer form, whereupon the participant would provide the colloquial variant"* (Kirby 2014). Such precautions are of the greatest importance to obtain reasonably homo-geneous data, otherwise the speaker's behaviour may fluctuate in unpredictable ways.

Keeping these difficulties in view, it is possible to create dialogue texts that integrate common reduction phenomena in the orthographic representation. Let your carefully selected and paired dialogue partners practice the texts in advance; allow them to adjust the texts slightly to their own way of expression by introducing, omitting or replacing words and phrases. Conceding this flexibility to speakers has proven effective to increase the comfort of speakers, which then positively affected the expressiveness and informality of their way of speaking (Kohler and Niebuhr 2007; Niebuhr 2010, 2012). A further means to control the way of speaking is the font type of the written prompts. According to the experience of the first author, expressiveness and informality are best elicited with font types other than the businesslike Times, Arial, Calibri and Tahoma fonts.

Be careful with translations. Translating experimental materials and instructions requires all the precautions usually associated with translation, which is a profession on its own. For example, the second author was asked on several occasions times to read question-answer pairs in French such as "Qui est allé au restaurant? – Jean est allé au restaurant", which had obviously been translated from English ("Who went to the restaurant? – John went to the restaurant"). Question-answer pairs like the one above aim at eliciting narrow focus. They aim to elicit a realization in which the name "Jean" stands out as the informative part (emphasis) whereas the rest of the sentence is backgrounded (post-focus compression). Whatever the validity of the original English,

its translation sounds decidedly weird to native speakers of French: more appropriate answers would be simply "Jean", or "c'est Jean" ("it is Jean"), or – at a push – the cleft sentence "C'est Jean qui est allé au restaurant". Bad translation also ignores cultural factors. For example, the sentence "He decided to move house, but not to leave the town" may make good sense to speakers of American English, but it becomes odd when translated and used for French speakers, since the population density and urban structure of France is very different from that of the United States. Practical and detailed instructions on the use of translations in fieldwork are provided by Mosel (2011) and references therein.

Furthermore, monotonous tasks are to be avoided when using written prompts. In order to be comparable, utterances do not only need to be identical in their written form: *"most importantly, they have to be performed with the intention of achieving the same illocutionary act"* (Himmelmann 2006:168). Especially in experiments with a larger number of cross-combined independent variables it is often necessary to elicit numerous targets (e.g., words) with specific phonetic properties in prosodically controlled environments. In addition to artefacts in the form of 'list intonation', readers may spontaneously establish semantic/pragmatic relations between individual sentences, interpreting them as successive episodes within a single narrative, as it were – even if these sentences are presented on separate sheets of paper, or on separate slides shown on a computer screen. For example, the sequence "Peter came by car" - "Meghan came by bus" - "Steve came by boat" may cause "bus" and "boat" to be realized with prosodies of contrastive topic. In the sequence "The plate is on the table" - "The glass is on the table", "table" becomes given information, and "glass" is likely to be realized in contrast to "plate". Unless this is controlled for at the stage of data collection, wavering interpretations of the recording task will be treated as random variance at the stage of statistical analysis.

As the largest prosodic changes seem to occur after the fiftieth sentence (see 3.3), it seems safe to use lists of less than fifty sentences per session. Repetitions of the same sentence within the same session is to be avoided – or at least the investigators should be aware of the potential bias introduced by this repetition.

Isolated syllables should be carefully randomized, to avoid contrast effects similar to those described above. A sequence of syllables arranged by vowel, such as "ta, ma, ba, da, na, pa, ra…" will lead to more attention being focused on the realization of the consonant than on that of the vowel; and the opposite bias will be present for sequences arranged by consonant: "ta, tu, to, ti…".

One way to limit such prosodic artefacts consists in interspersing dummy sentences in the sequence of sentences to be recorded, so that successive sentences will be as unrelated as possible. But this increases the length of the task, and hence the ever-present risk of fatigue, without providing any guarantee that the speaker will not invent links between successive sentences. Using dialogues appears a more powerful solution.

4.4 Training, dummy-runs, and debriefing techniques

Training of the subjects needs to be handled with care. A widely-cited article about "Intonational invariance under changes in pitch range and length" is based on an experiment in which *"...the pitch range instruction was varied in 10 steps, and six to*

eight repetitions of each pattern in each pitch range were recorded. In both of the experiments to be described, 'degree of overall emphasis or excitement' was the term used in the subjects' instructions, and the kind of variation desired was illustrated by example" (Liberman and Pierrehumbert 1984:169). Four speakers were recorded, including the two coauthors. *"For subjects other than the authors, the desired intonation patterns were demonstrated by example before the experiment, and the ability of the subjects to produce them naturally was checked"* (p. 172). This formulation exemplifies the problems mentioned above (cf. 3.1) concerning the notion of naturalness. Additionally, one may have a couple of minor quibbles about data collection here, concerning (i) the use of a metalinguistic indication, "degree of overall emphasis or excitement", allowing for a broad range of interpretations, and (ii) the example set by the co-authors for the other two speakers. These go a long way towards explaining why the admirably clear-cut result obtained in that study – namely, that the terminal point of the F0 curve remains almost unchanged, whereas the highest F0 value is proportional to the degree of emphasis – could not be replicated in a later study (Nolan 1995).

Such salient artefacts are sometimes identified clearly by the community of phoneticians, leading to the adoption of new principles: it would not currently be considered good practice for linguists to report analyses based on their own speech data. This general principle is useful, but should be complemented by investigators as befits each specific experimental setup. An ideal towards which it may be useful to tend would consist in making laboratory experiments a mutual learning and teaching process for all people involved – like linguistic fieldwork. Dummy-runs (or a warm-up time for "spontaneous" – i.e. unscripted – conversations) and training can allow for this communicative process to take place. It is for the investigator to make adjustments and preparations carefully before the recording starts, in order not to distract the subjects' attention during the experiment. For instance, if the data are to be used in fine-grained acoustic analysis, it may be useful to instruct the speakers to avoid shuffling their feet or rubbing their hands on their clothes too vigorously during recording; but if these gestures are habitual for the speaker, making conscious attempts to suppress them takes up part of the speaker's attention, and may have consequences such as an increased amount of disfluencies. *"The importance of good recording needs to be balanced against the importance of keeping the participants at ease"* (Souag 2011:66).

Debriefing is a useful (though currently nonstandard) way of finding out more about the subjects' interpretation of the task and the evolution of their behaviour in the course of the session. Participants may sometimes point out accidental omissions: a speaker of French recruited for a recording of nasal vowels became aware of the absence of any example of /œ̃/, which in his speech contrasted with /ɑ̃/, /ɛ̃/ and /ɔ̃/. This would not have been detectable on the basis of the recordings, and inclusion of data from this speaker would have detracted from the reliability of the results, since the study assumed a three-way contrast between /ɑ̃/, /ɛ̃/ and /ɔ̃/. In retrospect, this aspect of the phonological system should have been examined individually for each potential participant before they were selected to conduct recordings.

Debriefing plays a central role in the 'Kieler Sammlung Expressiver Lesesprache' (KIESEL Corpus, i.e. 'Kiel Collection of Expressive Read Speech', presented in Niebuhr 2010). The aim is to achieve a high degree of expressiveness in read speech.

Read speech allows for segmental and prosodic control of the data; expressiveness is approached by instructing the speakers to judge each other's production performances, and repeating the dialogue until they are both satisfied and agree that they have produced a dialogue that resembles their colloquial, everyday speaking style. This setup yields encouraging results, despite its limitations.

4.5 Minimizing the recorder's paradox

Every recording situation will inevitably raise the speakers' awareness about the way they speak. This, in turn, can make speakers change their speech behaviour, with the consequence that the analyzable object diverges from the actual research object. Our goal is to understand speech communication, and speech recordings make it at the same time easier and harder to reach this goal. Xu (2012) calls this the "recorder's paradox". In order to reduce influences of the recording situation, you may use headmounted microphones – which, unlike table microphones, are not present in the speaker's field of vision –, select experienced speakers, use dialogue rather then monologue scenarios (pairs of speakers distract each other more easily from the recording situation), hide or store away technical equipment, and avoid dark or dark-coloured recording rooms. For research questions that require a high degree of self-revelation from the speakers, as in the case of expressive or dialectal speech or when small children are to be recorded, it can sometimes be even better to conduct the recordings in silent rooms of the speakers' own homes, as can be seen in Figure 5. Suitable rooms have as few plane and sound-reflecting surfaces as possible (e.g., bed rooms or living rooms). The remaining plane and sound-reflecting surfaces can be covered with bed sheets, large towels or similar pieces of household linen (cf. Fig.7). Bookcases and bookshelves also improve a room's acoustic qualities.

Figure 5. Speakers of the endangered North Frisian dialect Fering produce isolated sentences and scripted dialogues; recordings were conducted in the speakers' homes on the small island Föhr in the Northern Sea off the German coastline, cf. Niebuhr and Hoekstra (2014).

In fieldwork in small villages, in typically less-developed area, it is uncommon to have access to a room that is padded with bookshelves or soft furniture. Bare rooms with cement or tile flooring are common, as exemplified in Figure 6(a) by the kitchen of a Naxi farm in Yunnan, China. In such a room, there is an amount of reverberation that makes it difficult to make out on a spectrogram to what extent an intervocalic stop is voiced: the complete silence during the closed phase of an unvoiced stop is masked by reverberation, which results in noise on spectrograms even at points where one would expect complete silence. Doing a recording out in the open is hardly an option: a field or a pasture are acoustically fine, as there is close to zero reverberation, but the speaker and the investigator are then exposed to the elements – including the wind, which can ruin a recording if no windscreen is available. The comical scene of speech data acquisition is also mercilessly exposed to the gaze of passers-by and the curiosity of wandering animals, making it pretty hopeless to achieve the required degree of concentration on the part of all concerned.

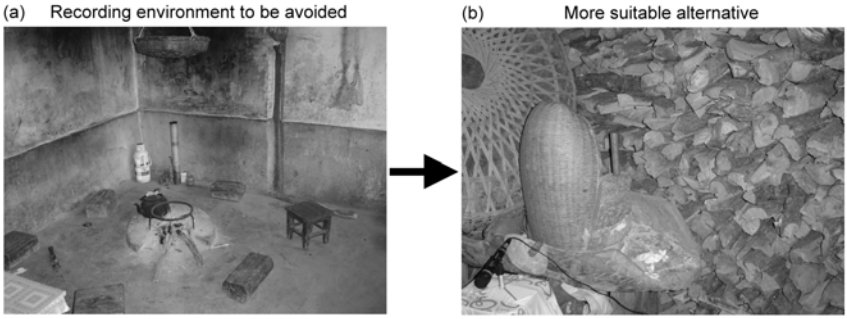


Figure 6. (a) shows an unsuitable recording room, a kitchen of a Naxi farm in Yunnan, China; (b) shows the more suitable recording environment on the farm, which was chosen instead.

Figure 6(b) shows the location that was chosen for recording inside the Yunnanese farm: in the courtyard, behind a thick stack of firewood which absorbs reverberation. This spot is located under a porch-roof, whose uneven tiling provides protection from the rain without creating strong reverberation. Large utensils of wood and stone partly cover the cement floor (a grindstone to make bean curd is seen on the photo, behind the microphone). When doing a recording, a piece of thick, rough cloth is propped across the open end of this makeshift recording booth, both contributing (at least minimal) acoustic improvement and providing a signal to the family members that noises are to be kept to a minimum, cf. Figure 7. The environment noises (chirping of birds and occasionally distant sounds of dogs barking, people shouting, or the rumble of an engine) are not a real issue for acoustic analysis, whereas reverberation is a major problem. Needless to say, any such changes in layout at someone's home needs to be discussed with one's hosts; good communication with one's consultants is absolutely essential.

Figure 7. Further optimation of the recording environment. A piece of thick, rough cloth is used to cover sound-reflecting surfaces and open space, which also dampens background noise.

4.6 Placing one's data within a broader context

Constant carrier sentences like "I don't know the word ___"; "The next word is ___"; or "I have seen ___ on the table" allow for a maximum degree of control, necessary for experimental and statistical purposes. It should be borne in mind that they represent a strong abstraction from everyday communication, however, so that one should be careful when attempting extrapolations. No set of data is complete in itself; no experiment provides a knock-out argument. Each phonetic data set represents a compromise between the competing demands of symmetry, on the one hand, and breadth of scope, on the other. One researcher, or even one team of researchers, cannot hope to gather an all-encompassing set of data.

One's data should therefore be placed within a broader context, and seen as a contribution to a broader set, whose gradual constitution can only be a collaborative endeavour. This constitutes a vision for the future of phonetics in the digital era, where the issue of data storage and accessibility looks very different from the predigital era.

Initial promising steps to a collaborative creation and use of annotated speech corpora and databases at an international level have already been made. Prominent representatives of this endeavour are the EU-funded `CLARIN´ network (Common Language Resources and Technology Infrastructure, Váradi et al. 2008), the 'Reciprosody', i.e. a repository for prosodically oriented and annotated speech corpora (founded by A. Rosenberg, http://jaguar.cs.qc.cuny.edu/), or archives for endangered languages data such as the 'DoBeS Archive' hosted by the Max-Planck-Institute for Psycholinguistics in Nijmegen (Drude et al. 2012) and the 'Pangloss Collection' at CNRS (Michailovsky et al. 2014). The advice provided in this paper, particularly concerning the need to use professional recording equipment and to collect detailed metadata, must also be seen in the light of such international collaborations, which all put minimum (and in tendency increasing) demands on hosted corpora.

5. Summary

Recordings have to be well planned, tested, and should not be conducted with the first available equipment. In short: Do not underestimate the challenge of speech data acquisition. Do not take recordings lightly! Many potential issues can be anticipated, managed, and controlled. This final section summarizes tips and recommendations on how to meet the demands of specific research questions and achieve results of lasting value for the scientific community.

5.1 The speaker

- (a) Do not select just any available speaker: screen your speakers carefully with respect to how they fit in with your research question (e.g., language skills, expressiveness), and in order to create a homogeneous sample (e.g., age, gender, smoking habits, musical experience)

- (b) If you record dialogues, either control phonetic entrainment or exploit it to implicitly change the speech patterns you get into a certain direction.

- (c) Collect a comprehensive set of metadata from your speakers; put special emphasis on linguistic experience and personality characteristics/habits.

- (d) Your sample size should be large enough to allow looking for individual preferences and between-group differences, for example, with respect to gender, dialectal background, and daytime. This probably means recruiting at least 10 speakers.

- (e) When you prepare the recording environment, the elicitation materials, and the task instructions, keep in mind that speakers are no "vending machines", which produce representative speech by paying and pressing a button, and that speech is essentially a social phenomenon. Treat your speakers with respect and bond with them.

- (f) For practical reasons, phoneticians often record speakers living away from the area where the target language (or dialect) is spoken. The consequences of language contact can be reduced by selecting people who have recently arrived from their homeplace, but you should remain on the watchout for effects of language contact nonetheless.

5.2 The task

- (a) Be as detailed as you can in the instructions, and use the same instructions for all speakers. So, prepare them as a sound file or in written form.
- (b) Make sure that the semantic-pragmatic context in which you embed the elicitation task is as rich and specific as possible. Don't be afraid to be

creative and integrate multimedia resources and/or aspects of everyday life. The richer, clearer, and more specific the elicitation context is, the more homogeneous, replicable, and valid your speech material will be.

- (c) Include a debriefing during or after the recording session and take the feedback of your speakers seriously. In terms of everyday communication, they are no less an expert than you, so do not force them to produce utterances they reject, for example, because of their wording or grammar.

- (d) Be extremely careful when translating speech material or instructions from a different study.

- (e) Conduct pilot recordings to test your material and instructions, and don't use the same speakers for the following main recording.

- (f) If you need initial data to study the basic characteristics of a new field of segmental or prosodic phenomena, it is advisable to make use of the high event density and control offered by read speech, i.e. elicit isolated sentences or read monologues.

- (g) If, on the other hand, you would like to go into the details of a well-understood segmental or prosodic phenomenon, or if you are interested in the phonetic exponents of discourse functions, make use of functionally rich and ecologically more valid dialogue tasks, i.e. elicit scripted or unscripted dialogues.

- (h) If you are not sure whether (g) applies to your research topic, the best solution is always to elicit different types of speech materials, for example a combination of read sentences and unscripted dialogues.

- (i) Role-play, quiz, and instruction-giving tasks allow eliciting target words with reasonably high frequencies even in unscripted speech. Similarly, read dialogues also strike a reasonable compromise between event density, experimental control, expressiveness, and ecological validity.

- (j) In dialogues, you can also use a privy dialogue partner to channel the speech behaviour of your speakers towards a certain direction.

- (k) If you elicit monologues – be they read or spontaneous – give your speakers somebody to address. Even if s/he does not say anything, it will make your speaker feel more comfortable and increase the quality, diversity, and validity of your speech materials.

- (l) If your recording session involves two subjects, use good friends in order to create a more informal atmosphere during the recording session.

- (m) Avoid monotonous tasks. For example, limit sentence lists to no more than 50 tokens; randomize sentences. Avoid repetitions of the same sentence (structure); and if this is not possible, insert dummy sentences that clearly deviate from your target sentences.

5.3 The recording

- (a) Take time to look for a suitable recording environment, if you make recordings outside the laboratory, and further optimize the environment by reducing open space, background noise, and reverberant surfaces, if possible.

- (b) Use professional recording equipment, and digitalise your speech with 44.1 kHz and 16-bit or higher recording settings. Monitor your recording carefully, especially if you are not yet thoroughly familiar with the equipment.

- (c) Allow speakers to familiarize themselves with the recording situation. That is, include a warm-up task – for example, let the speaker summarize the previous weekend/dinner or ask his/her hobbies – before you start with the actual recordings. Use the warm-up phase for adjusting the recording level in order to avoid that sections of the actual recording are distorted by clipping, cf. (b) above.

- (d) Try using a head-mounted microphone or a microphone with cardioid or shotgun characteristic to reduce environmental noise in your recordings. Head-mounted microphones have the further advantage that the distance between speaker/mouth and the microphone remains constant. Hence the intensity level is independent of head or body movements and becomes an analyzable acoustic parameter.

- (e) Hide away technical equipment and other things that have the potential to intimidate or distract your speakers.

- (f) Do not record "disposable data". Make your data available to the scientific community by depositing them in institutional repositories that ensure their long-term preservation and access. This requires prior design of forms to be signed by the speakers, to indicate their informed consent to participate in the experiment and to give copyrights (in certain fieldwork settings, oral consent can be substituted as appropriate). 'CreativeCommons' licences have many advantages for data sharing in scientific research.

6. Conclusion

As a speech scientist, you record data as you think fit, in view of your immediate research purposes. The above review suggests that you stand to gain a lot by considering a range of options at each of the three main stages of speech data collection: different types of procedures offer different insights, and their combination yields an in-depth, well-rounded view of speech. Time constraints and tight deadlines make it appear unreasonable to lavish your time on seemingly preliminary tasks such as contextualizing data and exchanging at leisure with your consultants before and after experiments. But the time spent on preparing data collection is in fact well invested, yielding considerable benefits for research. You will get a handle on major sources of variability, instead of unwittingly leaving important parameters uncontrolled and treating the ensuing variability as random. Painstaking data collection makes for reliable and enduring documents, which can profitably be shared – not only re-used, but also enriched collaboratively. In this optimistic perspective, data collection (language documentation) and research can progress hand in hand, allowing for a cumulative approach to research in the phonetic sciences.

We are well aware that the present article, which is essentially intended to provide some practical suggestions, only scratches the surface of data collection methodology. Further work would require scrutinizing and comparing the full range of recording tasks, conditions, and instructions on the basis of systematic experimental studies. Pending such in-depth work, our provisional morality is that perfecting elicitation methods requires keeping a constant eye on function, meaning, and the individual; this holds true of all types of research in phonetics/phonology, over and above the great diversity of research goals and methods.

7. Acknowledgments

Many thanks to Jacqueline Vaissière for useful comments. Further thanks are due to Gu Wentao and Klaus Kohler, and two anonymous reviewers for their constructive comments on an earlier draft of this paper. Needless to say, they are not to be held responsible for the views expressed here. We are also indebted to Sarah Buchberger and Jana Bahrens for helping us with formatting the paper and cross-checking the references. Support from Agence Nationale de la Recherche (HimalCo project, ANR-12-CORP-0006) and from LabEx "Empirical Foundations of Linguistics" (EFL) is gratefully acknowledged.

8. References

ABBERTON, E. / A.J. FOURCIN. 1984. Electroglottography. Experimental Clinical Phonetics, 62–78.

ABRAMSON, A.S. 1972. Tonal Experiments with Whispered Thai. In: A. Valdman (Ed.), Papers on Linguistics and Phonetics in Memory of Pierre Delattre (pp. 31–44). The Hague: Mouton.

AMBRAZAITIS, G. 2005. Between Fall and Fall-Rise: Substance-function Relations in German Phrase-final Intonation Contours. Phonetica 62 (2-4), 196–214.

AMDAL, I. / T. SVENDSEN. FonDat1: A Speech Synthesis Corpus for Norwegian. Proc. 5th International Conference on Language Resources and Evaluation, Genova, Italy, 2096-2101.

AMDAL, I. / O. STRAND / J. ALMBERG / T. SVENDSEN. 2008. RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus. Proc. 5th International Conference on Language Resources and Evaluation, Marrakech, Morocco, 1907-1913.

ANDERSON, A.H. / M. BADER / E. GURNAN BARD / E. BOYLE / G. DOHERTY / S. GARROD / S. ISARD / et al. 1991. The HCRC Map Task Corpus. Language and Speech 34, 351–366.

BARNLUND, D.C. 2008. A transactional model of communication. In: C. D. Mortensen (Ed.), Communication theory (pp. 47-57). New Brunswick, New Jersey: Transaction.

BARNES, J. / A. BRUGOS / E. ROSENSETIN / S. SHATTUCK-HUFNAGEL / N. VEILLEUX. 2013. Segmental sources of variation in the timing of American English pitch accents. Paper presented at the annual meeting of the Linguistic Society of America, Boston, USA.

BRAUN, B. / D.R. LADD. 2003. Prosodic correlates of contrastive and non-contrastive themes in German. Proc. 8th Eurospeech Conference, Geneva, Switzerland, 789-792.

BRUNELLE, M. 2012. Dialect Experience and Perceptual Integrality in Phonological Registers: Fundamental Frequency, Voice Quality and the First Formant in Cham. Journal of the Acoustical Society of America 131, 3088–3102.

BÜHLER, K. 1934. Sprachtheorie. Die Darstellungsfunktion Der Sprache. Jena: Gustav Fischer.

CAMPBELL, N. / P. MOKHTARI. 2003. Voice Quality: The 4th Prosodic Dimension. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2417–2420.

CLARKE, C.M. / M.F. GARRETT. 2004. Rapid Adaptation to Foreign-accented English. Journal of the Acoustical Society of America 116, 3647–3658.

COOKE, M. / O. SCHARENBORG. 2008. The Interspeech 2008 Consonant Challenge. Proc. 7th Interspeech Conference, Brisbane, Australia, 1-4.

CRAWFORD, M.D. / G.J. BROWN / M.P. COOKE / P.D. GREEN. 1994. Design, collection and analysis of a multisimultaneous- speaker corpus. Inst. Acoustics 16, 183-190.

CULIOLI, A. 1995. Cognition and Representation in Linguistic Theory. Current Issues in Linguistic Theory. Amsterdam: John Benjamins.

DELVAUX, V. / A. SOQUET. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 64, 145-173.

DÔ, T.D. / T.H. TRÂN / G. BOULAKIA. 1998. Intonation in Vietnamese. In: D. Hirst & Albert Di Cristo (Eds), Intonation Systems: A Survey of Twenty Languages (pp. 295-416). Cambridge: Cambridge University Press

DOMBROWSKI, E. 2003. Semantic features of accent contours: effects of F0 peak position and F0 time shape. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 1217-1220.

DIMMENDAAL, G. J. 2001. Places and People: Field Sites and Informants. In: P. Newman & M. Ratliff (Eds.), Linguistic Fieldwork (pp. 55-75). Cambridge: Cambridge University Press.

DRUDE, S. / P. TRILSBEEK / D. BROEDER. 2012. Language Documentation and Digital Humanities: The (DoBeS) Language Archive. Proc. International Conference of Digital Humanities, Hamburg, Germany, 169-173.

EDMONDSON, J.A. / K.J. GREGERSON. 1993. Western Austronesian Languages. In: J.A. Edmondson & K.J. Gregerson (Eds), Tonality in Austronesian Languages (pp. 61-74). Honolulu: University of Hawai'i Press.

ELLIS, L. / W.J. HARDCASTLE. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. Journal of Phonetics 30, 373- 396.

ERNESTUS, M. 2000. Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface. Utrecht: LOT.

FABRE, P. 1957. Un Procédé Électrique Percutané D'inscription De L'accolement Glottique Au Cours De La Phonation: Glottographie De Haute Fréquence. Bulletin De l'Académie Nationale De Médecine 141, 66–69.

FERLUS, M. 1979. Formation Des Registres Et Mutations Consonantiques Dans Les Langues Mon-khmer. Mon- Khmer Studies 8, 1–76.

FON, J. 2006. Shape Display: Task Design and Corpus Collection. Proc. 3rd International Conference of Speech Prosody, Dresden, Germany, 181-184.

FÓNAGY, I. 2001. Languages Within Language: An Evolutive Approach. Foundations of Semiotics 13. Amsterdam/Philadelphia: Benjamins.

FOURAKIS, M. / G.K. IVERSON. 1984. On the 'Incomplete Neutralization' of German Final Obstruents. Phonetica 41, 140–149.

FRANCOIS, A. 2012. The Dynamics of Linguistic Diversity: Egalitarian Multilingualism and Power Imbalance Among Northern Vanuatu Languages. International Journal of the Sociology of Language 214, 85–110.

GARTENBERG, R. / C. PANZLAFF-REUTER. 1991. Production and Perception of F0 Peak Patterns in German. Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel (AIPUK) 25, 29–113.

GENDROT, C. / M. ADDA-DECKER. 2007. Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. Proc. 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 1417–1420.

GENZEL, S. / F. KÜGLER. 2010. The prosodic expression of contrast in Hindi. Proc. 5th International Conference of Speech Prosody, Chicago, USA, 1-4.

GILES, H. / N. COUPLAND. 1991. Language: Contexts and Consequences. Mapping Social Psychology. Belmont: Thomson Brooks/Cole Publishing Co.

GODIN, K.W. / J.H.L. HANSEN. 2008. Analysis and perception of speech under physical task stress. Proc. 8th Interspeech Conference, Brisbane, Australia, 1674-1677.

GÖRS, K. 2011. Von früh bis spät - Phonetische Veränderungen der Sprechstimme im Tagesverlauf. BA thesis, Kiel University, Germany.

GÖRS, K. / O. NIEBUHR. 2012. Hocus Focus - What the Eliciation Method Tells Us About Types and Exponents of Contrastive Focus. Proc. 6th International Conference of Speech Prosody, Shanghai, China, 262–265.

GOW, D. 2003. Feature parsing: Feature cue mapping in spoken word recognition. Perception and Psychophysics 65, 575-590.

GRINEVALD, C. 2006. Worrying about ethics and wondering about "informed consent": Fieldwork from an Americanist perspective. In: A. Saxena & L. Borin (Eds), Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology (pp. 339-370). Berlin: De Gruyter.

GROSS, J.J. / O.P. JOHN. 1997. Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. Journal of Personality and Social Psychology 72, 435-448.

GUENTHER, F.H. / C.Y. EPSY-WILSON / S.E. BOYCE / M.L. MATTHIES / M. ZANDIPOUR / J.S. PERKELL. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. Journal of the Acoustical Society of America 105, 2854-2865.

GUSSENHOVEN, C. 2004. The Phonology of Tone and Intonation. Cambridge: CUP.

HAAS, M. 1944. Men's and Women's Speech in Koasati. Language 20 (3): 142 149.

HAUDRICOURT, A.-G. 1961. Richesse En Phonèmes Et Richesse En Locuteurs. L'Homme 1, 5–10.

HENDERSON, E.J.A. 1952. The Main Features of Cambodian Pronunciation. Bulletin of the School of Oriental and African Studies 14, 149–174.

HENRICH, N., C. D'ALLESSANDRO, M. CASTELLENGO, and B. DOVAL. 2004. On the Use of the Derivative of Electroglottographic Signals for Characterization of Non-pathological Voice Phonation. Journal of the Acoustical Society of America 115, 1321–1332.

HERMES, A. / J. BECKER / D. MÜCKE / S. BAUMANN / M. GRICE. 2008. Articulatory gestures and focus marking in German. Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 457-460.

HIMMELMANN, N. 2006. Prosody in Language Documentation. In: J. Gippert, N.P. Himmelmann & U. Mosel (Eds), Essentials of Language Documentation (pp. 163-181). Berlin/New York: de Gruyter.

HIRST, D. / A. DI CRISTO. 1998. A Survey of Intonation Systems. In: D. Hirst & Albert Di Cristo (Eds), Intonation Systems: A Survey of Twenty Languages (pp. 1–43). Cambridge: Cambridge University Press.

HOOLE, P. 1999. On the lingual organization of the German vowel system. Journal of the Acoustical Society of America 106, 1020–1032.

HOUSE, J. 1989. Syllable structure constraints on F0 timing. Poster presented at LabPhon II, Edinburgh, Scotland.

HUALDE, J.I. 2003. Remarks on the diachronic reconstruction of intonational patterns in Romance with special attention to Occitan as a bridge language. Catalan Journal of Linguistics 2, 181–205.

HUFFMAN, F.E. 1976. The Register Problem in Fifteen Mon-Khmer Languages. Austroasiatic Studies. In: P.N. Jenner, L.C. Thompson & S. Starosta (Eds), Oceanic Linguistics Special Publication No 13 (pp. 575-589). Honolulu: Hawaii University Press.

ITO, K. / S.R. SPEER. 2006. Using interactive tasks to elicit natural dialogue. In: S. Sudhoff et al. (Eds), Methods in empirical prosody research (pp. 229-258). Berlin/ New York: de Gruyter.

IWASHITA, N. / T. MCNAMARA / C. ELDER. 2001. Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. Language Learning 21, 401-436.

JANSE, E. / P. ADANK. 2012. Predicting Foreign-accent Adaptation in Older Adults. Quarterly Journal of Experimental Psychology 65, 1563–1585.

JASSEM, W. / L. RICHTER. 1989. Neutralization of Voicing in Polish Obstruents. Journal of Phonetics 17, 317–325.

JOHANNES, B. / P. WITTELS / R. ENNE / G. EISINGER / C.A. CASTRO / J.L. THOMAS / A.B. ADLER / R. GERZER. 2007. Non-linear function model of voice pitch dependency on physical and mental load. European Journal of Applied Physiology 101, 267–276.

JOHNSON, K. / P. LADEFOGED / M. LINDAU. 1993. Individual differences in vowel production. Jorunal of the Acoustical Society of America 94, 701-714.

JOHNSTONE, T. / C.M. VAN REEKUM / K. HIRD / K. KISNER / K. SCHERER. 2005. Affective speech elicited with a computer game. Emotion 5, 513-518.

KEATING, P. / C. ESPOSITO / M. GARELLEK / S. UD DOWLA KHAN / J. KUANG. 2010. Phonation Contrasts Across Languages. UCLA Working Papers in Phonetics 108, 188–202.

KIRBY, J. 2014. Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. Journal of Phonetics 43. 69–85.

KIM, M. 2012. Phonetic Accommodation after Auditory Exposure to Native and Nonnative Speech. PhD thesis, Northwestern University, IL, USA.

KLEBER, F. / T. JOHN / J. HARRINGTON. 2010. The Implications for Speech Perception of Incomplete Neutralization of Final Devoicing in German. Journal of Phonetics 38, 185–196.

KOHLER, K.J. 1983. F0 in speech timing. AIPUK 20, 55-97.

KOHLER, K.J. 1990. Macro and micro F0 in the synthesis of intonation. In: J. Kingston & M.E. Beckman (Eds), Papers in Laboratory Phonology I (pp. 115-138). Cambridge: Cambridge University Press.

KOHLER, K.J. 1991. A model of German intonation. AIPUK 25, 295-360.

KOHLER, K.J. 1995. Einführung in die Phonetik des Deutschen. Berlin: Erich Schmidt.

KOHLER, K.J. 2004. Categorical speech perception revisited. Proc. of the Conference "From Sound to Sense: 50+ years of discoveries in speech communication", MIT Cambridge, USA, 1-6.

KOHLER, K.J. 2006. Paradigms in experimental prosodic analysis: From measurement to function. In S. Sudhoff et al. (Eds.), Methods in empirical prosody research (pp. 123-152). Berin/New York: de Gruyter.

KOHLER, K.J. / O. NIEBUHR. 2007. The phonetics of emphasis. Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken, Germany, 2145-2148.

KOHLER, K.J. / O. NIEBUHR. 2011. On the Role of Articulatory Prosodies in German Message Decoding. Phonetica 68, 1–31.

KOHTZ, L.-S. 2012. Datenerhebung mittels Leselisten - Eine kritische phonetische Evaluation. BA thesis, Kiel University, Germany.

KÜHNERT, B. / P. HOOLE. 2004. Speaker-specific kinematic properties of alveolar reductions in English and German. Clinical Linguistics and Phonetics 18, 559-575.

LADD, D.R. 2003. Phonological conditioning of F0 target alignment. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 249-252.

LAVER, J. 1994. Principles of Phonetics. Cambridge: Cambridge University Press.

LIBERMAN, M. / J. PIERREHUMBERT. 1984. Intonational Invariance Under Changes in Pitch Range and Length. In R.T. Oehrle & M. Aronoff (Eds), Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students (pp. 157–233). Cambridge: MIT Press.

MARCHAL, A. 2009. From Speech Physiology to Linguistic Phonetics. London: Wiley.

MICHAILOVSKY, B. / M. MAZAUDON / A. MICHAUD / S. GUILLAUME / A. FRANCOIS / E. ADAMOU. 2014. Documenting and researching endangered languages: the Pangloss Collection. Language Documentation and Conservation 8, 119–135.

MICHAUD, A. 2004a. Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. Phonetica 61, 119–146.

MICHAUD, A. 2004b. A Measurement from Electroglottography: DECPA, and Its Application in Prosody. Proc. 2nd International Conference of Speech Prosody, Nara, Japan, 633–636.

MICHAUD, A. 2012. Monosyllabicization: Patterns of Evolution in Asian Languages. In: N. Nau, T. Stolz & C. Stroh (Eds), Monosyllables: From Phonology to Typology (pp. 115–130). Berlin: Akademie Verlag.

MICHAUD, A. / T. VU-NGOC / A. AMELOT / B. ROUBEAU. 2006. Nasal Release, Nasal Finals and Tonal Contrasts in Hanoi Vietnamese: An Aerodynamic Experiment. Mon-Khmer Studies 36, 121–137.

MITHUN, M. 2001. Who Shapes the Record: The Speaker and the Linguist. In: P. Newman & M. Ratliff (Eds), Linguistic Fieldwork (pp. 34–54). Cambridge: Cambridge University Press.

MIXDORFF, H. 2004. Qualitative analysis of prosody in task-oriented dialogs, Proc. 2nd International Conference on Speech Prosody, Nara, Japan, 283-286.

MIXDORFF, H. / H.R. PFITZINGER. 2005. Analysing fundamental frequency contours and local speech rate in map task dialogs. Speech Communication 46, 310-325.

MOSEL, U. 2006. Field Work and Community Language Work. In: J. Gippert, N.P. Himmelmann & U. Mosel (Eds), Essentials of Language Documentation (pp. 67–83). Berlin/New York: de Gruyter.

MOSEL, U. 2011. Morphosyntactic analysis in the field - a guide to the guides. In: N. Tieberger (Ed.), The Oxford handbook of linguistic fieldwork (pp. 72-89). Oxford: OUP.

NEWMAN, P. / M. RATLIFF. 2001. Linguistic Fieldwork. Cambridge: Cambridge University Press.

NIEBUHR, O. 2007a. The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. Phonetica 64, 174-193.

NIEBUHR, O. 2007b. Perzeption und kognitive Verarbeitung der Sprechmelodie – Theoretische Grundlagen und empirische Untersuchungen. Berlin/New York: de Gruyter.

NIEBUHR, O. 2008. Coding of Intonational Meanings Beyond F0: Evidence from Utterance-final /t/ Aspiration in German. Journal of the Acoustical Society of America 142, 1252–1263.

NIEBUHR, O. 2009. Intonation Segments and Segmental Intonations. Proc. 10th Interspeech Conference, Brighton, UK, 2435–2438.

NIEBUHR, O. 2010. On the Phonetics of Intensifying Emphasis in German. Phonetica, 170–198.

NIEBUHR, O. 2012. At the edge of intonation - The interplay of utterance-final F0 movements and voiceless fricative sounds. Phonetica 69, 7-27.

NIEBUHR, O. 2013. On the acoustic complexity of intonation. In: E.-L. Asu & P. Lippus (Eds), Nordic Prosody XI (pp. 15-29). Frankfurt: Peter Lang.

NIEBUHR, O. / J. BERGHERR / S. HUTH / C. LILL / J. NEUSCHULZ. 2010. Intonationsfragen hinterfragt - Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. Zeitschrift für Dialektologie und Linguistik 77, 304-346.

NIEBUHR, O. / M. D'IMPERIO / B. GILI FIVELA / F. CANGEMI. 2011a. Are There 'Shapers' and 'Aligners'? Individual Differences in Signalling Pitch Accent Category. Proc. 17th International Congress of Phonetic Sciences, Hong Kong, China, 120–123.

NIEBUHR, O. / M. CLAYARDS / CH. MEUNIER / L. LANCIA. 2011b. On Place Assimilation in Sibilant Sequences - Comparing French and English. Journal of Phonetics 39, 429–451.

NIEBUHR, O. / CH. MEUNIER. 2011. The phonetic manifestation of French /s#ʃ/ and /ʃ#s/ sequences in different vowel contexts - On the occurrence and the domain of sibilant assimilation. Phonetica 68, 133-160.

NOLAN, F. 1992. The descriptive role of segments: evidence from assimilation. In D.R. Ladd & G.J. Docherty (Eds), Papers in Laboratory Phonology 2 (pp. 261–280). Cambridge: CUP.

NOLAN, F. 1995. The Effect of Emphasis on Declination in English Intonation. In: J.W. Lewis (Ed.), Studies in General and English Phonetics. Essays in Honour of Professor J.D. O'Connor (pp. 241–254). London & New York: Routledge.

ODGEN, R. 2006. Phonetics and social action in agreements and disagreements. Journal of Pragmatics 38, 1752–1775.

OHL, C.K. / H.R. PFITZINGER. 2009. Compression and Truncation Revisited. Proc. 10th Interspeech Conference, Brighton, UK, 2451–2454.

PASTEUR, L. 1939. OEuvres, tome VII: mélanges scientifiques et littéraires. Paris: Masson. http://catalogue.bnf.fr/ark:/12148/cb37416454q.

PETERS, B. 1999. Prototypische Intonationsmuster in Deutscher Lese- Und Spontansprache. AIPUK 34, 1–177.

PETERS, B. 2005. The Database 'The Kiel Corpus of Spontaneous Speech'. AIPUK 35a, 1–6.

PETERS, J. 2006. Intonation deutscher Regionalsprachen (Linguistik - Impulse & Tendenzen, Vol. 21). Berlin/New York: de Gruyter.

PIERREHUMBERT, J.B. / J. HIRSCHBERG. 1990. The meaning of intonation contours in the interpretation of discourse. In: P.R. Cohen, J. Morgan, and M.E. Pollack (Eds), Intentions in communication (pp. 271-311). Cambridge, Mass.: MIT Press.

PORT, R. / M. O'DELL. 1985. Neutralization of Syllable-final Voicing in German. Journal of Phonetics 13: 455–471.

PROBERT, PH. 2009. Comparative philology and linguistics. In: B. Graziosi, Ph. Vasunia & G. Boys-Stones (Eds.), The Oxford Handbook of Hellenic Studies (pp. 697-708). Oxford: Oxford University Press.

RIETVELD, T. / C. GUSSENHOVEN. 1995. Aligning pitch targets in speech synthesis: Effects of syllable structure. Journal of Phonetics 23, 375-385.

SCHERER, K. / D. GRANDJEAN / T. JOHNSTONE / G. KLASMEYER / T. BÄNZIGER. 2002. Acoustic correlates of task load and stress. Proc. International Conference on Spoken Language Processing, Denver, USA, 2017-2020.

SCHWAB, S. 2011. Relationship between Speech Rate Perceived and Produced by the Listener. Phonetica 68, 243- 255.

SHEN, R. 2013. Tones and consonants in Shibei Min Chinese. Proceedings of International Conference on Phonetics of the Languages in China (ICPLC-2013), Hong Kong, China, 171–174.

SIMPSON, A. 2012. The First and Second Harmonics Should Not Be Used to Measure Breathiness in Male and Female Voices. Journal of Phonetics 40, 477–490.

SIMPSON, A. / K.J. KOHLER / T. RETTSTADT. 1997. The Kiel Corpus of Read/Spontaneous Speech: Acoustic Data Base, Processing Tools, and Analysis Results. Kiel. AIPUK 32.

SOUAG, L. 2011. Review of: Claire Bowern. 2008. Linguistic Fieldwork: A Practical Guide. Language Documentation and Conservation 5, 66–68.

SPILKOVÁ, H. / D.S. BRENNER / A. ÖTTL / P. VONDRICKA / W. VAN DOMMELEN / M. ERNESTUS. 2010. The Kachna L1/L2 Picture Replication Corpus. Proc. 7th International Conference on Language Resources and Evaluation, Malta, Spain, 2432-2436.

STEPPLING, M.L. / A.A. MONTGOMERY. 2002. Perception and production of rise-fall intonation in American English. Perception and Psychophysics 64, 451-461.

SUNDBERG, J. 1979. Maximum Speed of Pitch Changes in Singers and Untrained Subjects. Journal of Phonetics 7, 71–79.

TITZE, I.R. 1989. Physiologic and Acoustic Differences Between Male and Female Voices. Journal of the Acoustical Society of America 85, 1699–1707.

TORREIRA, F. / M. ADDA-DECKER / M. ERNESTUS. 2010. The Nijmegen Corpus of Casual French. Speech Communication 52, 201-221.

TURK, A. / S. NAKAI / M. SUGAHARA. 2006. Acoustic segment durations in prosodic research: A practical guide. In: S. Sudhoff et al. (Eds.), Methods in empirical prosody research (pp. 1-28). Berlin/New York: de Gruyter.

VAISSIÈRE, J. 2004. The Perception of Intonation. In: D.B. Pisoni & R.E. Remez (Eds.), Handbook of Speech Perception (pp. 236–263). Oxford: Blackwell.

VAISSIÈRE, J. / K. HONDA / A. AMELOT / SH. MAEDA / L. CREVIER-BUCHMAN. 2010. Multisensor Platform for Speech Physiology Research in a Phonetics Laboratory. Journal of the Phonetic Society of Japan 14, 65–77.

VÁRADI, T. / P. WITTENBURG / S. KRAUWER / M. WYNNE / K. KOSKENNIEMI. 2008. CLARIN: Common language resources and technology infrastructure. Proc. 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco, 1244-1248.

WAGENER, P. 1986. Sind Spracherhebungen paradox? Über die Möglichkeit, natürliches Sprachverhalten wissenschaftlich zu erfassen. In: A. Schöne (Ed.), Akten des VII. IVG-Kongresses, Vol. 4 (pp. 319-327). Tübingen: Niemeyer.

WATSON, I. 2002. Convergence in the Brain; the Leakiness of Bilinguals' Sound Systems. In: M. Jones & E. Esch (Eds), Language Change: The Interplay of Internal, External and Extra-linguistic Factors (pp. 243–266). Berlin/New York: Mouton de Gruyter.

WELBY, P. / H. LOEVENBRUCK. 2006. Anchored down in Anchorage: Syllable structure and segmental anchoring in French. Italian Journal of Linguistics 18, 74-124.

XU, J. 2012. Problems and coping strategies of speech data collection - Insights from a special-purpose corpus of situated adolescent speech. Manuscript, University of Science and Technology of China. http://fld.ustc.edu.cn/123/xujiajin/index.htm
XU, Y. 2011. Speech prosody: A methodological review. Journal of Speech Sciences 1, 85-115.

ZELLERS, M. / B. POST. 2012. Combining Formal and Functional Approaches to Discourse Structure. Language and Speech 55, 119-139.

Kieler Arbeiten zur Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

2 Tone and Intonation – Introductory Notes and Practical Recommendations

Alexis Michaud
International Research Institute MICA
Hanoi University of Science and Technology, CNRS, Grenoble INP, Vietnam
Langues et Civilisations à Tradition Orale, CNRS/Sorbonne Nouvelle, France
alexis.michaud@mica.edu.vn

Jacqueline Vaissière
Laboratoire de Phonétique et Phonologie
Université Paris 3 - Sorbonne Nouvelle, CNRS, Paris
jacqueline.vaissiere@univ-paris3.fr

The present article aims to propose a simple introduction to the topics of (i) lexical tone, (ii) intonation, and (iii) tone-intonation interactions, with practical recommendations for students. It builds on the authors' observations on various languages, tonal and non-tonal; much of the evidence reviewed concerns tonal languages of Asia. With a view to providing beginners with an adequate methodological apparatus for studying tone and intonation, the present notes emphasize two salient dimensions of linguistic diversity. The first is the nature of the lexical tones: we review the classical distinction between (i) contour tones that can be analyzed into sequences of level tones, and (ii) contour tones that are non-decomposable (phonetically complex). A second dimension of diversity is the presence or absence of intonational tones: tones of intonational origin that are formally identical with lexical (and morphological) tones.

1. Introduction

There are many brilliant publications on methodological issues relating to prosody. Strongly recommended readings include Volume 8 of the journal "Language Documentation and Conservation" (2014), which contains high-level introductions to the study of tone systems ("How to Study a Tone Language" article series, edited by Steven Bird and Larry Hyman). Concerning intonation, suggested places to start

include the textbooks by Cruttenden (1986) and Hirst and Di Cristo (1998). But we feel that there is still room for a beginner-friendly introduction to the topics of (i) lexical tone, (ii) intonation, and (iii) tone-intonation interactions. The aim of the present notes is to provide beginners with an adequate methodological apparatus for studying tone and intonation, and to offer practical recommendations for engaging in research on these topics.

In particular, we wish to convey to students a sense of two salient dimensions of linguistic diversity. The first is the nature of the lexical tones. Section 2 reviews the classical distinction between (i) contour tones that can be analyzed into sequences of level tones, and (ii) contour tones that are non-decomposable (phonetically complex). A second dimension of diversity is the presence or absence of **intonational tones**: tones of intonational origin that are formally identical with lexical (and morphological) tones; section 3 examines the issue to what extent the various components of a language's intonation are structured in tonal terms. Many examples come from the languages with which the authors have greatest familiarity, but the aim is to clarify a few basic distinctions which we think allow for a clear view of prosodic phenomena in tonal and non-tonal languages alike.

2. Complex tones and level tones

2.1 A synchronic view

2.1.1 Fundamental notions: level tones and decomposable contours

The notions of "level tone" and "contour tone" are used in two different ways. Let us first introduce what can be broadly termed as "Africanist" usage, in which "level tone" refers to **a tone that is defined simply by a discrete level of relative pitch.**

Level-tone systems have **two to five levels of relative pitch**: L(ow) vs. H(igh); L vs. M(id) vs. H; L vs. M vs. H vs T(op); or B(ottom) vs. L vs. M vs. H vs. T. Systems with two levels are most widespread; systems with more than three levels are relatively uncommon (Bariba: Welmers 1952; Bench, a.k.a. Gimira: Wedekind 1983, 1985). One single case of six-level system has been reported: Chori (Dihoff 1977), for which reanalysis as a five-level system is possible (Odden 1995). These languages are spoken in Subsaharan Africa, a domain where level tones are especially common. However, level-tone representations have proved useful beyond the Subsaharan domain, for which they were initially developed.[1]

In level-tone systems, a phonetic contour results from the combination of two or more level tones: typically, a LH sequence realized phonetically as a rise in F_0, or a HL sequence realized as a fall. The contours are phonologically decomposable; **the observed movement in F_0 is the result of interpolation between the successive levels**. For instance, in Yongning Na (Sino-Tibetan), which has three level tones,

[1] On languages of the Americas: Gomez-Imbert 2001; Hargus and Rice 2005; Girón Higuita and Wetzels 2007; Michael 2010; on languages of Asia: Ding 2001; Hyman and VanBik 2002, 2004; Donohue 2003, 2005; Evans 2008; Jacques 2011.

H(igh), M(id) and L(ow), the compound noun /boɹ-ɬvɬ/ ‘pig’s brains’ in association with the copula /ɲiɻ/ yields /boɹ-ɬvɬ ɲiɽ/ ‘is (a/the) pig’s brains’: the rising contour present on ‘pig’s brains’ in isolation unfolds as M+H over the two syllables.

2.1.2. Notes on the phonetic realization of level tones

After phonological facts have been established through listening, practising and developing a feel for the language, it is revealing to explore phonetic detail in the realization of the tones. For instance, Figure 1 shows one token of /boɹ-ɬvɬ/ ‘pig’s brains’ and /boɹ-ɬvɬ ɲiɽ/ ‘is (a/the) pig’s brains’, from a set of recordings of compound nouns in Yongning Na.[2]

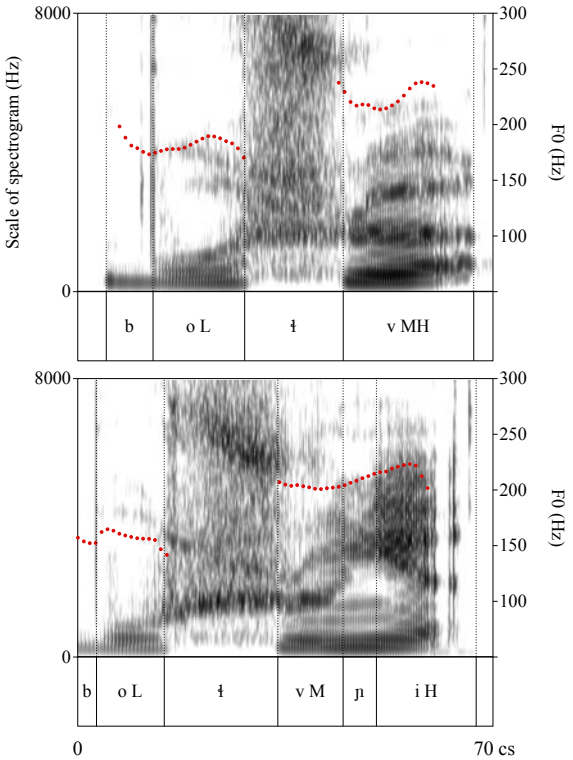


Figure 1. Spectrogram and F_0 tracing of two Na phrases. The time scale of both spectrograms is the same (70 centiseconds).

[2] The entire data set is available online from the Pangloss Collection (Michailovsky et al. 2014); direct link: http://lacito.vjf.cnrs.fr/pangloss/languages/Na_en.htm
For more information on the tone system of Yongning Na, see Michaud (2008, 2013).

The clear rise in F_0 on the rhyme /v/ in the top part of the figure is consistent with phonological description as a MH tone; and the flatter shape on that same rhyme in the bottom part of the figure, followed by higher F_0 on the copula, is consistent with phonological description as a sequence of M on one syllable and H on the next.

On the other hand, a fundamental point is that **there is no way to read phonological tones off F_0 tracings** (as emphasized e.g. by Cruz and Woodbury 2014 and Morey 2014). Figure 1, like any piece of experimental evidence, illustrates variability in the realization of tone. For instance, glottalization is found at the end of both tokens, exerting a detectable (lowering) influence on F_0 towards the offset of voicing: /boɭ-ɬvɬ/ 'pig's brains' and /boɭ-ɬvɬ ɲiʔ/ 'is (a/the) pig's brains' are both found in absolute final position (constituting entire sentences on their own), and glottalization is common in Na at this juncture of an utterance. Also, /boɭ/ 'pig' is realized with noticeably different F_0 in the top part of the figure and in the bottom part. Tonal realizations have some range of variation within tonal space (F_0), like vowels have some range of variation within the acoustic space (as characterized essentially by the first three formants, F1-F2-F3). The linguistic comment that can be proposed about the slight initial rise in the realization of the L tone of /boɭ/ 'pig' in the top part of Figure 1 is based on a phonological observation: in Na, rising tones are never found in initial position within a tone group (phonological phrase), and hence the identification of an initial L tone is not jeopardized by its realization with a slight rise as in the top part of Figure 1. Seen in this light, the existence of slightly rising 'allotones' does not come as a surprise: it makes sense in view of the phonological system – in the same way as, in a language that does not have contrastive aspirated consonants, plain (unaspirated) unvoiced consonants may sometimes be realized phonetically with some aspiration.

Back in the 1970s, at a time when F_0 tracings were difficult to obtain – requiring help from a specialized engineer –, a specialist of Bantu tone asked the second author of this paper to create an F_0 tracing from a recording illustrating a specific phonological phenomenon. After receiving the desired tracing, this famous specialist of tonology said that there must be a mistake, as the F_0 tracing did not correspond to the tone pattern that his trained ear discerned clearly in the recording. In fact, there was no error in F_0 detection; the issue lay in this colleague's expectation of a neatly binary F_0 tracing, straightforwardly reflecting the sequence of H and L tones. **Experimental examination of spoken language reveals that, even in languages with relatively straightforward prosodic systems**, such as Standard Japanese, **F_0 curves are shaped by a number of factors, and do not reflect phonological tone in a crystal-clear, transparent way**. (For auditory observations on phonetic realizations of tone in a two-tone language, see e.g. Guthrie 1940.)

2.1.3. Another fundamental notion: unitary contours

An important complement to the above discussion of "levels" and "contours" (2.1.1) is that there exist **tone systems where contours are not phonologically decomposable**. The use of the term "contour tone" to refer to a unitary contour is sometimes referred to as an "Asianist" use of terms, because of the wealth of well-documented examples from East and Southeast Asia. A contour tone in the Asianist sense is a tone defined

phonologically in terms of **an overall template specifying the time course of F_0 over the tone-bearing unit**.

The two types of phonological contour tones – sequences of levels on the one hand, unitary contours on the other – can be phonetically indistinguishable. The evidence for distinguishing the two types of contours is morpho-phonological.

In many languages, there is a wealth of evidence for the analysis of contour tones into sequences of level tones. A rising contour in an African language will typically exhibit phonological behaviour showing that it consists of two levels: a low tone followed by a high tone (see in particular Clements and Goldsmith 1984; Clements and Rialland 2007). An example from Yongning Na (a language of China) was presented in paragraph 2.1.1. There are some languages for which attempts at the decomposition of contours into levels has been less successful, however, to the point of casting doubt on the relevance of decomposition for these languages. The discussion below will focus on tonal systems of East and Southeast Asia.

Chao Yuen-ren's work on Mandarin Chinese in the early 20th century (Chao Yuen-ren 1929, 1933) brought to the attention of linguists the complexities of its tone system. Following sustained exchanges with Chao Yuen-ren, Kenneth Pike proposed a typological divide between two types of tones: (i) register-tones, defined simply in terms of discrete pitch levels, and (ii) contour-tones, about which he concludes: *"the glides of a contour system must be treated as unitary tonemes and cannot be broken down into end points which constitute lexically significant contrastive pitches"* (Pike 1948:10). This is echoed by recent observations about the Tai-Kadai family: *"I do not find the idea of binary features necessary or helpful in analysing the languages that I have worked on. In these languages, I do not believe that reducing the analysis of tones to a binary choice of H and L will assist in the understanding of the tonal system"* (Morey 2014:639). Likewise, in Vietnamese (Austroasiatic), *"there are no objective reasons to decompose Vietnamese tone contours into level tones or to reify phonetic properties like high and low pitch into phonological units such as H and L"* (Brunelle 2009a, 94; see also Brunelle et al. 2010; Kirby 2010).

In the description of contour-tone systems, the term "level tone" is used to refer to a tone that does not exhibit any salient fluctuations in F_0. For instance, Mandarin tone 1 and Vietnamese tone A1 (orthographic <ngang>) can be referred to as "level tones" because, unlike the other tones of Mandarin and Vietnamese, their F_0 curve is relatively stable in the course of the syllable. This does not entail that they are phonologically defined by a discrete level of relative pitch (on Mandarin: see Xu and Wang 2001).

Later studies have brought out the importance of **durational properties and phonation-type characteristics**. In some systems, phonation types are a redundant, low-level phonetic characteristic of some tones (see e.g. an investigation into the role of creaky voice in Cantonese tonal perception: Yu and Lam 2014). In others, phonation types are a distinctive feature orthogonal to tone, as in the Oto-Manguean languages Mazatec (Garellek and Keating 2011) and Trique (DiCanio 2012). Finally, in a third type of system, phonation-type characteristics are part and parcel of the definition of tones. Experimental studies of this third type of tone system include Rose (1982, 1989a, 1990) for the Wu branch of Sinitic; Edmondson et al. (2001) for Yi and Bai; Mazaudon and Michaud (2008) for Tamang; and Andruski and Ratliff (2000), Andruski and Costello (2004), Kuang (2012) for Hmong.

A famous example of this type of system is Hanoi Vietnamese, whose tones contrast with one another through a set of characteristics that include specific phonation types in addition to the time course of F_0 (Mixdorff et al. 2003; Brunelle, Nguyễn Khắc Hùng and Nguyễn Duy Dương 2010). As an example, tone C2 has medial glottal constriction. Figures 2 and 3 shows two tokens of this tone. (In transcriptions, superscript indications A1, C2 etc transcribe lexical tone.) The syllable at issue is the last of the sentence /ɓa^{A1}.đi^{A1}.hok^{D2}.ɗa^{C2}/ 'First, Ba (a person name) goes/will go to class' (orthography: <Ba đi học đã>). The top and bottom parts of the figures correspond to the realizations by a male speaker in two different contexts:

- top: Ba's friend asks him whether he has any plan for that morning. Ba says, in casual, conversational style, that first he will go to class. This reading will be referred to as **Declaration**.
- bottom: as Ba is on his way to school, some friends ask him to come and hang out with them. Ba answers in such a way as to clarify that he obviously can't join them, as he is going to class. This reading will be referred to as **Obviousness**.

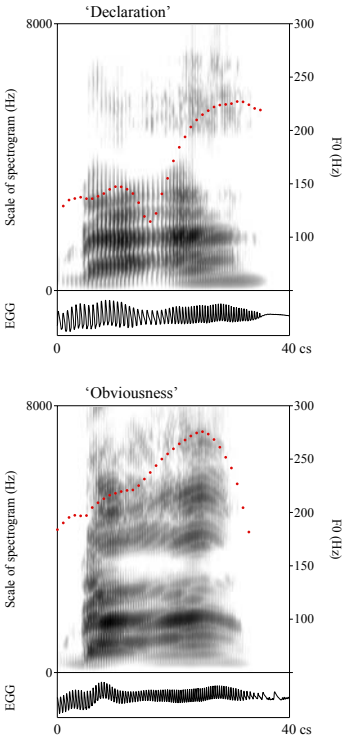


Figure 2. Spectrograms, F_0 tracings and electroglottographic signals for the Vietnamese syllable /ɗa^{C2}/ under two reading conditions. Same tokens as in Figure 3.

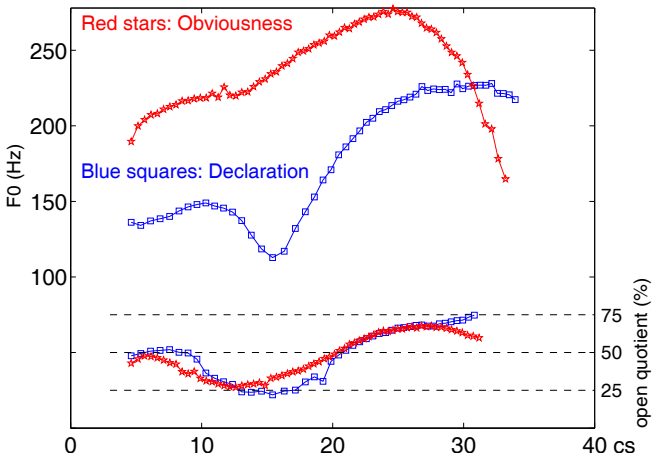


Figure 3. F_0 and open quotient values during the rhyme of the Vietnamese syllable /da^{C2}/ under two reading conditions. Same tokens as in Figure 2.

Figures 2 and 3 illustrate the presence of glottalization in the syllable /da^{C2}/. In the top part of Figure 2 ('Declaration'), the syllable is realized with a visible medial glottal constriction, evidenced by a sharp dip in F_0. A few longer glottal cycles are visible on the electroglottographic signal below the spectrogram. Then F_0 rises again, to high values. In the bottom part of Figure 2 ('Obviousness'), F_0 is much higher, and the F_0 is smoother, without a clear dip. Figure 3 reveals that laryngeal constriction is still present, however. Figure 3 shows the glottal open quotient (O_q) for these tokens, excluding the portion corresponding to the initial consonant, which is less relevant for the study of tone. (On the glottal open quotient and its estimation from electroglottography, see seminal research by Henrich et al. 2004.) The figure reveals that the glottal open quotient dips to extremely low values – on the order of 25 % – in the first half of the rhyme, under the 'Declaration' reading condition as well as under the 'Obviousness' reading condition. Thus, laryngeal constriction is present in both realizations, despite the great difference in terms of overall F_0 range. Figures 2 and 3 reveal a succession of rapid changes in phonation types. Open quotient values span a considerable range: the lowest values, under 25 %, are indicative of strong vocal fold adduction: extremely pressed voice. The highest values, on the order of 75 %, are indicative of highly relaxed phonation.

Additionally, the two tokens in Figures 2 and 3 illustrate the intonational plasticity of complex tones. Vietnamese tone C2 is specified for phonation type, in that it has glottal constriction mid-way through the syllable rhyme. On the other hand, there is no specification on its phonation type at the offset of voicing. Said differently, speakers are free to realize tone C2 with or without **final** glottalization: whether there is final glottalization or not has no bearing on tonal identification, so long as the syllable has the telltale characteristic of tone C2, namely **medial** constriction. This degree of freedom is exploited intonationally, in the expression of the speaker's attitude. Under the 'Declaration' reading condition, the syllable has a soft offset of voicing: the vocal

folds separate, and airflow decreases. This is evidenced by the gradual decrease in the amplitude of the electroglottographic signal. On the other hand, under the 'Obviousness' condition, the syllable ends in glottalization, as evidenced by the final decrease in open quotient: the downward tilt in the O_q tracing in red stars in Figure 3. No O_q values are displayed for the last few glottal cycles, as O_q could not be reliably estimated for that portion of the electroglottographic signal, due to the absence of well-defined opening peaks in the derivative of the signal; but glottalized cycles are visible in the EGG signal at the bottom right-hand corner of Figure 2.[3]

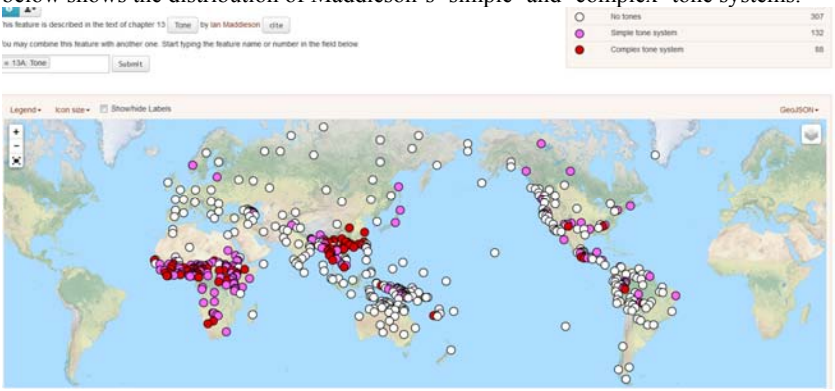
In view of such phenomena, the "contour tones" of Pike's typology can be considered part of a larger set: "complex tones", including tones that comprise phonation-type characteristics.

A quick summary:

- The term 'level tone' is used in different ways by different authors, and you should keep this polysemy in mind when reading papers. 'Level tone' can either refer to a discrete level of relative pitch, such as L or H; or it can refer to the plateau-like shape of a complex tone.
- To recapitulate the terms used in the present discussion: **level-tone systems** are based on discrete levels of relative pitch, unlike **complex-tones systems**, defined by an overall template specifying the time course of F0 over the tone-bearing unit.[4] Among complex-tone systems, a further distinction is whether they comprise phonation-type characteristics.

[3] For more data on how speaker attitude is reflected in the realization of tones in Vietnamese, see Nguyen et al. (2013). The robustness of glottalization as a correlate of certain tones is confirmed statistically (Michaud and Vu-Ngoc 2004; Michaud 2004).

[4] Note that this differs from the definition used in the World Atlas of Language Structures, where "complex" refers to the number of oppositions, not to the nature of the tones: *"[t]he languages with tones are divided into those with a simple tone system — essentially those with only a two-way basic contrast, usually between high and low levels — and those with a more complex set of contrasts"* (Maddieson 2011). The map below shows the distribution of Maddieson's 'simple' and 'complex' tone systems.



The recognition of this dimension of typological diversity is currently hampered by some difficulties, however, as outlined below.

2.1.2. The paucity of phonological models for complex tones: 'What is the alternative?'

Specialists typically lend more weight to the geographical areas or phylogenetic groups with which they are most familiar, with the consequence that they tend to grant universal status to the characteristics that they repeatedly observe in these languages, and to grant prototype status to the languages they are most familiar with. While many authors are not fully aware of this bias, Eugénie Henderson makes her choice clear:

> "My preference, derived both from professional training and experience, would be to present only material of which I have first-hand personal knowledge, since, though this may be fallible, one may at least suppose the same bias to run through the whole of it." (Henderson 1965:403)

This can lead one to consider with some suspicion the work of colleagues who, on the basis of evidence from different languages, reach different conclusions. The problem is by no means specific to tonal typology, but it is made more acute by the limited geographical distribution of complex-tone systems. Eugénie Henderson, who is familiar with complex tones, warns that *"… 'tone' is seldom, if ever, a matter of pitch alone"* (Henderson 1965:404); this "if ever" amounts to casting doubt on the validity of tonal descriptions that do not mention phonation type and other potential phonetic correlates of tone (see also Jones 1986; Rose 1989b). Conversely, the "autosegmental" representations initially developed for the level tones of African languages (Clements and Goldsmith 1984; Hyman 1981) are promptly raised by some authors to the status of universal representation of tone (e.g. Yip 2002).

It hardly sounds realistic to expect researchers in the field of tonal studies to acquire first-hand familiarity with languages from all over the world, including within the sample a Tai-Kadai, Hmong-Mien, Sino-Tibetan or Austroasiatic language with complex tones. The next best option is to obtain second-hand familiarity through readings and exchanges with colleagues. But at this point, another obstacle crops up: little has been proposed in the way of phonological modelling for systems of complex tones. A course about level tones can start out from compelling examples of alternations, followed by an exposition of the autosegmental model.

> "Here the evidence is clear: at least some contour tones must be analyzed as sequences of level tones because they can be seen to be derived from that source. In Hausa (Newman 1995; Jaggar 2001), some words have two variants, bi-syllabic and mono-syllabic. If the bisyllabic word is HL, then the monosyllable has a fall. If the fall is analyzed as simply a HL on a single vowel, then we can understand this as vowel deletion, with retention and reassociation of the remaining tone: mínì or mîn 'to me'." (Yip 2007:234)

It is much harder to convey a feel for tones that are phonetically complex and do not behave phonologically as levels or sequences of levels. The autosegmental model is an economical model, for which there is ample evidence from a broad range of languages;

it understandably creates hopes that it can eventually apply to all of the world's languages. This goes a long way towards explaining the existence of quite a substantial literature on the autosegmental analysis of the tones of Mandarin, Thai... (e.g. Bao 1999; Yip 1980, 1989, 2002; Morén and Zsiga 2007), and its continuing popularity despite major concerns about the absence of language-internal evidence (e.g. Barrie 2007:345; Sun 1997:516; Morey 2014:639). From a cross-linguistic point of view, level tones are reported to constitute a vast majority (Maddieson 1978:364); *"the most 'normal' tone system is one with only two level tones"* (Maddieson 1978:369). Phonetically complex tones thus look 'abnormal'; they have been presented as a geographical exception – a case of *"Chinese and the 'Sinosphere' (Matisoff 1999) vs. the world"* (Hyman 2011a:190).

It must be acknowledged that much remains to be done in proposing phonological models of complex tones, and bringing them to the stage where they can be tested through computer implementation (on the importance of computer implementation as a benchmark: Karttunen 2006; for proposals concerning Thai tones: Prom-on 2008). Even so, instead of positing that all tones can be decomposed into levels, it is at least as reasonable to adopt the opposite standpoint, **viewing contours as nondecomposable units unless there is positive evidence to the contrary** (Nick Clements, p.c. 2008). The systematic reduction of tones to levels may well turn out to be an example of a research agenda that eventually proves fruitless, even though it originally appeared highly desirable from a theory-internal point of view. This would be like the attempt to propose universal tone features, by analogy to the features commonly used in segmental phonology (Wang 1967): recent work re-examining the discussions that have taken place in the course of four decades suggests that the extension of feature analysis to tones is not warranted, and that tones may not call for an analysis into features (Clements, Michaud and Patin 2011; Hyman 2011b; a general critique of universal features is found in Chapter 1 of Ladd 2014).

This may come as a disappointment – the failure of an attempt at increasing theoretical parsimony. But it can also be viewed as a positive result, based on converging evidence, and opening into a new research agenda. As you grapple with the difficulties of tone and intonation systems, you will have ample occasion to verify the observation *"that structural categories of language are language-particular, and we cannot take pre-established, a priori categories for granted"* (Haspelmath 2007:129). **Learning about a rich tonal typology** – not restricted to level-tone models – **is a promising starting-point**.

This situation could be compared to that of click consonant phonemes: the International Phonetic Alphabet's [ʘ ǀ ǃ ǂ ǁ]. They have a limited geographical distribution, being found essentially in Khoe-San languages (Southern Africa); but that does not detract from their phonemic status, and clicks are therefore part of the IPA as a matter of course. Clicks constitute an interesting component of student's phonetic/phonological apprenticeship, and a rich field for phonetic studies of coarticulation, speech aerodynamics and other topics (Miller et al. 2009). Likewise, phonetically complex tones are firmly attested, and interesting for phonetic/phonological research. Concerning their distribution, if they have been described as a geographical exception – *"Chinese and the 'Sinosphere' vs. the world"* (Hyman 2011a:190) – this may partly be because of researchers' greater interest in level tones,

which did not encourage an active search for instances of phonetically complex tones outside Asia. A knowledge of Asian facts is clearly a useful part of the prosodist's toolbox, e.g. when grappling with the tone systems of Oto-Manguean languages (Cruz and Woodbury 2014).

Various disciplines and techniques hold promise for gaining evidence on representations of lexical tones in different languages. Among these, some, such as brain imaging, are still in early stages of development, and will not be discussed here; others can already contribute a wealth of evidence on phonetically complex tones. This constitutes the topic of the following paragraphs.

2.2. Evidence from speech errors and word games

Evidence for non-decomposable contour tones can be found in the analysis of speech errors and word games. If a contour tone consists of a sequence of level tones, one would expect to find speech errors in which only one of these levels is omitted or replaced by another: e.g., for a L+H tone, a L or H realization – the slip of the tongue consisting in omission of one of the levels – or a L+L or H+H realization, by accidental substitution of another level. But **observations on complex tones suggest that, in tonal speech errors, one of the tones of the system is substituted holus-bolus for another** (Wan and Jaeger 1998 on Mandarin). This leads to the conclusion that *"phonological theories which require that all contour tones in every language must be represented as a sequence of level tones underlyingly may be missing an insight into the possible underlying differences among tone languages"* (Wan and Jaeger 1998:458). Similar results emerge from word games where $C_1V_1C_2V_2$ is changed to $C_1V_2C_2V_1$ or $C_2V_2C_1V_1$: speakers of Bakwiri, Dschang-Bamileke and Kru (i.e. languages with level-tone systems) leave the tone pattern unchanged, whereas speakers of Mandarin, Cantonese, Minnan and Thai (i.e. languages with complex-tone systems) tend to move the tones with the syllables (Hombert 1986:180–181).

2.3. Diachronic insights into the phonetically complex tones of Asian languages

It is extremely useful to compare dialectal variants and consult diachronic evidence where possible. This section recapitulates salient facts about the evolutions leading up to the creation of complex tones. Since the early 20th century, an increasing number of languages have come under linguistic scrutiny; together with synchronic descriptions, historical studies have attained an increasing degree of precision, and the diachronic origin of tones in many of the languages of the area is now well-understood. In addition to reviewing these findings, a hypothesis will be set out (in paragraph 2.3.3) about a possible relationship between the historical phasing of the various stages of tonogenesis and the properties of the resulting tones.

2.3.1. Background knowledge about phonation-type register systems

An important contribution to prosodic typology and to the evolutionary study of prosodic systems was made by studies of contrastive phonation-type registers, sometimes referred to as "voice quality registers" or "voice registers". **In languages**

with a phonation-type register system, phonation type has a lexically distinctive role. Thus, the Mon language has a 'clear' voice (also called 'modal' voice) register contrasting with a breathy/whispery voice register; this was still the case of Khmer less than a century ago (Henderson 1952). **Even more than other linguistic features, phonation-type registers tend to have multiple correlates**: mode of vibration of the vocal folds, but also greater duration of syllables carrying nonmodal phonation, differences in vowel articulation, and differences in F_0 (instrumental studies of phonation-type register systems include Lee 1983; Thongkum 1987, 1988, 1991; Edmondson et Gregerson 1993; Hayward et al. 1994; Watkins 2002; Wayland et Jongman 2003; DiCanio 2009; Brunelle 2009b, 2012). Phonation-type registers are one of the possible precursors of tone, as explained below.

2.3.2. Tonogenesis and registrogenesis

Tonogenesis can result from the loss of various phonemic oppositions, through a mechanism of compensation (**transphonologization**): lexical contrasts are preserved – at least in part – by means of a new phonological opposition, as illustrated in Table 1. These processes are now well understood (see Kingston 2011 for a worldwide survey). Taking the textbook case of Vietnamese (Haudricourt 1954, 1972), Table 1 recapitulates the evolution from a stage when the language did not have tone (Table 1a) up to the present-day system (1c) via a stage where there were three tonal categories (with stop-final syllables as a distinct, fourth set).

Table 1a. Late Proto-Viet-Muong: toneless stage. Open syllables without glottalization; final glottal constriction; final /h/; final /p/, /t/ or /k/ (after Ferlus 2004:24).

ta	taˀ	tah	tap, tat, tak
da	daˀ	dah	dap, dat, dak

Table 1b. First stage of tonogenesis in Vietnamese: three tones originating in earlier laryngeal finals; no tonal oppositions on stop-final syllables.

ta	tone A	ta	tone B	ta	tone C	tap, tat, tak	category D
da		da		da		dap, dat, dak	(toneless)

Table 1c. The tone system of contemporary Hanoi Vietnamese, after a tone split conditioned by the voicing feature of initial consonants. Tonal categories are provided in etymological notation, and tone names given in the orthography. Tones A1 to C2 only appear on open or nasal-final syllables, and tones D1 and D2 on stop-final syllables.

ta	tone A1 *(ngang)*	ta	tone B1 *(sắc)*	ta	tone C1 *(hỏi)*	tap, tat, tak	tone D1 *(sắc)*
ta	tone A2 *(huyền)*	ta	tone B2 *(nặng)*	ta	tone C2 *(ngã)*	tap, tat, tak	tone D2 *(nặng)*

Table 1. Vietnamese tones in diachronic perspective.

The details of the process whereby these transphonologizations take place are increasingly well understood. **Voicing oppositions have, as one of their phonetic correlates, slight differences in fundamental frequency on the following vowel**: F_0 is slightly depressed after voiced consonants, as compared with unvoiced consonants, as brought out by classical studies (see in particular Hombert 1978). In diachrony, **voicing oppositions can trigger a split of the tone system**, as from 1b to 1c in Table 1 above. The study of Khmer by Henderson (1952) provided key insights on how a voicing opposition on initial consonants could become a phonation-type register opposition, formerly voiced consonants yielding breathy/whispery phonation on the following vowel, and formerly unvoiced consonants yielding modal or tense phonation on the following vowel. Henderson's article sheds light on the diachronic links between consonants, phonation-type registers and tones: the loss of voicing oppositions on initial consonants resulted in the creation of a phonation-type register opposition, as can still be observed in Mon and Suai; the phonation-type register opposition evolved further into a two-way split of the vowel system in Khmer, and into a multiplication of the number of tones in Sinitic, Vietnamese and Thai. Here is a summary of the consequences of the weakening of the lenis (voiced) series of consonants during a voiced-unvoiced merger among initial stops:

> "...the relative laxness of the laryngeal-oral muscles keeps the consonants 'soft' during their oral closure, and this type of articulation gets prolonged into the following vowel. The relaxation of the larynx lets breathy voice come through and lowers the pitch of the voice, while the relaxation of the muscles of the mouth results in a 'lax' vowel quality. When this process takes place in languages without tones, it is the difference in vowel quality that eventually becomes distinctive. A correlation of consonants disappears as a correlation of vowels appears, decreasing the number of consonants, while increasing that of vowels. On the other hand, when this happens in a tone language, it is the change of register that becomes relevant: the correlation between consonants disappears and a tonal correlation appears, decreasing the number of consonants while causing a two-way split of the tone system." (Haudricourt 1965, translated by Paul Sidwell; about phonation types and tonogenesis, see also Egerod 1971 and Pulleyblank 1978: 173)

There is no hard-and-fast dividing line between phonation-type register systems and tone systems (Abramson and Luangthongkum 2009). It appears sufficient to adopt the following criterion: a system will be referred to as tonal if F_0 is the main cue to the opposition at issue – 'main' in the sense of primus inter pares, 'first among equals': not necessarily as the only cue. On this basis, experimental procedures can be devised to support the classification of a dialect as tonal or non-tonal (on Kammu, which has tonal and non-tonal dialects, see Svantesson and House 2006, and Karlsson, House and Svantesson 2012; on Kurtöp, which is currently undergoing tonogenesis, see Hyslop 2009).

2.3.3. The origin of complex tones vs. level tones: a hypothesis

Level tones can have various diachronic origins: they can result from the transphonologization of oppositions on initial consonants, as in Oceanic languages (the initial opposition was simple-vs.-geminated in some languages (Rivierre 1993, 2001) and voiced vs. voiceless in others (Ross 1993), or of laryngeal features of coda consonants, as in Athabaskan (Hargus and Rice 2005). As for complex tones – tones comprising phonation-type characteristics –, our conjecture is that they obtain in cases when the second stage of tonogenesis (Table 1c) begins before the first stage (Table 1b) is fully completed: before the transphonologization of final laryngeals reaches the pure-pitch tonal stage which would be its logical endpoint. Said differently, the conjecture is that **complex tones arise when there is a temporally overlapping conjunction of syllable-initial and syllable-final phonational effects**.

This would shed light on the limited geographic extent of complex tones. Their appearance requires a specific conjunction of structural properties: the inception of a split of the tone system at a stage when a previous tonogenetic process is still in progress, i.e. when the tones still preserve lingering phonation-type characteristics associated with the earlier consonantal oppositions in which they originate. This conjunction took place in a number of East and Southeast Asian languages because of similar evolutions in their syllable structure: monosyllabicization resulted in the creation of consonant-replete monosyllabic morphemes, whose gradual consonantal depletion led to the development of phonation-type registers and tones, and to an increase in the number of vowels (see Michaud 2012 for an overview).

After they come into existence, tones involving phonation-type characteristics may change, and their specific phonation type may disappear: for instance, Hanoi Vietnamese has glottalization in two of its tones, whereas Southern Vietnamese does not retain any phonation-type characteristics (Brunelle 2009a). In complex-tone systems, various types of changes in phonation types can take place over time, not just the loss of phonation-type characteristics: as can be seen from Table 1, the two tones that are currently glottalized in Hanoi Vietnamese are B2 and C2; etymologically, the syllables with a final glottal constriction correspond to present-day B1 and B2. This shows that glottalization was lost in one of these categories (B1), whereas glottalization appeared in category C2, which did not possess it originally: C2 originates in syllables with a voiced initial and a final /h/. In the vast domain of Sinitic languages ("Chinese dialects"), the tones of some varieties clearly have specific phonation (see in particular Rose 1989a, 1990) whereas for others it can be debated to what extent the nonmodal phonation which is occasionally present for some tones is part of their phonological definition: e.g. to what extent the occasional presence of laryngealization for Mandarin tone 3 is a low-level phonetic consequence of its low F_0, and to what extent it is a language-specific, phonological characteristic.

The boundary between level-tone systems and complex-tone systems appears more clear-cut from a theoretical point of view, but borderline situations are likely to exist here, too. It has been suggested that *"tone languages change type in the wake of change in morphological structure"* (Ratliff 1992a:241); language contact also plays a major role, as when a level-tone system is in contact with a complex-tone system. Such is currently the case of all the level-tone systems of China, e.g. Pumi, Naxi and Na, in a

context of non-egalitarian bilingualism (Haudricourt 1961) where the national language, Mandarin, enjoys considerable prestige. Experimental investigation into such situations of contact between languages with different tone systems appears as a promising research direction.

2.3.4. Differences in evolutionary potential between level tones and complex tones

There appear to be salient differences in evolutionary potential between different types of tone systems. **Non-decomposable tones such as those of Vietnamese, Thai and Mandarin undergo a gradual phonetic evolution** – apart from tone mergers, which are categorical and irreversible: e.g. etymological tones C1 and C2 have merged in Southern Vietnamese, so that the language has only five tones, as against six in Hanoi Vietnamese. **The evolution of level-tone systems, on the other hand, is punctuated by categorical changes**: under given circumstances, noncontrastive details in the realization of tone – i.e. conditioned allotonic variation – can be reinterpreted as differences between tonal categories; as a result, the phonological system is modified.

For instance, a comparison of Moba and Gulmancema (Rialland 2001) shows how a Top tone (super-high tone) can be created, leading to a change from a three-level system to a four level system. Gulmancema is more conservative than Moba: it has a three-level system (H, M, and L). In Gulmancema, there exists a phonetic precursor to the creation of a fourth level: a H tone preceding a L tone is phonetically raised. For instance, the syllable /^{H}kan/ will be realized phonetically higher in the sequence / ^{LM}o ^{H}kan ^{L}di / 'he stepped over', where it precedes a L tone, than in / ^{LM}o ^{H}kan ^{H}di / 'he steps over', where it precedes another H tone. This phonetic phenomenon does not affect the phonological nature of the tones. The closely related language Moba, on the other hand, is innovative: word-final vowels disappeared, as shown in Table 2; but the opposition between sentence pairs such as 'he stepped over' vs 'he steps over' is maintained. **The super-high phonetic variant of the high tone has gained contrastive status**: a lexical Top tone has emerged.

meaning	Gulmancema	Moba
he stepped over	^{LM}o ^{H}kan ^{L}di (ŏ kándì)	^{LM}u ^{XH}kant (ŭ kấnt)
he steps over	^{LM}o ^{H}kan ^{H}di (ŏ kándí)	^{LM}u ^{H}kant (ŭ kánt)

Table 2. A comparison showing the origin of the Top tone of Moba. Data and analysis from Rialland (2001:317). Tone is indicated in superscript at the beginning of the tone-bearing syllable. XH=extra-high tone (Top tone).

This is a case of transfer of distinctiveness, from the tone of the word-final vowel to the tone that precedes. Allotonic variation (as evidenced by Gulmancema) paves the way for diachronic change, but the change itself – the modification of the tone system – is triggered by the loss of final vowels.

Level-tone systems (like all linguistic systems) evolve in time, in ways which are increasingly well-documented; Boyeldieu (2009) sheds light on the development of new categories in a level-tone system. But we have not come across reports showing that, in a level-tone system, the tones could undergo a gradual change in phonetic shape, e.g. a H tone gradually acquiring a final downward tilt and eventually becoming H+L. By contrast, unitary contours are subject to gradual change. **The evolution of unitary contours can take place without any conspicuous phonological change**. These tones are defined in terms of an overall contour – as well as phonation-type characteristics in some cases –, which can vary somewhat so long as the oppositions among the tones present in the language are preserved. The Tamangic group of Sino-Tibetan is an especially well-documented example, revealing various evolutionary stages reflected in the spatial diversity of dialects, as well as a remarkable amount of cross-speaker differences within the same village, and even for one and the same speaker. Risiangku Tamang illustrates an early stage in the gradual phonetic evolution of complex tones: the four tones of this language, breaking off their last ties with the earlier voicing correlation on initial consonants, become free to evolve away from their original F_0 range, namely: relatively lower tones after former voiced initials, higher after former unvoiced initials. The evolution is more advanced in Marphali and in Taglung Tamang, where tone 4, which etymologically belongs in the low series, is now phonetically high; likewise for tone 3 in Manangke (Mazaudon 2005, 2012; Hildebrandt 2003). ***"Once it is established, the tonal system evolves without regard for its old etymological pitch levels"*** (Haudricourt 1972:63; see also Ratliff 2010:224).

These arguments drawn from dialectology are confirmed by phonetic evidence, in the case of tone systems for which there is a sufficient time depth in experimental studies (although special precautions must be taken in interpreting phonetic data that were collected with widely different setups). A well-described example is Bangkok Thai, which has been documented experimentally at intervals for a hundred years. For instance, the tone which in 1908 was the highest, with a final fall, has now become rising (Pittayaporn 2007:fig. 2). The number of distinctive tones has remained the same; their phonetic evolution is gradual, and the evolution of one tone has consequences on that of the other tones with which the risks of confusion are greatest. In this process, one sees at play the familiar antagonist forces of (i) the tendency towards simplification, on the one hand, and (ii) the pressure towards the preservation of distinctive oppositions, on the other. (On the evolution of the tone systems of Tai languages in Northeast India: Morey 2005.)

From a synchronic point of view, there is no difficulty in proposing a level-tone analysis for any system, for instance labelling the five tones of Bangkok Thai as H, L, H+L, L+H and zero on the basis of a stylization of F_0 tracings (for an example of such analyses: Morén and Zsiga 2007). But the linguist is then at a loss to describe the diachronic change mentioned above: how come Thai tone 4 changed from high-falling to rising? Dialectal data offer a geographical projection of different stages of the evolution of tone systems; they provide confirmation of the view that complex tones undergo a gradual evolution, as was mentioned above for Tamang. Here as in many other cases, dialectology and diachrony provide precious insights for phonological modelling. (See also a detailed argument concerning the limitations of a flatly synchronic description of the Cantonese tone system: Yu 2003.)

3. Beyond lexical tone: How common are intonational tones?

3.1. Some definitions

As a preliminary to the discussion of intonation, some terminological clarifications appear useful. To take the example of the term 'tone', in some models it is synonymous with F_0: Hyman and Monaka (2008) define the term 'tonal' in a phonetic sense, to mean 'realized by F_0', and 'non-tonal' to mean 'realized by parameters other than F_0' (such as phonation types). The equation between 'tone' and 'F_0' (and its perceptual counterpart: pitch) appears so self-evident that it could seem unnatural to try to define tone in any different way. But from a classical linguistic perspective, **it appears crucial to make a distinction between F_0, which is an acoustic parameter, and linguistic tone, which is a functional concept**. For the present discussion, a central point is the division of prosody into several levels, distinguishing **intonation** from **lexical prosody** and **morphological prosody** (Rossi 1999; Vaissière 2002, 2004; see also the discussion in Zerbian 2010).

Lexically distinctive prosody includes **stress**, as in English, Russian, and Spanish, **tone**, as in Yoruba, Vietnamese, and Puinave, and **phonation-type register**, as in Mon, Kammu and Cham. There are also languages without any form of lexical prosody, e.g. Newar (Genetti 2007:69–89), Hindi and French. Morphologically distinctive prosody is found in fewer languages than lexically distinctive prosody; however, a survey shows that *"tonal morphology (...) exhibits essentially the same range of morphological properties as in all of segmental morphology"* (Hyman and Leben 2000:588; about East/Southeast Asian languages, see Henderson 1967; Downer 1967; Ratliff 1992b). 'Tone' is therefore used here mainly in the sense of lexical and morphological tone, although (as will be discussed in this section) **there exist some languages that possess intonational tones: tones that encode intonation, and that are formally identical with lexical tones**.

Tone has the function of lexical and morphological differentiation, and intonation the functions of speech phrasing, of coding prominence and sentence mode, and of expressing emotions and attitudes towards the listener. Intonation is, in Bolinger's phrase, a *"half-tamed savage"* (Bolinger 1978:475). **Phrasing** is on the tamer, more intellectual side; it surfaces at its clearest in deliberate oral renderings of elaborately composed texts. **Prominence** is a less tame dimension of intonation: it can still be described in terms of a linguistic system, with clear cross-linguistic differences, but the intrusion of the stronger manifestations of prominence can interfere with phrasing as determined by syntactic structure. As for the expression of sentence mode, attitudes and emotions, it can partly be described in terms of ethological principles, such as the "Frequency Code" (Ohala 1983).

3.2. Instances of intonational tones

There are some well-established cases where intonation is encoded by tones that are treated on a par with lexical tones and morphological tones: **in some tonal languages, tone can serve as a marker for functions at the phrasal level. These will be referred to as intonational tones.** This extension of the notion of tone beyond its

primary meaning (lexical and morphological tone) is made in view of the structural similarities between lexical and morphological tone, on the one hand, and certain intonational phenomena, on the other hand; it does by no means amount to a broadening of the concept of tone to intonational phenomena in general, as is the case in some versions of autosegmental-metrical models of intonation (discussed in section 3.4).

First, tone may indicate sentence mode.

> "The most commonly encountered cases involve a tonal means to distinguish interrogatives from declaratives. In Hausa, a L is added after the rightmost lexical H in a yes/no question, fusing with any pre-existing lexical L that may have followed the rightmost H (…). As a result, lexical tonal contrasts are neutralized. In statements, [kái] 'head' is tonally distinct from [kâí] 'you [masculine]'. But at the end of a yes/no question, they are identical, consisting of an extra-H gliding down to a raised L." (Hyman and Leben 2000:61)

This is described as a case of intonational tone, rather than a case of superimposition of an intonational pattern onto an underlyingly unchanged tone sequence. On superficial examination, one could be tempted to say that intonation in English or German is likewise expressed by tones: the final rise often found in question would be a rising tone, contrasting with a different tone for statements. But in-depth phonetic investigation reveals that the perception of sentence mode in German is influenced by *"shape, slope, and alignment differences of the preceding prenuclear pitch accent"* (Petrone and Niebuhr 2013), i.e. **question intonation in German is distributed over the utterance, quite unlike the addition of a final L tone in Hausa yes/no questions**.

Second, **tone may serve phrasing functions**. In some languages, certain junctures of the utterance are characterized by the addition of boundary tones, which, though introduced by post-lexical rules, are integrated into the tone sequence of the utterance on a par with lexical tones. L. Hyman (personal communication) points out that such phenomena are *"rampant in African tone systems"*, taking the example of a phrase-final boundary tone in Luganda: this tone is transcribed as H%, where the %, representing a boundary, is a functional indication of the tone's origin. It acts just like any level tone, except that it is inserted into the tonal string later than the lexical tones. Any sequence of preceding toneless moras will be raised to that H level (though there has to remain at least one L before it). For example, /omulimi/ 'farmer' is pronounced all-L as subject of a sentence (/òmùlìmì/), but at the end of an utterance marked by this H%, it is pronounced L-H-H-H: /òmúlímí/.

A third intonational function that may be served by tone is to convey prominence. A clear example of intonational tone (a tone of intonational origin) is encountered in Naxi, a Sino-Tibetan tonal language: a word that carries lexical L or M tone on its last syllable can be focused by addition of a H tone that aligns at the right edge of the word, causing the tone of the last syllable to become rising (Michaud 2006:72).

In order to understand how intonational tones emerge and evolve, it appears interesting to examine not only clear-cut cases as those reviewed in this paragraph, but also doubtful cases of intonational tones.

3.3. Doubtful cases of intonational tones: crossing the fine line between intonation and tone?

Scholars have long been aware of the phonetic similarities between (phrasal) intonation and (syllable-based) tones. In the mid-17th century, the European authors who devised a Latin-based writing system for Vietnamese (de Rhodes 1651) had to develop a notation for a six-way tonal contrast. One of the tones was left unmarked; grave and acute accents were used for two others, and tilde for a fourth one. For the remaining two tones, symbols from sentence-level punctuation were used: the full stop was added (below the vowel) to indicate tone B2 (orthographic *nặng*) on the basis of the perceived similarity between its final glottal constriction and the intonational expression of **finality**; and the **question mark** (in reduced form, on top of the vowel) was used for tone C1 (orthographic *hỏi*) due to its final rise (Haudricourt [1949] 2010). To the authors of this system, the newly coined tone marks served as mnemonic cues to the pronunciation of tone, via an analogy with intonation in Romance languages. In this instance, there is no possible confusion between lexical tone and intonation; but there exist cases where a language's lexical tones are reported to serve intonational purposes. Phake (Tai-Kadai language family) exemplifies the diversity of situations found in Asian languages.

3.3.1. The expression of negation and sentence mode in Phake

Phake, a Tai-Kadai language of Assam (India), has six lexical tones, and cases of *"changed tones"* (Morey 2008:234–240). There are two different processes.

(i) If a verb has the second tone (High falling), it changes to rising when negated. This rising tone is identical in form to the rising tone (no. 6); this is perceived by the speakers as a categorical tone change. This process appears to be spreading to verbs carrying other tones (S. Morey, p.c.).

(ii) According to observations made in the 1960s and 1970s, changing the lexical tone of the last syllable in a sentence to the sixth tone (a rising tone) would express a question (Banchob 1987).

More recent fieldwork reports the same phenomenon, but instead of identifying the "changed tone" with one of the six lexical tones, it is suggested that it is *"a special questioning tone (...). This questioning tone first rises and then falls, and here is arbitrarily notated as 7"* (Morey 2008:234). Finally, an eighth tone is reported: an "imperative tone", *"that exhibited glottal constriction and creaky voice"* (Morey 2008:239).

Observation (ii) can be reinterpreted as cases of neutralization of tonal oppositions: it does not appear implausible that question intonation in Phake overrides the lexical tone of the sentence's last syllable in questions. Likewise, imperative intonation in Phake has a salient influence which may go so far as to override the lexical tone of some syllables. *"The fluctuating needs of communication and expression are reflected more directly and immediately in intonation than in any other section of the phonic system"* (Martinet 1957). The phonation type associated to imperative mode – a

contraction of the laryngeal sphincter, to convey an attitude of authority – appears to have a clear iconic motivation (see Fónagy 1983:113–126).

It is perhaps significant that "changed tones" are reported in an area where the dominant languages are non-tonal. Speakers of Phake are also fluent in Assamese, a non-tonal language, which may create a pressure towards the simplification of the Phake tone system, e.g. through neutralization of tonal contrasts in some contexts. Overall, it would seem that intonation does not easily win the day over lexical tone. Some experimental evidence on this topic comes from a study of the Austroasiatic language Kammu, one of few languages with two dialects whose only major phonological difference is the presence or absence of lexical tones. A comparison of the two dialects concludes that the intonational systems of the two Kammu dialects are basically identical, and that the main differences between the dialects are adaptations of intonation patterns to the lexical tones when the identities of the tones are jeopardized (Karlsson, House and Svantesson 2012).

3.3.2. Mandarin interjections: a case of spurious tonal identification

The treatment of the interjection /a/ (transcribed as 啊 in Chinese writing) in a learner's dictionary of Standard Mandarin offers a clear case of spurious tonal identification. This dictionary treats the interjection as if it had lexical tone, and sets up four distinct entries for it, corresponding to the four tones of Standard Mandarin: with tone 1, the interjection would mean "speaker gets to know something pleasant"; with tone 2, it would signal a "call for repetition"; with tone 3, "surprise or disbelief"; and with tone 4, the "sudden realization of something" (Huangfu Qinglian 1994, entry "a"). This categorization is based on phonetic similarities between the pitch patterns of the four tones and intonational variants of the interjection, as recapitulated in Table 3.

There is in fact a considerable phonetic difference between the four-way division of the Mandarin tonal space, on the one hand, and the intonational gradations in the realization of interjections, on the other – involving not only F_0, but also length and other parameters. Interestingly, the authors of the dictionary gloss the "tone-4" realization of the interjection /a/ as the "**sudden** realization of something" (emphasis added). The interjection /a/ can just as well convey the realization of something, without any specific hint of suddenness (Lin Yutang 1972, entry "啊"). The F_0 of the interjection decreases gradually, in a manner that does not resemble tone 4 (an abruptly falling tone). The mention of suddenness was added because the intonational signalling of this extra nuance tends to shorten the interjection, thereby creating a surface similarity with tone 4. From the point of view of linguistic functions, there should be no confusion: **the phonetic realization of interjections in Standard Mandarin is purely intonational, *"with varying, indeterminate accent, like English Oh! ah! aha!"*** (Lin Yutang 1972, entry "啊"). Mandarin interjections bypass tonal coding; the interjection /a/ has a wide range of possible realizations, and of expressive effects. The four entries set up for this interjection in the dictionary single out four of these realizations, and grant them separate status merely because they happen to be phonetically close to the language's four lexical tones. This example illustrates the potential for a **misinterpretation of intonational phenomena as tonal**.

We now turn to an examination of "autosegmental-metrical" models of intonation. We would like to warn students about pitfalls of these models, which base the representation of intonation on "tones".

tone	characterization in dictionary	example	translation of example	F_0 on interjection	canonical realization of tone
1	"speaker gets to know something pleasant"	啊！我考过了！ ā ! wǒ kǎo guò-le !	Wow! I passed the exam!	overall high F_0	level, in the upper part of the speaker's range
2	"call for repetition"	啊，是吗？ á, shì ma?	Oh, is that right?	rising	rising
3	"surprise or disbelief"	啊？ 你在这儿干吗? ǎ? nǐ zài zhèr gànmá?	Huh? What are you doing here?	falling-rising	falling from mid-low to lowest, with final rise in isolation
4	"sudden realization of something"	啊，现在我知道了。 à, xiànzài wǒ zhīdào-le.	Aha! Now I understand.	falling	sharply falling, from high starting-point

Table 3. Phonetic basis for the four-way categorization of the nuances expressed by the interjection /a/ in Mandarin, as proposed in a dictionary: Huangfu Qinglian 1994.

3.4. How autosegmental-metrical models blur the distinction between tone and intonation

Autosegmental-metrical models are based on concepts from Subsaharan tonology (for a review: Rialland 1998). **In autosegmental-metrical models, "*lexical pitch variations and intonational pitch variations are phonologically represented as tones, like H(igh) and L(ow)*"** (Gussenhoven 2004:xvii). The exciting paradox of these models is summarized in John Goldsmith's paradoxical description of *"English as a tone language"* (Goldsmith 1981): describing intonation patterns (in languages such as English) with the same tools as the tones of Bantu languages. In the Tones and Break Indices (ToBI) system, proposed as *"a standard for labelling English prosody" (Silverman et al. 1992), intonation is transcribed "as a sequence of high (H) and low (L) tones marked with diacritics indicating their intonational function as parts of pitch accents or as phrase tones marking the edges of two types of intonationally marked prosodic units"* (Beckman and Elam 1997:8).

In the wake of ToBI, adaptations for a wide range of other languages were developed. Leading researchers, some of whom initially argued against the modelling of intonation into discrete levels (Ladd 1978), now advocate the autosegmental-metrical model, which has become the mainstream model of intonation (Ladd 1996; Gussenhoven 2004; Jun 2005). Autosegmental-metrical models operate with the same concepts for all languages, as a matter of definition.[5]

[5] Notions from autosegmental studies of tone that are carried over into intonation studies include **downstep**: the categorical lowering of a High tone with respect to preceding High tones, typically due to an intervening Low tone, either overt or 'floating' (Connell 2001; Rialland 1997, 2001). (An argument for the usefulness of the

At this point it is worth pausing and asking to what extent the use of the same concepts for tone and intonation is useful and enlightening. Let us take as an example the notation "H%", for a H boundary tone. It is used for tonal languages such as Kinande (Hyman 2010:207); for English (Pierrehumbert and Steele 1989); for French (Fougeron and Jun 1998); and for Vietnamese (Ha and Grice 2010), among other languages.

The linguistic facts are much less homogeneous than the use of the same label suggests, however.

- In Kinande, the H% which marks the end of a phrase is **a real (*bona fide*) tone, which triggers the same phonological processes as H tones of lexical origin**, e.g. causing neutralization of certain lexical tone oppositions on nouns when they are said in isolation (Hyman 2010:207).
- In French, in the absence of tones at the lexical level or at the morphological level, there is no language-internal evidence to decide whether the phenomenon at issue is tonal or not. (On the description of French intonation, without recourse to the notion of tone: see Delattre 1966; Rossi 1999; Vaissière and Michaud 2006; Martin 2009.)
- For Vietnamese, H% is used for "rising final pitch movements" by Ha and Grice (2010); like in French, this choice of label is based on theory-internal motivations. Unlike in Kinande, there are no reasons to identify the phenomenon labelled as H% with one of the lexical tones of Vietnamese.

The adoption of the same label, H%, may appear economical from the point of view of a universal model, but this leads to an artificial and counter-intuitive description. This is analogous to the use of the feature /ATR/ (Advanced Tongue Root) in the description of vowel systems. This feature was initially proposed for a set of languages of Africa; a crucial phonological test is the presence of ATR vowel harmony (see e.g. Hess 1992). The feature was later used to describe the four-way opposition in a vowel system containing /i-e-ɛ-a/ (Calabrese 2000). The argument is theory-internal: combining a binary /ATR/ feature with a binary /open/ feature yields 2x2=4 values, and thus obviates the need for a multi-valued /open/ feature for vowels, considered uneconomical. **The pinch comes when typological considerations come in**: should French, and other Romance languages, be included in cross-linguistic studies of ATR? A common-sense answer is that it is best to begin by identifying a core set of languages that uncontroversially possess ATR systems, and to apply due caution when considering extensions of the concept beyond this core domain.

To return to intonation, the generalized use of boundary-tone labels such as H% and L%, and of tonal notations such as H*, L*, L+H*, L*+H for "pitch accents", **veils typological differences** (Ladd 2008). In most East and Southeast Asian languages, the available literature suggests that **intonation does not seem to be implemented by the addition of tones** in the way described for Kinande, Hausa, Luganda and Naxi.

---

notion of downstep in the study of intonation is proposed by Ladd 1993.) Laniran and Clements (2003) point out that there are paradoxically more phonetic studies of downstep in widely-studied languages (e.g. English, Swedish, German, Dutch, or French) where it is posited as a component of intonation, than in the African languages where downstep was initially reported as a component of the tone system.

The widely-studied case of Standard Mandarin provides an example. Mandarin has salient intonational phenomena, which have a strong influence on the phonetic realization of tones, to the extent of making the automatic recognition of tone in continuous speech a great challenge. But **these intonational phenomena do not affect the phonological identity of the lexical tones**. Instead, intonation is superimposed on tone sequences. From the point of view of linguistic structure, intonation remains on an altogether different plane from tones: it does not modify the phonological sequence of tones, even in cases where it exerts a considerable influence on their phonetic realization. This has been studied since the pioneering work of Chao Yuen-ren (1929).

Relevant evidence on this issue comes from the field of speech synthesis: some specialists choose to specify (i) full templates of the time course of F_0 for each lexical tone, and (ii) a "strength coefficient" for each syllable (Kochanski, Shih Chilin and Jing Hongyan 2003; Kochanski and Shih Chilin 2003). The strength coefficient, which correlates with informational prominence, plays a major role in the final shape of the synthesized F_0 curve. This synthesis system provides indirect confirmation for the observation that, although intonational parameters interact with the phonetic realization of tone, they do not modify the underlying phonological sequence of tones: there is no insertion or deletion of tones. The informational prominence of a syllable is indicated by local phenomena of curve expansion and lengthening on the target syllable, as well as some modifications in supraglottal articulation; conversely, a degree of phonetic reduction is found on other syllables, including a degree of post-focus compression of F_0 range (see in particular Xu 1999).

3.5. Some suggested topics and directions for prosody research

Generalized use of the notion of 'tone' in autosegmental-metrical models of intonation has paradoxically slowed down the identification of real examples of intonational tones. The observation that intonational tones are uncommon across languages opens into interesting linguistic issues such as

(i) their relative frequency across languages: how many languages have tonal encoding for (some) intonational functions?

(ii) the factors facilitating their development, and their relationship to a language's lexical tones, morphological tones, and intonation system.

Concerning possible relationships between tone systems and intonation systems, it seems intuitively clear that multilevel tone systems (e.g. Ngamambo, Wobe) cannot allow the type of intonational flexibility in the realization of tone which is pervasive in Mandarin or Vietnamese, because such flexibility would jeopardize the identification of the utterance's underlying tonal string. All other things being equal, it would seem that level tones constrain intonation to a greater extent than complex tones. The typological hypothesis could be phrased as follows: prosodic systems based on discrete pitch levels allow less allotonic variation, so that less information about phrasing and prominence can be encoded as modulations of F_0 superimposed on the tonal string. This would create a pressure towards privileging other means to convey phrasing and prominence: either by integration into the tonal string (i.e., **intonational tones** as defined here), or

by the use of nonintonational means, such as the use of topic/focus morphemes to convey information structure.

Another typological parameter that may favour the development of intonational tones is the presence of **morphological tone**. In Kifuliiru (Van Otterloo 2011), for instance, tone serves not only a lexically distinctive function but also complex morphophonological functions, so that the surface-phonological tone sequence for the utterance obtains through the application of a large set of categorical processes. In this language, prosodic structure hinges on the **calculation** of a tone sequence; the phonetic implementation of this tone sequence is reported to be relatively straightforward. At another extreme of the typological continuum, in Vietnamese, no distinction needs to be made between a lexical and a surface-phonological level – the tones are the same at both levels. (For a review of tone rule systems in Africa and Asia: Chen 1992.) **There may exist a relationship between the functional load of morphological tone and the degree of development of intonational tones**. There exists a well-documented tendency for segmental morphology to become tonal in languages that have lexical tone: it has even been proposed that *"tone languages seem to start to lose segmental morphology, with consequent transfer of its function to the tonal plane, almost as soon as they begin to acquire it"* (Ratliff 1992a:242). In turn, **the presence of morphological tone is likely to facilitate the development of intonational tones**: a tonal reinterpretation of certain aspects of intonation can be structurally economical in systems where tone already plays morphological functions. It does not seem to be coincidental that tonal change in Burmese, which serves a few grammatical functions, is also reported to serve an intonational function. In Burmese, which has a phonetically complex system of four tones, shift from the "low" and (less often) "heavy" tone to "creaky" can indicate grammatical relationships such as possession ("ŋa" '1SG', "ŋá" '1SG.POSS; my'); it is also reported that the Burmese vocative morpheme *ye* has its tone changed from "heavy" to "creaky" to express impatience (Okell 1969:20; Wheatley 2003:198).[6]

Conversely, the conspicuous absence of any categorical processes operating over tone sequences in Vietnamese, a language which has neither tonal morphology nor categorical tone sandhi, may go a long way towards explaining why this language does not develop intonational tones.[7]

By paying attention to such topics, you will be able to make a contribution towards the long-term tasks of (i) describing the diversity of the world's intonation systems, and (ii) putting together a set of parameters that will allow prosodic typology to capture the entire spectrum of the world's prosodic systems.

[6] Needless to say, experimental verification would be useful to verify whether this attitudinal contrast is really encoded as a categorical tone change to "creaky" tone.

[7] This hypothesis is not intended as a generalization about all languages of East/Southeast Asia. For instance, languages of the Wu branch of Sinitic display great diversity in their tone systems, and the issue of whether the notion of intonational tones usefully applies to them is here left open.

4. Conclusion

This article attempted to step back and take stock of the available evidence on tone and intonation. Two observations were emphasized: (i) not all lexical tones lend themselves to an analysis into levels; and (ii) not all languages have intonational tones: tones of intonational origin that are formally identical with lexical tones. The first observation can be considered fairly uncontroversial, in view of the wealth of converging evidence available in the literature. The second, on the other hand, remains controversial: it amounts to calling into question some assumptions that underpin "autosegmental-metrical" models of intonation.

Intonation is about conveying shades and nuances through fine modulation, not only of fundamental frequency, but of all aspects of speech production. This versatility is at odds with the linguist's aim to arrive at hard-and-fast conclusions: *"to establish those pitch movements that are interpreted as relevant by the listener"*, *"characterized by discrete commands to the vocal cords" and recoverable "as so many discrete events in the resulting pitch contours"* (Cohen and 't Hart 1967:177–178). When doing research on prosody, you too may become convinced, at some point, that you have managed to pinpoint intonational facts that have a cognitive basis and constitute primitives of intonation as a human faculty. If so, pause and think twice about the nature of the discovery, remembering that, in phonetics and phonology, "the devil is in the detail" (Nolan 1999; for caveats on the acoustic complexity of intonation: Niebuhr 2013).

In conclusion, what we propose as a "point to take home" is that the distinction between tone and intonation should serve as a backbone of prosody research. **Tone** is a broad field of linguistics, not a homogeneous and well-defined concept. **Intonation** in each language should be described on its own terms; cross-linguistic comparison and cross-linguistic notions are an important part of the method, but one should refrain from jumping to conclusions and generalizations on the basis of concepts proposed as universals, such as H and L tones, downstep, "pitch accents" (such as "L+H*"), and boundary tones. You should judge for yourself to what extent notions carried over from research on certain types of tone systems can usefully be applied to the data under study.

The evidence reviewed here clearly establishes (in our view) that there are tone systems for which autosegmental models do not tell the full story. New developments in the study of these tones could provide impetus for new developments in intonation models.

5. Acknowledgments

Many thanks to Marc Brunelle, Christian DiCanio, Donna Erickson, Michel Ferlus, Alexandre François, Wentao Gu, David House, Larry Hyman, Guillaume Jacques, James Kirby, Đăng-Khoa Mạc, Martine Mazaudon, Stephen Morey, Oliver Niebuhr, Pittayawat Pittayaporn, Bert Remijsen, Mario Rossi, and four anonymous reviewers, who, each in their own way, generously provided insightful comments and suggestions. Needless to say, none of them is to be held responsible for the views expressed here: the authors assume full responsibility for the contents of this article.

Support from Agence Nationale de la Recherche (HimalCo project, ANR-12-CORP-0006, and LabEx "Empirical Foundations of Linguistics", ANR-10-LABX-0083) is gratefully acknowledged.

6. References

ABRAMSON, ARTHUR / THERAPAN LUANGTHONGKUM. 2009. A fuzzy boundary between tone languages and voice-register languages. In: Gunnar Fant / Hiroya Fujisaki / J. Shen (Eds.), Frontiers in phonetics and speech science (pp. 149–155). Beijing: Commercial Press.

ANDRUSKI, JEAN E. / JAMES COSTELLO. 2004. Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong. Journal of the International Phonetic Association 34, 125–140.

ANDRUSKI, JEAN / MARTHA RATLIFF. 2000. Phonation types in production of phonological tone: the case of Green Mong. Journal of the International Phonetic Association 30, 39–62.

BANCHOB, BANDHUMEDHA. 1987. Phake-Thai-English dictionary. Assam, India: manuscript published by the author.

BAO, ZHIMING. 1999. The Structure of Tone. New York/Oxford: Oxford University Press.

BARRIE, MICHAEL. 2007. Contour tones and contrast in Chinese languages. Journal of East Asian Linguistics 16, 337–362.

BECKMAN, MARY / G. ELAM. 1997. Guidelines for ToBI labeling, version 3.0. The Ohio State University Research Foundation.

BOLINGER, DWIGHT LE MERTON. 1978. Intonation across languages. In: Joseph H. Greenberg (Ed.), Universals of Human Language, Vol. 2, Phonology (pp. 471–524). Stanford: Stanford University Press.

BOYELDIEU, PASCAL. 2009. Le quatrième ton du yulu. Journal of African languages and Linguistics 30, 193–230.

BRUNELLE, MARC. 2009a. Tone perception in Northern and Southern Vietnamese. Journal of Phonetics 37, 79–96.

BRUNELLE, MARC. 2009b. Contact-induced change? Register in three Cham dialects. Journal of Southeast Asian Linguistics 2. 1–22.

BRUNELLE, MARC. 2012. Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. Journal of the Acoustical Society of America 131, 3088–3102.

BRUNELLE, MARC / NGUYE KHAC HUNG / NGUYEN DUY DUONG. 2010. A Laryngographic and Laryngoscopic Study of Northern Vietnamese Tones. Phonetica 67, 147–169.

CALABRESE, ANDREA. 2000. The feature [Advanced Tongue Root] and vowel fronting in Romance. In: Lori Repetti (Ed.), Phonological theory and the dialects of Italy (pp. 59–88). Amsterdam: John Benjamins.

CHAO, YUEN-REN. 1929. Beiping intonation [in Chinese]. A.A. Milne: The Camberley Triangle, (Appendix). Shanghai: Zhonghua Bookstore.

CHAO, YUEN-REN. 1933. Tone and intonation in Chinese. Bulletin of the Institute of History and Philology, Academia Sinica 4, 121–134.

CHEN, MATTHEW Y. 1992. Tone rule typology. Proceedings of the Annual Meeting of the Berkeley Linguistics Society 18, 54–66.

CLEMENTS, NICK / JOHN GOLDSMITH. 1984. Autosegmental studies in Bantu Tone. In: Nick Clements / D.L. Goyvaerts (Eds), Publications in African Languages and Linguistics. Dordrecht: Foris.

CLEMENTS, NICK / ALEXIS MICHAUD / CEDRIC PATIN. 2011. Do we need tone features? In: Elizabeth Hume / John Goldsmith / W. Leo Wetzels (Eds), Tones and Features (pp. 3–24). Berlin: De Gruyter Mouton.

CLEMENTS, NICK / ANNIE RIALLAND. 2007. Africa as a phonological area. In: Bernd Heine / Derek Nurse (Eds.), A linguistic geography of Africa (pp. 36–85). Cambridge: Cambridge University Press.

COHEN, A. / JOHAN 'T HART. 1967. On the anatomy of intonation. Lingua 19, 177–192.

CONNELL, BRUCE. 2001. Downdrift, Downstep, and Declination. Typology of African Prosodic Systems Workshop. Bielefeld University, Germany.
CRUTTENDEN, ALAN. 1986. Intonation. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

CRUZ, EMILIANA / TONY WOODBURY. 2014. Finding a way into a family of tone languages: The story and methods of the Chatino language documentation project. Language Documentation and Conservation 8, 490–524.

DELATTRE, PIERRE. 1966. Les dix intonations de base du français. The French Review 40, 1–14.

DiCanio, Christian. 2009. The phonetics of register in Takhian Thong Chong. Journal of the International Phonetic Association 39, 162–188.

DiCanio, Christian. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. Journal of Phonetics 40, 162–176.

Dihoff, Ivan R. 1977. Aspects of the tonal structure of Chori. University of Wisconsin.

Ding, Picus Sizhi. 2001. The pitch accent system of Niuwozi Prinmi. Linguistics of the Tibeto-Burman Area 24, 57–83.

Donohue, Mark. 2003. The tonal system of Skou, New Guinea. Proc. Symposium on Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone, Tokyo, Japan, 329–364.

Donohue, Mark. 2005. Tone and the Trans New Guinea languages. Proc. Symposium on Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone, Tokyo, Japan, 33–54

Downer, Gordon B. 1967. Tone change and tone shift in White Miao. Bulletin of the School of Oriental and African Studies 30, 589–599.

Edmondson, Jerold A. / John Esling / Jimmy G. Harris / Li Shaoni / Lama Ziwo. 2001. The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. Mon-Khmer Studies 31, 83–100.

Edmondson, Jerold A. / Kenneth J. Gregerson. 1993. Western Cham as a register language. In: Jerold A. Edmondson / Kenneth J. Gregerson (Eds), Tonality in Austronesian Languages (pp. 61–74). Honolulu: University of Hawai'i Press.

Egerod, Søren Christian. 1971. Phonation types in Chinese and South East Asian languages. Acta Linguistica Hafniensia: International Journal of Linguistics 13, 159–171.

Evans, Jonathan. 2008. "African" tone in the Sinosphere. Language and Linguistics 9, 463–490.

Ferlus, Michel. 2004. The Origin of Tones in Viet-Muong. Proc. 11th Annual Conference of the Southeast Asian Linguistics Society, Tempe, Arizona, USA, 297–313.

Fónagy, Ivan. 1983. La vive voix: essais de psycho-phonétique. Paris: Payot.

Fougeron, Cécile / Sun-Ah Jun. 1998. Rate effects on French intonation: prosodic organization and phonetic realization. Journal of Phonetics 26, 45–69.

GARELLEK, MARC / PATRICIA KEATING. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. Journal of the International Phonetic Association 41, 185–205.

GENETTI, CAROL. 2007. A Grammar of Dolakha Newar. Berlin: De Gruyter Mouton.

GIRÓN HIGUITA, JESUS-MARIO / W. LEO WETZELS. 2007. Tone in Wãnsöhöt (Puinave). In: W. Leo Wetzels (Ed.), Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area (pp. 129–156). Leiden: Publications of the Research School of Asian, African and Amerindian Studies (CNWS).

GOLDSMITH, JOHN. 1981. English as a tone language. In: D.L. Goyvaerts (Ed.), Phonology in the 1980s (pp. 287–308). Ghent: Story-Scientia.

GOMEZ-IMBERT, ELSA. 2001. More on the Tone versus Pitch Accent Typology: Evidence from Barasana and Other Eastern Tukanoan Languages. Proc. Symposium on Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics, Tokyo, Japan, 369–412.

GUSSENHOVEN, CARLOS. 2004. The phonology of tone and intonation. (Research Surveys in Linguistics). Cambridge: Cambridge University Press.

GUTHRIE, MALCOM. 1940. Tone ranges in a two-tone language (Lingala). Bulletin of the School of Oriental and African Studies 10, 469–478.

HA, KIEU-PHUONG / MARTINE GRICE. 2010. Modelling the interaction of intonation and lexical tone in Vietnamese. Proceedings of Speech Prosody 2010. Chicago, USA.

HARGUS, SHARON / KEREN RICE. 2005. Athabaskan Prosody. (Current Issues in Linguistic Theory 269). Amsterdam/Philadelphia: John Benjamins.

HASPELMATH, MARTIN. 2007. Pre-established categories don't exist: consequences for language description and typology. Linguistic Typology 11, 119–132.
HAUDRICOURT, ANDRÉ-GEORGES. 1954. De l'origine des tons en vietnamien. Journal Asiatique 242, 69–82.

HAUDRICOURT, ANDRÉ-GEORGES. 1961. Richesse en phonèmes et richesse en locuteurs. L'Homme 1, 5–10.

HAUDRICOURT, ANDRÉ-GEORGES. 1965. Les mutations consonantiques des occlusives initiales en môn-khmer. Bulletin de la Société de Linguistique de Paris 60, 160–72.

HAUDRICOURT, ANDRÉ-GEORGES. 1972. Two-way and three-way splitting of tonal systems in some Far Eastern languages (Translated by Christopher Court). In: Jimmy

G. Harris / Richard B. Noss (Eds), Tai phonetics and phonology (pp. 58–86). Bangkok: Central Institute of English Language, Mahidol University.

HAUDRICOURT, ANDRÉ-GEORGES. 2010. The origin of the peculiarities of the Vietnamese alphabet. Mon-Khmer Studies 39, 89–104.

HAYWARD, KATRINA / D. GRAFIELD-DAVIES / B.J. HOWARD / J. LATIF / R. ALLEN. 1994. Javanese stop consonants: the role of the vocal folds. London: School of Oriental and African Studies, University of London.

HENDERSON, EUGÉNIE J.A. 1952. The main features of Cambodian pronunciation. Bulletin of the School of Oriental and African Studies 14, 149–174.

HENDERSON, EUGÉNIE J.A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. Lingua 15, 400–434.

HENDERSON, EUGÉNIE J.A. 1967. Grammar and tone in Southeast Asian languages. Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig 16, 171–178.

HENRICH, NATHALIE / CHRISTOPHE D' ALESSANDRO / MICHÈLE CASTELLENGO / BORIS DOVAL. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. Journal of the Acoustical Society of America 115, 1321–1332.

HESS, S. 1992. Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan. Journal of Phonetics 20, 475–492.

HILDEBRANDT, KRISTINE. 2003. Manange tone: scenarios of retention and loss in two communities. Santa Barbara: University of California at Santa Barbara Ph.D.

HIRST, DANIEL / ALBERT DI CRISTO. 1998. A survey of intonation systems. Intonation Systems: A Survey of Twenty Languages. Cambridge: Cambridge University Press.

HOMBERT, JEAN-MARIE. 1978. Consonant types, vowel quality and tone. In: Victoria A. Fromkin (Ed.), Tone : a Linguistic Survey (pp. 77–111). New York: Academic Press.
HOMBERT, JEAN-MARIE. 1986. Word games: some implications for analysis of tone and other phonological constructs. In: John Ohala / J.J. Jaeger (Eds.), Experimental Phonology (pp. 175–186). Orlando: Academic Press.

HUANGFU QINGLIAN. 1994. Jīngxuǎn fǎ/hàn - hàn/fǎ cídiǎn (精选法汉-汉法词典) [A concise French/Chinese-Chinese/French dictionary]. Beijing: Shangwu/Larousse.

HYMAN, LARRY M. 1981. Tonal accent in Somali. Studies in African Linguistics 12, 169–203.

HYMAN, LARRY M. 2010. How to study a tone language, with exemplification from Oku (Grassfields Bantu, Cameroon). UC Berkeley Phonology Lab Annual Report, 179–209.

HYMAN, LARRY M. 2011a. Markedness, faithfulness, and the typology of two-height tone systems. UC Berkeley Phonology Lab Annual Report, 189–199.

HYMAN, LARRY M. 2011b. Do tones have features? In: Elizabeth Hume / John Goldsmith / W. Leo Wetzels (Eds.), Tones and Features (pp. 50–80). Berlin: De Gruyter Mouton.

HYMAN, LARRY M. / WILLIAM R. LEBEN. 2000. Suprasegmental processes. In: Geert Booij / Christian Lehmann / Joachim Mugdan (Eds.), Morphology: an international handbook on inflection and word-formation (pp. 587–594). Berlin: de Gruyter.

HYMAN, LARRY M. / KEMMONYE C. MONAKA. 2008. Tonal and non-tonal intonation in Shekgalagari. UC Berkeley Phonology Lab Annual Report, 269–288.

HYMAN, LARRY M. / KENNETH VANBIK. 2002. Tone and stem2 formation in Hakha Lai. Linguistics of the Tibeto-Burman Area 25, 113–121.

HYMAN, LARRY M. / KENNETH VANBIK. 2004. Directional rule application and output problems in Hakha Lai tone. Language and Linguistics, Taipei: Academia Sinica, Special Issue: Phonetics and Phonology 5, 821–861.

HYSLOP, GWENDOLYN. 2009. Kurtöp tone: a tonogenetic case study. Lingua 119, 827–845.

JACQUES, GUILLAUME. 2011. A panchronic study of aspirated fricatives, with new evidence from Pumi. Lingua 121, 1518–1538.

JAGGAR, PHILIP J. 2001. Hausa. (London Oriental and African Language Library 7). Amsterdam: John Benjamins.

JONES, ROBERT B. 1986. Pitch register languages. In: John McCoy / Timothy Light (Eds.), Contributions to Sino-Tibetan Studies (pp. 135–143). Leiden: E.J. Brill.

JUN, SUN-AH. 2005. Prosodic Typology. In: Sun-Ah Jun (Ed.), Prosodic typology: the phonology of intonation and phrasing (pp. 430–458). Oxford: Oxford University Press.

KARLSSON, ANASTASIA / DAVID HOUSE / JAN-OLOF SVANTESSON. 2012. Intonation adapts to lexical tone: the case of Kammu. Phonetica 69, 28–47.

KARTTUNEN, LAURI. 2006. The insufficiency of paper-and-pencil linguistics: the case of Finnish prosody. In: M. Butt / M. Dalrymple / T.H. King (Eds.), Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan (pp. 287–300). Stanford: CSLI Publications.

KINGSTON, JOHN. 2011. Tonogenesis. In: Marc van Oostendorp / Colin J. Ewen / Elizabeth Hume / Keren Rice (Eds.), The Blackwell companion to phonology (pp. 2304–2333). Oxford: Blackwell.

KIRBY, JAMES. 2010. Dialect experience in Vietnamese tone perception. Journal of the Acoustical Society of America 127, 3749–3757.

KOCHANSKI, GREG P. / SHIH CHILIN. 2003. Prosody Modelling with Soft Templates. Speech Communication 39, 311–352.

KOCHANSKI, GREG P., SHIH CHILIN / JING HONGYAN. 2003. Hierarchical structure and word strength prediction of Mandarin prosody. International Journal of Speech Technology 6, 33–43.

KUANG, JIANJING. 2013. The tonal space of contrastive five level tones. Phonetica 70, 1–23..

LADD, ROBERT. 1978. Stylized intonation. Language 54, 517–540.

LADD, ROBERT. 1993. In defense of a metrical theory of intonational downstep. In Harry van der Hulst / K. Snider (Eds.), The Phonology of Tone: The Representation of Tonal Register (pp. 109–132). Berlin and New York: Mouton/ De Gruyter.

LADD, ROBERT. 1996. Intonational Phonology. Cambridge: Cambridge University Press.

LADD, ROBERT. 2008. Review of Sun-Ah Jun (ed.) (2005) Prosodic typology: the phonology of intonation and phrasing. Phonology 25, 372–376.

LADD, ROBERT. 2014. Simultaneous structure in phonology. Oxford: Oxford University Press.

LANIRAN, YETUNDE O. / NICK CLEMENTS. 2003. Downstep and high raising: interacting factors in Yorùbá tone production. Journal of Phonetics 31, 203–250.

LEE, THOMAS. 1983. An acoustical study of the register distinction in Mon. UCLA Working Papers in Phonetics 57, 79–96.

LIN YUTANG. 1972. Dāngdài Hàn-Yīng cídiǎn (當代漢英詞典) [Chinese-English Dictionary of Modern Usage]. Hong Kong: Hong Kong Chinese University.

MADDIESON, IAN. 1978. The frequency of tones. Proc. Annual Meeting of the Berkeley Linguistics Society 4, 360–369.

MADDIESON, IAN. 2011. Tone. In: Matthew S. Dryer / Martin Haspelmath (Eds.), The World Atlas of Language Structures online. Leipzig: Max Planck Digital Library. http://wals.info/chapter/13.

MARTINET, ANDRÉ. 1957. Phonetics and linguistic evolution. In: Louise Kaiser (Ed.), Manual of phonetics (pp. 252 – 273). Amsterdam: North Holland.

MARTIN, PHILIPPE. 2009. Intonation du français. Paris: Armand Colin.

MATISOFF, JAMES A. 1999. Tibeto-Burman Tonology in an Areal Context. Proc. Symposium on Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics, Tokyo, Japan, 3–31.

MAZAUDON, MARTINE. 2005. On tone in Tamang and neighbouring languages: synchrony and diachrony. Proc. Symposium on Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics, Tokyo, Japan, 79–96.

MAZAUDON, MARTINE. 2012. Paths to tone in the Tamang branch of Tibeto-Burman (Nepal). In: Gunther de Vogelaer / Guido Seiler (Eds.), Dialects as a testing ground for theories of language change (pp. 139–177). Amsterdam/Philadelphia: John Benjamins.

MAZAUDON, MARTINE / ALEXIS MICHAUD. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a "missing link" in tonogenesis. Phonetica 65, 231–256.

MICHAEL, LEV. 2010. The interaction of tone and stress in the prosodic system of Iquito (Zaparoan). UC Berkeley Phonology Lab Annual Report, 57–79.

MICHAILOVSKY, BOYD / MARTINE MAZAUDON / ALEXIS MICHAUD / SÉVERINE GUILLAUME / ALEXANDRE FRANÇOIS / EVANGELIA ADAMOU. 2014. Documenting and researching endangered languages: the Pangloss Collection. Language Documentation and Conservation 8, 119–135.

MICHAUD, ALEXIS. 2004. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica 61, 119–146.

MICHAUD, ALEXIS. 2006. Tonal reassociation and rising tonal contours in Naxi. Linguistics of the Tibeto-Burman Area, 61–94.

MICHAUD, ALEXIS. 2008. Phonemic and tonal analysis of Yongning Na. Cahiers de linguistique - Asie Orientale 37, 159–196.

MICHAUD, ALEXIS. 2012. Monosyllabicization: patterns of evolution in Asian languages. In: Nicole Nau / Thomas Stolz / Cornelia Stroh (Eds.), Monosyllables: from phonology to typology (pp. 115–130). Berlin: Akademie Verlag.

MICHAUD, ALEXIS. 2013. The tone patterns of numeral-plus-classifier phrases in Yongning Na: a synchronic description and analysis. In: Nathan Hill / Tom Owen-Smith (Eds.), Transhimalayan Linguistics. Historical and Descriptive Linguistics of the Himalayan Area (pp. 275–311). Berlin: De Gruyter Mouton.

MICHAUD, ALEXIS / TUÂN VU-NGOC. 2004. Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, F0 Curve is Modified. Proc. 2nd International Conference of Speech Prosody, Nara, Japan, 745–748.

MILLER, AMANDA / JOHANNA BRUGMAN / BONNY SANDS / LEVI NAMASEB / MATS EXTER / CHRIS COLLINS. 2009. Differences in airstream and posterior place of articulation among N|uu clicks. Journal of the International Phonetic Association 39, 129–161.

MIXDORFF, HANSJÖRG / NGUYEN HUNG BACH / HIROYA FUJISAKI / MAI CHI LUONG. 2003. Quantitative Analysis and Synthesis of Syllabic Tones in Vietnamese. Proc. Eurospeech, Geneva, Switzerland, 177–180.

MORÉN, BRUCE / ELIZABETH ZSIGA. 2007. The lexical and post-lexical phonology of Thai tones. Natural Language and Linguistic Theory 24, 113–178.

MOREY, STEPHEN. 2005. Tonal change in the Tai languages of Northeast India. Linguistics of the Tibeto-Burman Area 28, 139–202.

MOREY, STEPHEN. 2008. The Tai languages of Assam. In: Anthony Diller / Jerold A. Edmondson / Luo Yongxian (Eds.), The Tai-Kadai languages (pp. 207–253). London: Routledge.

MOREY, STEPHEN. 2014. Studying tones in North East India: Tai, Singpho and Tangsa. Language Documentation and Conservation 8, 637—671.

NEWMAN, PAUL. 1995. Hausa Tonology: Complexities in an "Easy" Tone Language. In: J. Goldsmith (Ed.), Handbook of Phonological Theory (pp. 762–781). Oxford: Blackwell.

NGUYEN, THI-LAN / ALEXIS MICHAUD / DO-DAT TRAN / DANG-KHOA MAC. 2013. The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particles. Proc. Interspeech 2013, Lyon, France, 1-5.

NIEBUHR, OLIVER. 2013. On the acoustic complexity of intonation. In: E.-L. Asu / P. Lippus (Eds.), Nordic Prosody XI (pp. 25–38). Frankfurt: Peter Lang.

NOLAN, FRANCIS. 1999. The devil is in the detail. Proc. 14th International Congress of the Phonetic Sciences, San Francisco, USA, 1–8.

ODDEN, DAVID. 1995. Tone: African languages. In: John Goldsmith (Ed.), Handbook of Phonological Theory (pp. 444–475). Oxford: Blackwell.

OHALA, JOHN. 1983. Cross-language use of pitch: an ethological view. Phonetica 40, 1–18.

OKELL, JOHN A. 1969. A Reference Grammar to Colloquial Burmese. London: School of African and Oriental Studies.

OTTERLOO, KAREN VAN. 2011. The Kifuliiru language, volume 1: phonology, tone, and morphological derivation. Dallas: SIL International.

PETRONE, CATERINA / OLIVER NIEBUHR. 2013. On the intonation of German intonation questions: the role of the prenuclear region. Language and Speech 57, 108–146.

PIERREHUMBERT, JANET / S. STEELE. 1989. Categories of tonal alignment in English. Phonetica 46, 181–196.

PIKE, KENNETH L. 1948. Tone Languages. A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion. Ann Arbor: University of Michigan Press.

PITTAYAPORN, PITTAYAWAT. 2007. Directionality of tone change. Proc. 16th International Conference of the Phonetic Sciences, Saarbrücken, Germany, 1421–1424.

PROM-ON, SANTITHAM. 2008. Pitch target analysis of Thai tones using quantitative target approximation model and unsupervised clustering. Proc. Interspeech 2008, Brisbane, Australia, 1116–1119.

PULLEYBLANK, EDWIN G. 1978. The nature of the Middle Chinese tones and their development to Early Mandarin. Journal of Chinese Linguistics 6, 173–203.

RATLIFF, MARTHA. 1992a. Tone language type change in Africa and Asia: !Xũ, Gokana, and Mpi. Diachronica 9, 239–257.

RATLIFF, MARTHA. 1992b. Meaningful tone: A study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong. De Kalb, Illinois: Northern Illinois University Center for Southeast Asian Studies.

RATLIFF, MARTHA. 2010. Hmong-Mien language history. Canberra: Pacific Linguistics.

RHODES, ALEXANDRE DE. 1651. Dictionarium Annamiticum Lusitanum et Latinum. Rome.

RIALLAND, ANNIE. 1997. Le parcours du “downstep”, ou l’évolution d’une notion. Bulletin de la Société de Linguistique de Paris XCII, 207–243.

RIALLAND, ANNIE. 1998. Systèmes prosodiques africains: une source d'inspiration majeure pour les théories phonologiques multilinéaires. Faits de langues 11-12, 407–428.

RIALLAND, ANNIE. 2001. Anticipatory Raising in Downstep Realization: Evidence for Preplanning in Tone Production. Proc. Symposium on Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics, Tokyo, Japan, 301–322.

RIVIERRE, JEAN-CLAUDE. 1993. Tonogenesis in New Caledonia. In: Jerold A. Edmondson / Kenneth J. Gregerson (Eds.), Tonality in Austronesian Languages (pp. 155–173). Honolulu: University of Hawaii Press.

RIVIERRE, JEAN-CLAUDE. 2001. Tonogenesis and Evolution of the Tonal Systems in New Caledonia, the Example of Cèmuhî. Proc. Symposium on Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics, Tokyo, Japan, 23–42.

ROSE, PHILIP. 1982. Acoustic characteristics of the Shanghai-Zhenhai syllable-types. In: David Bradley (Ed.), Papers in Southeast Asian Linguistics n°8: Tonation (pp. 1–53). Canberra: Australian National University Press.

ROSE, PHILIP. 1989a. Phonetics and phonology of Yang tone phonation types in Zhenhai. Cahiers de linguistique - Asie Orientale 18, 229–245.

ROSE, PHILIP. 1989b. On the non-equivalence of fundamental frequency and linguistic tone. In: David Bradley / Eugénie J.A. Henderson / Martine Mazaudon (Eds.), Prosodic Analysis and Asian Linguistics: to honour R.K. Sprigg (pp. 55–82). Canberra: Pacific Linguistics..

ROSE, PHILIP. 1990. Acoustics and phonology of complex tone sandhi: An analysis of disyllabic lexical tone sandhi in the Zhenhai variety of Wu Chinese. Phonetica 47, 1–35.

ROSSI, MARIO. 1999. L'intonation, le système du français: description et modélisation. Gap/Paris: Ophrys.

ROSS, MALCOM. 1993. Tonogenesis in the North Huon Gulf chain. In Jerold A. Edmondson and Kenneth J. Gregerson (eds.), Tonality in Austronesian Languages, 133–153. (Oceanic Linguistics Special Publications 24). Honolulu: University of Hawaii Press.

SILVERMAN, K. E. A. / MARY BECKMAN / J. PITRELLI / M. OSTENDORF / COLIN W. WIGHTMAN / P. PRICE / JANET PIERREHUMBERT / J. HIRSCHBERG. 1992. ToBI: A Standard for Labeling English Prosody. Proc. International Conference on Spoken Language Processing, Banff, Canada, 867–870..

SUN, JACKSON T.-S. 1997. The Typology of Tone in Tibetan. Chinese Languages and Linguistics IV: Typological studies of languages in China. Taipei, Taiwan: Symposium Series of the Institute of History and Philology-Academia Sinica.

SVANTESSON, JAN-OLOF / DAVID HOUSE. 2006. Tone production, tone perception and Kammu tonogenesis. Phonology 23, 309–333.

THONGKUM, THERAPAN L. 1987. Another look at the register distinction in Mon. UCLA Working Papers in Phonetics 67, 29–48.

THONGKUM, THERAPAN L. 1988. Phonation types in Mon-Khmer languages. In: Osamu Fujimura (Ed.), Voice production: Mechanisms and functions (pp. 319–333). New York: Raven Press.

THONGKUM, THERAPAN L. 1991. An instrumental study of Chong register. In: Richard J. Davidson (Ed.), Austroasiatic languages: essays in honour of H.L. Shorto (pp. 141–160). London: School of Oriental and African Studies, University of London.

VAISSIÈRE, JACQUELINE. 2002. Cross-linguistic prosodic transcription: French vs. English. In: N.B. Volskaya / N.D. Svetozarova / P.A. Skrelin (Eds.), Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko (pp. 147–164). St Petersburg: St Petersburg State University Press.

VAISSIÈRE, JACQUELINE. 2004. The Perception of Intonation. In: David B. Pisoni / Robert E. Remez (Eds.), Handbook of Speech Perception (pp. 236–263). Oxford: Blackwell.

VAISSIÈRE, JACQUELINE / ALEXIS MICHAUD. 2006. Prosodic constituents in French: a data-driven approach. In: Ivan Fónagy / Yuji Kawaguchi / Tsunekazu Moriguchi (Eds.), Prosody and syntax: Cross-linguistic perspectives (pp. 47–64). Amsterdam: John Benjamins.

WANG, WILLIAM. 1967. Phonological Features of Tones. International Journal of American Linguistics 33, 93–105.

WAN, I-PING / JERI JAEGER. 1998. Speech errors and the representation of tone in Mandarin Chinese. Phonology 15, 417–461.

WATKINS, JUSTIN. 2002. The Phonetics of Wa. (Pacific Linguistics 531). Canberra: Australian National University.

WAYLAND, RATREE / ALLARD JONGMAN. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. Journal of Phonetics 31, 181–201.

WEDEKIND, KLAUS. 1983. A six-tone language in Ethiopia: Tonal analysis of Benč non (Gimira). Journal of Ethiopian Studies 16, 129–156.

WEDEKIND, KLAUS. 1985. Why Bench' (Ethiopia) has five level tones today. In: Ursula Pieper / Gerhard Stickel (Eds.), Studia linguistica diachronica et synchronica (pp. 881–901). Berlin: Mouton.

WELMERS, WILLIAM E. 1952. Notes on the structure of Bariba. Language 28, 82–103.

WHEATLEY, JULIAN K. 2003. Burmese. In: Graham Thurgood / Randy LaPolla (Eds.), The Sino-Tibetan languages (pp. 195–207). London: Routledge.

XU, YI. 1999. Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics 27, 55–106.

XU, YI / EMILY Q. WANG. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication 33, 319–337.

YIP, MOIRA. 1980. The Tonal Phonology of Chinese. Cambridge, Massachusetts: Indiana University Linguistics Club. Published 1990 by Garland Publishing, New York.

YIP, MOIRA. 1989. Contour tones. Phonology 6, 149–174.

YIP, MOIRA. 2002. Tone. Cambridge: Cambridge University Press.

YIP, MOIRA. 2007. Tone. In: Paul De Lacy (Ed.), The Cambridge Handbook of Phonology (pp. 229–252). Cambridge: Cambridge University Press.

YU, ALAN C.L. 2003. Some methodological issues in phonetic typology research: Cantonese contour tone revisited. Proc. 29th Annual Meeting of the Berkeley Linguistics Society, 623–634.

YU, KRISTINE / HIU WAI LAM. 2014. The role of creaky voice in Cantonese tonal perception. Journal of the Acoustical Society of America 136, 1320–1333.

ZERBIAN, SABINE. 2010. Developments in the study of intonational typology. Language and Linguistics Compass 3, 1–16.

Kieler Arbeiten zur Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

3 ExperimentMFC – Erstellung und Auswertung eines Perzeptionsexperiments in Praat

Julia Beck
Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
www.isfas.uni-kiel.de/de/linguistik

Diese Arbeit beschäftigt sich mit der Erstellung, Durchführung und Auswertung von Perzeptionsexperimenten mit Hilfe von *ExperimentMFC*, einem Objekttyp der der Software *Praat*. *ExperimentMFC* ermöglicht es dem Versuchsleiter Identifikations- und Diskriminationsexperimente zu erstellen, mit Versuchspersonen am Computer durchzuführen und die Ergebnisse anschließend auszuwerten. Im Fokus der Arbeit steht eine detaillierte Anleitung zur Erstellung einer *ExperimentMFC*-Steuerungsdatei, die die Inhalte und den Ablauf eines Experiments spezifiziert. Die Anleitung wird anhand eines zuvor durchgeführten Perzeptionsexperiments zur sprachübergreifenden Wahrnehmung des tschechischen Wortes *jasně* durch explizite Beispiele und zusätzliche Hinweise zur Behandlung von Fehlermeldungen unterstützt.

1. Motivation

Die hauptsächlich von Paul Boersma und David Weenik entwickelte Software *Praat* bietet neben vielen anderen Features auch die Möglichkeit, mittels *ExperimentMFC* eigene Perzeptionsexperimente zu erstellen, durchzuführen und die Ergebnisse anschließend zu verarbeiten. *ExperimentMFC* ist ein Objekttyp in *Praat*, der für die Entwicklung und Durchführung von Identifikations- und Diskriminationsexperimenten eingesetzt wird.

Im Rahmen des Hauptseminars ‚Kontrastive Phonetik: Perzeption' wurde die Software für ein Perzeptionsexperiment eingesetzt. Bei diesem Experiment wurde die sprachübergreifende Wahrnehmung des tschechischen Wortes *jasně* („klar", „deutlich", „alles klar!") untersucht. Dafür wurden tschechischen und deutschen Muttersprachlern *jasně*-Stimuli vorgespielt. Die Versuchspersonen sollten pro Stimulus entscheiden, ob der jeweilige Sprecher z.B. gleichgültig klingt oder nicht. Außerdem sollten sie ihre Antwort

jeweils in Bezug auf ihre Sicherheit bewerten. Auf diese Weise sollte geprüft werden, ob die deutschen Versuchspersonen dieselben funktionalen Assoziationen mit den Stimuli verbinden wie die tschechischen Muttersprachler. Und ob außerdem eine Interaktion zwischen der angebotenen und der assoziierten, kommunikativen Funktion besteht. Dieser Interaktionseffekt hat sich im Experiment als signifikant herausgestellt.[1]

Für die Durchführung von Perzeptionsexperimenten mit *ExperimentMFC* werden Stimuli, *Praat* und eine speziell aufgebaute, unter *Praat* ausführbare Experimentdatei benötigt. In dieser Experimentdatei werden verschiedene Parameter spezifiziert, die den Ablauf des Experiments steuern. Im Folgenden wird diese Datei daher als Steuerungsdatei bezeichnet. Die Steuerungsdatei ist kein Skript, sondern eine Textdatei und kann mit jedem beliebigen Texteditor in kurzer Zeit erstellt werden. Somit sind nur wenige Hilfsmittel für die Durchführung eines Experiments erforderlich. Da *Praat* zudem auf den gängigen Betriebssystemen Windows, MAC, Unix und Linux stabil läuft, sind die Experimente mit *ExperimentMFC* außerdem sehr portabel. Die benötigten Dateien und Programme können auf CD/DVD gebrannt oder auf einen USB-Stick geschrieben werden. So können die Experimente auf andere Medien übertragen und auf diesen ausgeführt werden, wodurch Experimente weniger ortsgebunden sind. Die Experimente können auch in verschiedenen Sprachen durchgeführt werden, da alle Folientexte und Button-Beschriftungen beliebig geändert werden können. Ein weiterer Vorteil von *ExperimentMFC* ist die Tatsache, dass es als Teil von *Praat* kostenlos unter der GNU General Public License verfügbar ist. Zudem wird *Praat* ständig gepflegt und weiterentwickelt.

Es muss jedoch auch erwähnt werden, dass *ExperimentMFC* der eigenen Kreativität viele Grenzen setzt. Die Auswahlmöglichkeiten für die Foliengestaltung und die Steuerung des Experiments sind eingeschränkt, was zu Problemen bei der Erstellung komplexerer Experimente führen kann. Der Einfluss auf den Kontrollfluss des Programms ist sehr gering, da der Aufbau der Steuerungsdatei sehr strikt[2] ist. Deshalb können z.B. Experimente mit bedingten Abfolgen, bei denen der nächste Stimulus in Abhängigkeit von der vorherigen Antwort der Versuchsperson ausgewählt wird, nur über Umwege und größeren Programmieraufwand erreicht werden (vgl. Mayer (2013: 241-262) für den Umstieg auf das *Praat Demo Window*). Ebenso sind Trainingsszenarien, bei denen Lernalgorithmen aufgrund von Erfahrung gezielt passende Stimuli für eine bestimmte Versuchsperson auswählen, mit *ExperimentMFC* noch nicht umsetzbar. Falls man ein komplexeres Experiment plant, sollte man also zunächst überdenken, ob nicht andere Programme wie *Presentation*[3] oder *E-prime*[4] geeigneter wären.

Das während des Experiments angezeigte Fenster, die sogenannte grafische Oberfläche, ist in *ExperimentMFC* kaum veränderbar und sehr simpel gestaltet. Es ist nur möglich, begrenzten Einfluss auf den Schriftsatz zu nehmen und Positionen gewisser Objekte

[1] Detaillierte Ausführungen zum *jasně*-Experiment finden sich in: Volín, J., L. Weingartová und O. Niebuhr (2014). Between recognition and resignation - The prosodic forms and communicative functions of the Czech confirmation tag "jasne". Proc. 7th International Conference of Speech Prosody, Dublin, Ireland.

[2] Diese Striktheit ist die größte Fehlerquelle bei der Erstellung von Experimenten.

[3] https://www.neurobs.com/

[4] http://www.pstnet.com/eprime.cfm

zu ändern. Die Fenster sind zudem nicht beliebig dynamisch in der Größe veränderbar, weshalb es auf verschiedenen Rechnern zu Kompabilitätsproblemen kommen kann. Daher ist es wichtig, das jeweilige Experiment vor der eigentlichen Ausführung zu testen, damit während des Experiments trotz verschiedener Bildschirmauflösungen alle Textabschnitte sichtbar sind.

Für die Erstellung dieser Anwendungsbeschreibung von *ExperimentMFC* wurde die aktuelle *Praat* Version *5.3.64* vom 12.02.2014 verwendet. Die Experimente wurden auf den Betriebssystemen Windows 8.1 und dem linuxbasierten Ubuntu 13.10 getestet. Im Folgenden sollen zunächst kurz die Experimenttypen vorgestellt werden, die mit *ExperimentMFC* durchgeführt werden können.

2. Experimenttypen in *ExperimentMFC*

In *ExperimentMFC* steht das *MFC* für *Multiple Forced Choice*. Das heißt, dass zwar mehrere Antwortmöglichkeiten zur Verfügung stehen, die Versuchsperson sich aber auf jeden Fall für eine Antwort entscheiden muss. Für die Analyse ist es wichtig, dass z.B. durch das Einfügen von Störsignalen wie Rauschen vermieden wird, dass die Versuchspersonen zu sehr durch vorangegangene Stimuli beeinflusst werden. Variation und eine Erschwerung der Perzeption kann auch durch eine Vereinfachung oder Qualitätsverschlechterung des Stimulus sowie durch Änderung der Lautstärke erreicht werden (vgl. McGuire (2010: 12)).

Neben anderen Experimenttypen wird vor allem zwischen Identifikations- und Diskriminationsexperimenten unterschieden. In *ExperimentMFC* sind beide Experimenttypen im Allgemeinen umsetzbar. In Identifikationsexperimenten wird der Versuchsperson in der Regel ein Stimulus angeboten, und es ist ihre Aufgabe, aus einer Menge von explizit vorgegebenen Antworten die passende auszuwählen. Bei Identifikationsexperimenten wird oft die kategoriale Wahrnehmung untersucht, wobei z.B. mehrdeutige Stimuli oder künstlich erzeugte, kontinuierliche Übergänge zwischen zwei Lauten einer bestimmten Kategorie zugeordnet werden sollen. Demgegenüber geht es bei Diskriminationsexperimenten um die Fähigkeit einer Versuchsperson, zwischen Stimuli zu unterscheiden.

Oft enthalten die Experimente auch Bewertungsaufgaben, bei denen Versuchspersonen die Aufgabe haben, das Gehörte subjektiv auf einer Skala z.B. als eher laut oder eher leise zu klassifizieren. Im *jasně*-Experiment sollten die Versuchspersonen hingegen nicht die Güte des Stimulus, sondern die Sicherheit der eigenen Antwort bewerten. Es ist in *ExperimentMFC* allerdings nicht möglich, mehr als zwei Bewertungsskalen anzuzeigen und pro Stimulus mehr als zwei Antworten zu geben. Für detaillierte Beschreibungen folgender und weiterer Experimenttypen siehe McGuire (2010).

2.1 Identifikationsexperiment

Bei einem Identifikationsexperiment hört die Versuchsperson in der Regel nur einen Stimulus und muss aus einer Menge von Antwortmöglichkeiten eine zu dem Gehörten passende Antwort auswählen.

Die einfachste Form ist die Ja-Nein-Identifikation. Ein klassisches Beispiel dafür sind Hörtests, bei denen ein Ton mit einer bestimmten Frequenz und Lautstärke einer Versuchsperson vorgespielt wird. Die Versuchsperson hat dann die Aufgabe mit (JA) oder (NEIN) zu beantworten, ob sie den Ton gehört hat oder nicht. Dabei sind die einfache Erstellung und Auswertung des Experiments von Vorteil, da es nur zwei Antwortmöglichkeiten gibt. Außerdem sind Fehler durch die Versuchsperson aufgrund der einfach erklärbaren, eindeutigen Aufgabenstellung eher unwahrscheinlich und die Messung der Reaktionszeit entsprechend zuverlässig.

Diese Aufgabenstellung war auch im *jasně*-Experiment gegeben, in dem die Versuchsperson aufgefordert wurde, anzugeben, ob der Sprecher im vorgegebenen Stimulus z.B. überrascht klingt (JA) oder nicht (NEIN). Dafür hat eine tschechische Muttersprachlerin zuvor jeweils vier Stimuli einer von acht unterschiedlichen Funktionen zugeordnet (*neutral, impatient, indifferent, eager, realizing, surprised, reassuring, resignated*). Die Stimuli stammten aus zuvor aufgenommenen Dialogen. Für das Experiment mit den deutschen Hörern wurden die einzelnen Funktionen entsprechend ins Deutsche übersetzt (*neutral, abwürgend, gleichgültig, eifrig, erleuchtet, überrascht, beruhigend, resigniert*). Bei der Übersetzung musste so gut wie möglich, darauf geachtet werden, dass die Semantik der englischen Begriffe nicht verloren geht bzw. verändert wird.

Die Stimuli wurden nacheinander tschechischen und deutschen Muttersprachlern vorgespielt. Pro Stimulus bekamen die Versuchspersonen auf dem Computer eine Experimentfolie angezeigt, auf denen der Satz ‚Der Sprecher klingt... ' und eine der acht Kategorien abgebildet waren sowie die Antwortmöglichkeiten (JA) / (NEIN) und eine Bewertungsskala (vgl. Abschnitt 3.2). Die 32 vorhandenen Stimuli, vier pro funktionaler Kategorie, wurden dabei gleichmäßig in zweimal 16 Stimuli aufgeteilt, sodass es zwei Teilexperimente gab. Jeder Stimulus wurde mit jeder Kategorie kombiniert, sodass es 16 x 8 = 128 verschiedene Kombinationen gab. Diese wurden zudem noch je dreimal in zufälliger Reihenfolge wiederholt, sodass die Versuchspersonen insgesamt pro Teilexperiment 384 Mal einen Stimulus einer funktionalen Kategorie zuordnen mussten. Außerdem sollten sie ihre Antwort jeweils in Bezug auf ihre Sicherheit bewerten.

In anderen Experimenten können natürlich auch andere Antworten außer (JA) und (NEIN) zur Verfügung stehen. Diese Art von Identifikation wird mit *Labeling* bezeichnet. Bei solchen Experiment wird der Versuchsperson immer nur je ein Stimulus präsentiert, und die Versuchsperson muss aus einer fest vorgegebenen Menge von Antworten die richtige oder passende wählen. Falls es nur zwei Antwortmöglichkeiten gibt, z.B. (S) vs. (SCH), erfolgt die Analyse analog zur Auswertung der Ja-Nein-Identifikation. In *ExperimentMFC* ist es jedoch nicht möglich, dass die Versuchsperson selbst eine beliebige Antwort notiert, indem sie ihre Antwort z.B. per Tastatur in ein Textfeld eingibt. Identifikationsexperimente erfordern daher immer die Erstellung einer Menge an *Labels*, aus denen die Versuchsperson wählen kann. Hierbei gilt, je mehr Auswahlmöglichkeiten es gibt, desto komplexer wird auch die Analyse. Allerdings ist eine rein quantitative Auswertung der Ergebnisse der einzelnen Kategorien laut McGuire (2010: 6) meist adäquat. Es muss auch bedacht werden, dass die vorgegebenen Antworten bzw. Kategorien die Versuchsperson auch beeinflussen oder in eine bestimmte Richtung lenken können. Deshalb war im *jasně*-Experiment die Korrektheit der Übersetzung auch

besonders wichtig.

Es gibt allerdings in *ExperimentMFC* auch die Möglichkeit, dass die Antworten keine Kategorien oder Labels wie (LAUT) oder (LEISE) sind, sondern Stimuli. Der passende Stimulus kann dort aus verschiedenen Stimuli ausgewählt werden, die die Versuchsperson sich jeweils einzeln anhören kann (vgl. 3.3.1).

2.2 Diskriminationsexperiment

Bei Diskriminationsexperimenten werden in der Regel mehrere Stimuli pro Urteil abgespielt. Die Aufgabe ist dabei meist ein Vergleich zwischen den Stimuli, ob ein Stimulus beispielsweise einem anderen präsentierten Stimulus ähnelt oder gar entspricht. Typisch sind Gleich-Verschieden-Experimente, die entsprechend der Anzahl der vorliegenden Stimuli bezeichnet werden (z.B. AX bzw. 2IAX, ABX, 4IAX, usw.).

Bei AX- bzw. 2IAX-Experimenten hört die Versuchsperson pro Teilaufgabe zwei Stimuli, A und X, die nacheinander abgespielt werden, wobei A und X durch eine bestimmte Zeitspanne (ISI: Interstimulus Intervall) voneinander abgegrenzt werden (vgl. McGuire (2010: 3)). Die Versuchsperson soll dabei entscheiden, ob die Stimuli gleich oder verschieden sind. Die verschiedenen Kombinationen, die dabei zustande kommen können sind: (AX), (XA), (AA) und (XX). Die Reihenfolge der Paare spielt dabei eine nicht unerhebliche Rolle und darf möglichst nicht außer Acht gelassen werden. Da in *ExperimentMFC* auch die Namen und Reihenfolge der Stimuli in den Experimentergebnissen dokumentiert werden, kann dieser Aspekt dort berücksichtigt werden. Weitere Messergebnisse von Diskriminationsexperimenten sind die Trefferquote, also ob die Versuchsperson richtig lag oder nicht, und Reaktionszeiten.

Aufgrund der möglichen Änderung der Reihenfolge kann die Analyse allerdings komplex werden. Vorteil ist hier, wie beim Identifikationsexperiment, die Einfachheit der Aufgabenstellung und die damit verbundene Zuverlässigkeit der Reaktionszeiten. Der Unterschied zweier Stimuli muss der Versuchsperson zudem nicht erklärt werden, wenn die Entscheidung auch ohne dieses Wissen erfolgen kann. Ein Nachteil ist laut McGuire (2010: 3), dass Paare mit gleichen Stimuli für die Analyse oft uninteressant sind und somit knapp die Hälfte der Ergebnisse wegfällt. Die Anzahl dieser Paare kann bei umfangreichen Diskriminationen reduziert werden. Wenn die Diskrimination jedoch einfach ist, geht dies nicht, da die Versuchsperson sonst eventuell aufgrund der Erwartungshaltung, dass die Anzahl der gleichen und verschiedenen Paarungen gleich ist, seine Strategie ändert. Schwierigere Diskriminationen bergen allerdings das Problem, dass Versuchspersonen dazu tendieren, die Stimuli als gleich wahrzunehmen (vgl. ebenda).

Eine Variante von AX-Experimenten sind die Speeded-AX-Experimente. Dabei sind die ISI besonders kurz[5] und die Versuchsperson steht zudem unter Zeitdruck. Die Reaktionszeiten sind dadurch sehr konstant und der Zeitaufwand für die Durchführung der Experimente gering. Die Schwierigkeit und der Stressfaktor der Aufgabe können jedoch zu Antwortverlusten führen. Außerdem ist die Aufgabe sehr von den Fähigkeiten der Versuchsperson abhängig, weshalb meist eine Bewertungsskala in Bezug auf die Sicherheit

[5] Unter 500 ms (meist 100 ms).

der gegebenen Antwort eingesetzt wird.

Eine weitere Variante mit zwei Stimuli sind 2AFC-Experimente (*Two Alternative Forced Choice*), bei denen die Reihenfolge der Stimuli erkannt werden soll. McGuire betont, dass die Versuchsperson dabei wisse, dass beide Reihenfolgen möglich sind. Trotzdem kann dies zu einer Erwartungshaltung bei der Versuchsperson führen. Zudem sind Instruktionen eventuell schwieriger zu verstehen, da sie verdeutlichen müssen, inwiefern Reihenfolge in Bezug auf die Stimuli definiert ist.

In ABX- bzw. AXB- bzw. XAB-Experimenten werden der Versuchsperson drei Stimuli A, B und X präsentiert. Es muss dann die Entscheidung getroffen werden, ob Stimulus A oder B Stimulus X ähnlicher oder gleich Stimulus X ist. Dabei sind zwei Intervalle zwischen den Stimuli nötig, die jedoch laut McGuire (2010: 4) meist die gleiche Länge haben. Ein zusätzlicher Vorteil im Vergleich zu AX-Experimenten sei, dass die Versuchsperson hierbei wisse, dass entweder A oder B die richtige Antwort ist. Problematisch ist die Tendenz, den Stimulus zu wählen, der aufgrund der Reihenfolge von A, B und X direkt vor dem X gehört wurde. Dieser Effekt kann durch eine wechselnde, zufällige Reihenfolge von A und B vermieden werden. Die eingefügten Intervalle müssen bei der Reaktionszeitmessung berücksichtigt werden.

In 4IAX-Experimenten (*Interval Forced Choice*) werden vier Stimuli, entweder (ABAA), (AABA), (BABB) oder (BBAB), präsentiert. Die ISI zwischen den Stimuli sind meist gleich. Die Versuchsperson entscheidet nun, ob der zweite oder der dritte Stimulus sich von den anderen Stimuli unterscheidet bzw. ob die Stimuli der jeweiligen Paare gleich sind. Wie bei 2AFC-Experimenten sind die Optionen (gleich, verschieden) beide pro Teilaufgabe präsent und erscheinen daher für die Versuchsperson gleich wahrscheinlich, sodass es weniger wahrscheinlich zu Verzerrungen der Messwerte kommt als bei AX-Experimenten. Da der Zeitpunkt der Entscheidung für ein Szenario stark variieren kann, sind Reaktionszeiten hier wie bei ABX-Experimenten schwierig zu bestimmen (vgl. McGuire (2010: 5 f.)).

3. Erstellung eines Experiments

3.1 Erstellung von Stimuli mittels Praat

Um ein Perzeptionsexperiment mit Hilfe von *ExperimentMFC* zu erstellen, müssen passende Stimuli in Dateiform vorhanden sein oder z.B. in *Praat* erzeugt werden. Vorhandene *.wav*-Dateien können in Praat als Sound-Objekte eingelesen und verwendet werden. Stimuli können aber auch entweder mit dem *Praat-SoundRecorder* als Mono- oder Stereosignal aufgenommen (New › Record mono Sound... bzw. New › Record stereo Sound...) oder mit einer mathematischen Formel als künstliches Signal erzeugt werden (z.B. New › Sound › Create Sound from formula...). Bei der Aufnahme mit dem *SoundRecorder* ist es wichtig, welche Abtastrate für die Digitalisierung gewählt wird. Falls man verschiedene Stimuli für das Experiment verwenden möchte, ist es wichtig, dass alle Stimuli über dieselbe Abtastrate verfügen, da das Experiment sonst nicht mittels *ExperimentMFC* ausgeführt werden kann (s. Fehlermeldung im Anhang). Dies sollte auch bei der Erzeugung von künstlichen Signalen bedacht werden. Notfalls

bietet jedoch der sogenannte Modify-Befehl in *Praat* die Möglichkeit die Frequenz nachträglich anzupassen[6].

Auch für die Erstellung von Störsignalen, die vor oder nach dem eigentlichen Stimulus abgespielt werden sollen, eignet sich *Praat*. Wenn man z.B. den Kammerton a als Störsignal erzeugen möchte, muss man nur zum passenden Dialog navigieren (New 〉 Sound 〉 Create Sound as pure tone...) und bei (Tone frequency) 440.0 Hz eingeben. Für ein Rauschen wählt man entsprechend (New 〉 Sound 〉 Create Sound from formula...). In beiden Dialogen können Abtastrate und Dauer frei bestimmt werden. Zu weiteren wichtigen Aspekten bei der Bearbeitung natürlicher und der Erstellung synthetischer Stimuli siehe Mayer (2013: 28-34).

3.2 Aufbau der angezeigten, grafischen Oberfläche

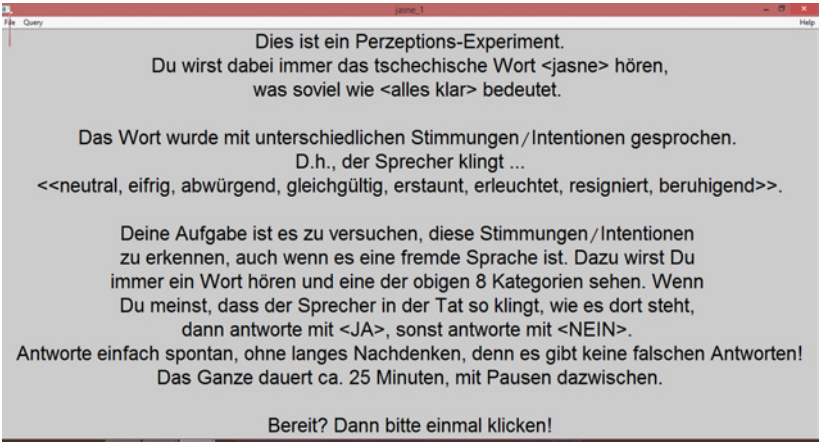


Abbildung 1. Titelfolie des jasně-Experiments.

Um den Aufbau der zusätzlich benötigten Steuerungsdatei nachvollziehen zu können, wird nun zunächst der Aufbau der während des Experiments angezeigten grafischen Oberfläche beschrieben. Während eines Experiments mit *ExperimentMFC* wird der Versuchsperson zuerst eine Titelfolie angezeigt, deren Text beliebig geändert werden kann. Nachfolgend erscheinen je nach Spezifikation der Steuerungsdatei eine gewisse Anzahl an Antwortfolien. Das Ende des Experiments wird durch eine entsprechende Endfolie angezeigt, die wiederum einen frei wählbaren Text enhält. Am Beispiel des *jasně*-Experiments können die Titel- und Endfolie wie in Abbildung 1 und 2 dargestellt aussehen.

[6] Öffne vorhandene Sounddatei mit (Open 〉 Read from file...), klicke in dem für das neue Soundobjekt rechts neu angebotenen Menü auf (Modify - 〉 Override sampling frequency...) und ändere die Frequenz im erscheinenden Dialogfenster.

Abbildung 2. Endfolie des jasně-Experiments.

Die Antwortfolien können je nach Experimenttyp verschieden gestaltet werden. In Abbildung 3 ist beispielhaft eine der 128 möglichen verschiedenen Antwortfolien des *jasně*-Experiments dargestellt.

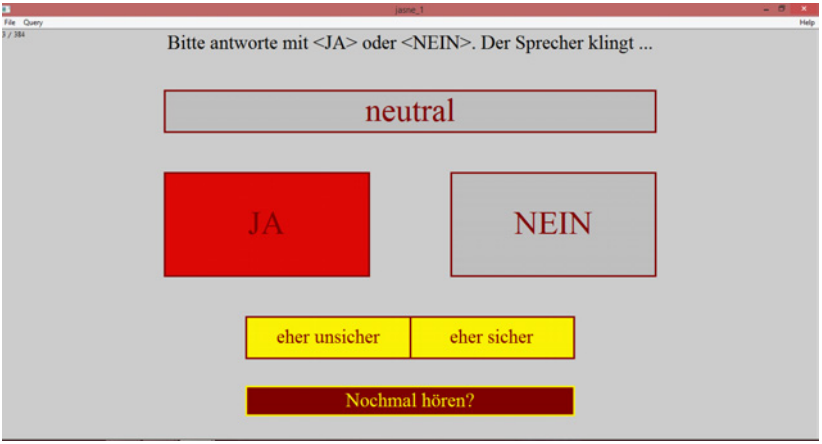


Abbildung 3. Eine Antwortfolie des jasně-Experiments.

Im oberen Bereich kann ein frei wählbarer Text mit Instruktionen angezeigt werden. Die Position des Textes ist nicht beliebig. Es folgen Antwortmöglichkeiten in Form von rechteckigen Feldern, die entweder grau, gelb oder rot eingefärbt sind. Ausgegraute Antwortfelder können nicht angeklickt werden. So kann das Klicken auf Buttons oder Felder zu unerwünschten Zeitpunkten vermieden werden. Im *jasně*-Experiment war dies z.B. bei

der Anzeige der jeweiligen kommunikativen Funktion sinnvoll. Hingegen können gelbe Felder angeklickt werden. Gelbe Felder werden rot, wenn sie angeklickt wurden, um der Versuchsperson anzuzeigen, welche Antwortmöglichkeit aktuell gewählt wurde. Pro Antwortfolie muss mindestens eine Antwort, im Fall einer Bewertungsskala zwei Antworten, gegeben werden. Dies spiegelt den *forced choice* in *ExperimentMFC* wider, da auf jeden Fall eine Antwort gegeben werden muss, damit es weitergeht. Die Beschriftung der Antwortfelder ist pro Antwortfolie änderbar (vgl. 3.3.3), d.h. dass stimulusabhängige Beschriftungen möglich sind. Außerdem können die Antwortmöglichkeiten in einen beliebigen Satz eingebaut werden, wobei es üblich ist, dass der Stimulus nicht das letzte Wort ist.

Wie in Abbildung 4 dargestellt, ist es in den neueren *Praat*-Versionen unter Windows und Mac auch möglich, Bilder als Antwortmöglichkeiten anzuzeigen.

Abbildung 4. Antwortfolie mit Bildern.

Dabei muss aber bedacht werden, dass Bilder die Versuchsperson beeinflussen können. Videos können in ExperimentMFC nicht eingefügt werden, sodass aufwändigere audiovisuelle Experimente wie für den McGurkEffekt (vgl. McGurk and MacDonald (1976)) nicht möglich sind. Weiterhin ist es möglich, dass Felder Sounds und nicht Beschriftungen als Antworten enthalten und die Versuchsperson diese Sounds abspielen und den richtigen auswählen kann.

Neben den möglichen Antwortfeldern kann den Antwortfolien auch je eine Bewertungsskala für die Bewertung der Antwort oder eine Bewertung der Stimuli durch die Versuchsperson hinzugefügt werden. Die Positionierung der Antwortfelder und Skalen erfolgt aufgrund eines Koordinatensystems und wird in Abschnitt 3.3.3 genauer erklärt. Antwortfolien können während des Abspielens des Stimulus als graue Fläche ohne Inhalt dargestellt werden, sodass die Antwortmöglichkeiten erst nach dem Hören sichtbar sind. Dies kann erwünscht sein, falls die Antworten die Versuchsperson ansonsten

in ihrer Entscheidung beeinflussen könnten. Die Antwortfolien bzw. Stimuli können mit Randomisierungsstrategien durchmischt werden, um eine möglichst zufällige Reihenfolge pro Versuchsperson zu erhalten und Reihenfolgeeffekte auszuschließen.

Auf den Antwortfolien können zusätzlich drei Buttons mit unterschiedlicher Funktionalität angezeigt werden. Die Beschriftung der Buttons kann geändert werden. Die Funktionalität ist allerdings nicht austausch- oder erweiterbar. Keiner der Buttons ist für ein Experiment mit *ExperimentMFC* erforderlich. Sie können aber nützlich sein, um der Versuchsperson die Möglichkeit zu geben, den Ablauf des Experiments zu beeinflussen. Mit dem *OK-Button* gelangt die Versuchsperson zu der jeweils nächsten Folie. Falls dieser Button definiert ist, muss er sogar gedrückt werden, damit das Experiment weitergehen kann. Bei Diskriminations- und Identifikationsexperimenten ist dieser Button eigentlich überflüssig. Der Button ist allerdings auf der Titelfolie sinnvoll, oder wenn die Antworten Sounds sind. Der *Replay-Button* ermöglicht das erneute Abspielen des aktuellen Stimulus. Ob und wie oft der Replay-Button gedrückt wurde, wird allerdings nicht gespeichert. Es kann aber festgelegt werden, wie oft der Replay-Button pro Stimulus gedrückt werden darf. Versuchspersonen können mit dem *Oops-Button* die Antwort für eine vorangegangene Folie korrigieren. Dies kann allerdings während des Experiments für die Versuchsperson schnell unübersichtlich werden. Wenn der *Oops-Button* nicht definiert ist, können Versuchspersonen ihre Antworten nicht im Nachhinein korrigieren, sondern müssen sich direkt für eine Antwort entscheiden.

Außer diesen drei Folienarten, Titel-, Antwort- und Endfolie, sind nur noch Pausenfolien möglich, um der Versuchsperson während des Experiments Pausen zu erlauben. Pausen sind wichtig, um den Stress, der auf der Versuchsperson lastet, möglichst gering zu halten. Falls der Wert dort 0 ist, gibt es in dem Experiment keine Pausen. In *ExperimentMFC* ist es nicht möglich, die Länge der Pause zu bestimmen. Die Versuchsperson bestimmt durch Klicken selbst, wann es weitergehen soll. Sonst sind keine weiteren Folien möglich. Es wäre also beispielsweise nicht möglich nach dem eigentlichen Experiment noch einen Feedback-Fragebogen anzuzeigen. Jedes Experimentfenster enthält zudem eine Menüleiste und das obligatorische, rote Kreuz in der oberen rechten oder linken Ecke, mit der das Experimentfenster geschlossen werden kann. Mittels dieses Buttons kann das Experiment beendet und das Experimentfenster geschlossen werden. Falls das Experiment jedoch vorzeitig, also noch vor der Endfolie, durch Klicken des Kreuzes abgebrochen wird, werden die Ergebnisse des Experiments nicht(!) gespeichert. In der Menüleiste befindet sich zum einen eine weitere Möglichkeit, das Experimentfenster zu schließen (File > Close) und Informationen über *Praat* (Query > Editor info) abzurufen.

3.3 Aufbau der *ExperimentMFC*-Steuerungsdatei

Die Implementierung der Steuerungsdatei kann in jedem beliebigen Texteditor erfolgen, da die Datei eine Textdatei mit der Dateiendung *.txt* ist. Die Steuerungsdatei kann auch mittels des in *Praat* eingebauten Skripteditors erstellt und dann als Textdatei abgespeichert werden. Allerdings bietet dieser keine Autovervollständigung oder ähnliche vorteilhafte Features, über die andere Texteditoren standardmäßig verfügen. Bei der

Erstellung des Skripts ist die Vermeidung von syntaktischen Fehlern besonders wichtig, damit *Praat* die Datei korrekt einlesen und ausführen kann. Am besten benutzt man daher eine Vorlage[7], um syntaktische Fehler zu vermeiden und sich zudem Schreibarbeit zu sparen. Im Anhang gibt es zusätzlich eine Übersicht zu typischen Fehlermeldungen von *Praat* und der möglichen Behebung dieser Fehler, da die Fehlermeldungen zum Teil wenig aussagekräftig sind.

Im Allgemeinen kann man die Steuerungsdatei als Liste von Parametern beschreiben, die den Ablauf des Experiments bestimmen. Die Parameter haben dabei eine feste, unveränderliche Reihenfolge. Es können auch keine Parameter weggelassen oder hinzugefügt werden, da die Datei sonst nicht mehr ausführbar ist. Für die Ausführbarkeit muss außerdem darauf geachtet werden, dass die jeweiligen Parameter den richtigen Parametertyp haben. In der Steuerungsdatei sind die Parametertypen *String* (Textparameter in doppelten Anführungszeichen für Beschriftungen oder Dateinamen, z.B. "Name"), *Integer* (numerische Parameter in Form von Ziffern für die Angabe einer Anzahl oder Position, z.B. 42) oder *Keyword* (Schlüsselwörter in spitzen Klammern: <yes>, <no> oder eine von fünf verschiedenen Randomisierungsstrategien, z.B. <PermutAll>) verfügbar. Die Parameter werden, wie in Programmiersprachen üblich, mit Bezeichnungen aufgeführt, die Binnenmajuskeln enthalten, und sie können u.a. Beschriftungen, Stimuli, die Randomisierungsstrategie, Speicherorte von Dateien oder die Position von Antwortfeldern spezifizieren. Alle Parameter können mit Kommentaren versehen werden, z.B. um gewisse Entscheidungen beim Aufbau des Experiments zu dokumentieren. Kommentare sind zwar in jeglicher Form erlaubt, sofern sie nicht wie die Parametertypen entweder in doppelten Anführunszeichen, in spitzen Klammern stehen oder Ziffern enthalten. Zur besseren Lesbarkeit der Datei sollten Kommentare aber z.B. durch ++ oder -- kenntlich gemacht werden. Die Bedeutung der einzelnen Parameter wird in den nachfolgenden Unterkapiteln erläutert.

Von besonderer Wichtigkeit ist auch der Speicherort der Steuerungsdatei und der Stimuli. Am besten erstellt man ein neues Verzeichnis für das Experiment und speichert dort die Steuerungsdatei. Die Stimuli – auch eventuelle Störsignale – sollten dann in dasselbe Verzeichnis oder in ein dazugehöriges Unterverzeichnis gespeichert werden. Zum einen werden dadurch Fehler vermieden, zum anderen kann so das gesamte Verzeichnis und somit das gesamte Experiment einfach portabel gemacht werden. Denn nun kann das Experimentverzeichnis z.B. auf einen USB-Stick kopiert werden.

3.3.1 Header und Stimuli-Spezifikation

Im Folgenden werden die einzelnen Zeilen der Steuerungsdatei genauer beschrieben, wobei austauschbare bzw. hier gewählte Textparameter und Ziffern rot und Schlüsselwörter blau dargestellt sind. Die ersten beiden Zeilen der Steuerungsdatei sind Textparameter

[7] Z.B. im Anhang dieser Arbeit oder auf der offiziellen Seite von ExperimentMFC: http://www.fon.hum.uva.nl/praat/manual/ExperimentMFC_2_1__The_experiment_file.html

und eigentlich[8] immer gleich:

```
"ooTextFile"
"ExperimentMFC 6"
```

Die erste Zeile gibt den Dateityp Textdatei an, während die zweite Zeile angibt, dass es sich bei der Datei um den *Praat*-Objekttyp *ExperimentMFC* handelt und um welche Version von *ExperimentMFC* es sich handelt. Die richtige Version ist wichtig, da die geforderte Parameteranzahl der neueren Version 6 sich von den älteren Versionen unterscheidet. Die nächste Zeile

```
blankWhilePlaying? = <yes>
```

gibt an, dass die Antwortmöglichkeiten während des Abspielens der Stimuli ausgeblendet werden und eine leere Folie angezeigt wird. Falls die Antwortmöglichkeiten immer angezeigt werden sollen, muss <yes> hier durch <no> ersetzt werden. In der nächsten Zeile

```
stimuliAreSounds? = <yes>
```

wird festgelegt, dass die Stimuli Sounds sind. Wenn hier <no> angegeben wird, kommt es zu einem fatalen Laufzeitfehler, der *Praat* zum Absturz bringt, weil dann die folgenden sieben Stimulusparamater fehlen müssen. Es müssen dann bei dem späteren `responsesAreSounds? = <yes>` und mehrere andere Parameter angegeben werden und die Antworten mit entsprechenden Dateinamen für Stimuli versehen werden. In dem Fall sind die Antworten Sounds (vgl. 3.3.3 und Beispiel im Anhang). Die nächsten beiden Zeilen geben den Dateipfad zum Verzeichnis an, in dem die Stimulidateien gespeichert sind und ihren Dateityp (inklusive vorausgehendem Punkt). Da man nur hier einen Dateipfad angeben kann, ist es wichtig, dass möglichst auch eventuelle Störsignale wie Rauschen im selben Verzeichnis wie die restlichen Stimuli abgespeichert sind. Ebenso kann man offensichtlich nicht verschiedene Dateitypen mischen.

```
stimulusFileNameHead = "C:/Experimente/Sounds/"
stimulusFileNameTail = ".wav"
```

Mit dem oben genannten beispielhaften Dateipfad könnte man also einen Stimulus ‚C:/Experimente/Sounds/jasne-sound1.wav' aufrufen. Wenn die Steuerungsdatei im selben Verzeichnis wie die Stimuli gespeichert ist, kann der Dateipfad auch mit `stimulusFileNameHead = ""` einfach weggelassen werden. Andere Sounddateitypen wie ".mp3" sind hier ebenfalls möglich. Allerdings ist die Kombination von verschiedenen Sounddateitypen nicht möglich, da die Dateiendung der Stimuli nur einmal spezifiziert werden kann. Die nächsten beiden Zeilen

```
stimulusCarrierBefore = "testton"
stimulusCarrierAfter = "endeton"
```

spezifizieren den Dateinamen von weiteren Stimuli wie Störsignalen, die vor (`Before`) oder nach (`After`) dem eigentlichen Stimulusabgespielt werden. So können auch

[8] Falls eine Steuerungsdatei X Teilexperimente enthält, kann dies mit "Collection" X angegeben werden, wobei X die Anzahl der Teilexperimente in Form einer Ziffer ist. In dem Fall beginnt die Steuerungsdatei einmal mit "ooTextFile", in der zweiten Zeile kommt die Angabe, dass eine Collection aus X Teilexperimenten vorliegt und dann folgen die Teilexperimente. Diese beginnen ab der Zeile "ExperimentMFC 6", der noch ein Namensparameter wie "jasne_1" hinzugefügt wird. Vgl. dazu auch Mayer (2013: 235 ff.).

aufgenommene Trägersätze wie „Es ist“ ... „Uhr!“, die als zwei Dateien im gleichen Verzeichnis wie die anderen Stimuli (`stimulusFileNameHead`) abgespeichert sind, um den eigentlichen Stimulus herum gebaut werden. Wie oben vermerkt, müssen diese Dateien die gleiche Dateiendung haben wie die anderen Stimuli. Entsprechend dürfen in diesen beiden Zeilen jeweils keine Dateiendung angegeben werden, da diese schon bei `stimulusFileNameTail` angegeben ist. Wenn es keine Störsignale geben soll, können die Parameter jeweils mit "" beschrieben werden. Die Parameter dürfen aber nicht einfach weggelassen werden, da die Datei sonst nicht mehr in *Praat* ausführbar ist. Im *jasně*-Experiment wurde nur vor dem eigentlichen Stimulus das Störsignal "Stimbeep.wav" bei `stimulusCarrierBefore` eingefügt und der `stimulusCarrierAfter` blieb mit "" frei. Es ist also auch möglich, dass nur einer der beiden Parameter spezifiziert wird. Die nächsten drei Zeilen geben die Dauer der Stille vor, zwischen und nach den Stimuli in Sekunden an:

```
stimulusInitialSilenceDuration = 0.5
stimulusMedialSilenceDuration = 1
stimulusFinalSilenceDuration = 0
```

Die erste Zeile gibt also an, wie lange es in Sekunden ab dem Zeitpunkt des Folienwechsels dauert bis der Stimulus bzw. das Störsignal oder die Trägerphrase, falls es diese gibt, abgespielt wird. Die zweite Zeile gibt das Interstimulusintervall ISI an (hier eine Sekunde), das sinnvoll wird, falls ein Diskriminationsexperiment erstellt werden soll und pro Antwortfolie mehr als ein Stimulus abgespielt wird. Ein ISI von 0 ist nicht sinnvoll, da dies laut McGuire (2010: 4) zu längeren Reaktionszeiten führen kann. Man sollte nur 0 angeben, wenn nur ein Stimulus pro Antwortfolie abgespielt werden soll. Kürzere ISI seien allerdings aufgrund von Effekten der Funktionalität des menschlichen Gedächtnisses besser, wobei die Werte normalerweise zwischen 0.1 und 1 Sekunde liegen. In Experimenten, in denen die Reihenfolge der Stimuli eine Rolle spielt, erschwert ein kürzeres ISI aber die Aufgabe für die Versuchsperson. Die dritte Zeile gibt die Dauer in Sekunden zwischen dem Ende des Stimulus und der Anzeige der Antwortfolie an, falls oben bei `blankWhilePlaying?` <yes> angegeben wurde und die Antwortfolien zunächst leer sind. Wenn dort <no> angegeben wurde, wird die Ziffer dort ignoriert, da sie nicht sinnvoll ist. Diese Zeiten können später bei der Analyse der Reaktionszeitmessung mit bedacht werden, da die Reaktionszeitmessung beginnt, sobald eine neue Antwortfolie angezeigt wird. Falls vor dem Stimulus ein Intervall eingefügt wurde, ist diese Zeit für die eigentliche Messung nicht unbedingt relevant, da sie vor dem zu beurteilenden Reiz liegt. Es kommt auch auf die Länge des aktuellen Stimulus an, ob dieser in die Reaktionszeit mit einbezogen werden soll. Bei längeren Stimuli kann die Versuchsperson die Entscheidung schon während des Hörens treffen und sofort nach dem Ende des Stimulus eine Antwort wählen. Während bei kurzen Aufnahmen eventuell noch die Zeit der eigentlichen Entscheidung mitgerechnet werden muss, die aber auch durch ein nachfolgendes Intervall abgesichert werden kann.

```
numberOfDifferentStimuli = 4
```

Der Wert der in der nächsten Zeile angeben wird, bestimmt die Anzahl der nächsten Zeilen und spezifiziert die Anzahl der verschiedenen Stimuli bzw., im Fall eines Diskriminationsexperiments, Stimulikombinationen. Für das Beispiel bedeutet das, dass

nun genau vier Zeilen, nicht mehr und nicht weniger, folgen müssen. Diese vier Zeilen bestehen jeweils aus zwei Textparametern, bei denen der erste Textparameter für den Dateinamen ohne Dateiendung eines Stimulus steht und der zweite Textparameter für die stimulusabhängige Beschriftung eines Antwortfeldes. Falls das Antwortfeld immer gleich beschriftet sein soll, z.B. mit (JA) und (NEIN), muss der zweite Textparameter mit "" beschrieben werden. Falls man für jeden Stimulus eine bestimmte Antwortbeschriftung erreichen möchte, könnten die vier Zeilen wie folgt aussehen:

```
"H2a_2_FEJA_jasne" "|neutral|JA|NEIN"
"H2a_2_JANA_jasne" "|eifrig|JA|NEIN"
"H1c_5_JARS_jasne" "|abwürgend|JA|NEIN"
"H1c_5_JARS_jasne" "|neutral|JA|NEIN"
```

Dabei ist z.B. `"H2a_2_FEJA_jasne"` der Dateiname ohne Dateiendung, und der zweite Textparameter gibt jeweils die Beschriftungen der einzelnen Antwortfelder an, wobei jeder senkrechte Strich | angibt, dass hier eine neue Beschriftung beginnt. Der Trennstrich wird nicht als Beschriftung angezeigt. Entsprechend sind in diesem Beispiel drei Antwortfelder vorhanden. Ein Antwortfeld für die funktionale Kategorie, eines für die Antwort (JA) und eines für die Antwort (NEIN). Für das *jasně*-Experiment bedeutete dies `numberOfDifferentStimuli = 128` und nachfolgend 128 Zeilen mit den verschiedenen Kombinationen von Kategorien und Stimuli. Die Kombinationen wurden dabei mit Hilfe eines eigens geschriebenen Permutationsprogramms in der Programmiersprache Haskell erstellt, um Zeit und Schreibarbeit zu sparen.

Falls es sich um ein Experiment handelt, bei dem die Antworten Sounds sind, gibt der erste Textparameter an, was in die Ergebnistabelle eingetragen werden soll, also für welche Antwort die Versuchsperson sich entschieden hat. Der zweite Textparameter steht dann für mögliche Instruktionen oder Fragen (vgl. Beispiel im Anhang).

Falls ein Diskriminationsexperiment erstellt werden soll, in dem mehrere Stimuli pro Antwortfolie abgespielt werden soll, werden diese im ersten Textparameter getrennt durch Kommata angegeben. Die Reihenfolge ist dabei relevant, d.h. im unten angeführten Beispiel würde immer zuerst FEJA und dann JARS abgespielt werden. Eine solche Kombination, die auch aus drei, vier oder mehr Stimuli bestehen kann, wird dann bei `numberOfDifferentStimuli` als ein Stimulus gezählt. Der zweite Textparameter ist hier beispielhaft leer:

```
"H2a_2_FEJA_jasne,H1c_5_JARS_jasne" ""
```

3.3.2 Spezifikation des Ablaufs und Folienbeschriftungen

Es folgen Angaben zum Ablauf des Experiments, die maßgeblich die Dauer des Gesamtexperiments bestimmen. Die nächsten zwei Zeilen nach den Angaben zu den Stimuli spezifizieren Wiederholungen und Pausen:

```
numberOfReplicationsPerStimulus = 3
breakAfterEvery = 64
```

Die erste Zeile gibt an, wie oft jede Antwortfolie und damit wie oft jeder Stimulus bzw. jede Stimulikombination im Experiment vorkommen soll. Der Wert sollte dabei größer Null sein, da es sonst keine Antwortfolien gibt und während des Experiments nur ein

Hinweis, dass es keine Antwortfolien gibt, angezeigt wird. Wiederholungen sind für Perzeptionsexperimente wichtig, da sie Fehler reduzieren, das Ergebnis festigen und weniger zufällig machen. Für das *jasně*-Experiment wurden wie oben drei Wiederholungen gewählt, sodass es 3*128=384 Antwortfolien gab. Die zweite Zeile gibt an, nach wie vielen Folien jeweils eine Pausenfolie eingeblendet wird. Im *jasně*-Experiment wurde nach jeder 64. Folie eine Pausenfolie angezeigt.

In der nächsten Zeile wird eine von fünf Randomisierungsstrategien gewählt:

```
randomize = <PermuteBalancedNoDoublets>
```

Bei den Randomisierungsstrategien muss man sich die Stimuli, die nach `numberOfDifferentStimuli` aufgeführt sind, als eine Einheit oder einen Block von Stimuli vorstellen. Im Folgenden wird als Rechenbeispiel davon ausgegangen, dass bei `numberOfReplicationsPerStimulus` wie oben 3 eingetragen wurde. Das heißt, dass es in dem Fall drei Blöcke gibt. Bei `<PermuteBalancedNoDoublets>` werden die Stimuli innerhalb dieser drei Blöcke zufällig durchmischt. Dabei bedeutet `NoDoublets`, dass vermieden wird, dass bei Übergängen von einem Block zum nächsten zwei gleiche Stimuli direkt hintereinander stehen. Diese Strategie wurde auch im *jasně*-Experiment eingesetzt. `<PermuteBalanced>` nimmt eine ähnliche Randomisierung vor, nur ohne das Vermeiden von Duplikaten an Übergängen. Demgegenüber steht die Strategie `<CyclicNonRandom>`, bei der die Blöcke ohne Durchmischung und somit ohne Änderung der Reihenfolge nacheinander abgearbeitet werden. Wenn man also möchte, dass genauso so vorgegangen wird, wie in der Steuerungsdatei angegeben, sollte man sich für diese Strategie entscheiden. Meist wird `<CyclicNonRandom>` eher für Testzwecke eingesetzt.

Für die Randomisierungsstrategie `<PermuteAll>` werden keine Blöcke gebildet. Alle Stimuli werden durchmischt und tauchen so oft auf, wie in `numberOfReplicationsPerStimulus` angegeben, wobei es aber auch zu doppeltem Vorkommen kommen kann. In `<WithReplacement>` kann es passieren, dass manche Stimuli öfter als andere und manche Stimuli gar nicht vorkommen. Dazu muss man sich die Abfolge des Experiments rundenbasiert vorstellen. Jeder Stimulus hat in jeder Runde die gleiche Wahrscheinlichkeit, an die Reihe zu kommen. Insgesamt kommen nur so viele Stimuli vor, wie durch die `numberOfDifferentStimuli` multipliziert mit der Anzahl der `numberOfReplicationsPerStimulus` angegeben. Beispielsweise ist `numberOfDifferentStimuli` als 4 und `numberOfReplicationsPerStimulus` als 3 angegeben, wodurch es insgesamt zwölf Antwortfolien bzw. Runden gibt. Allerdings wird pro Runde für jede Antwortfolie neu entschieden, welcher Stimulus an die Reihe kommt. Es kann also passieren, dass ein Stimulus fünfmal an die Reihe kommt, obwohl nur 3 Wiederholungen angegeben wurden. Oder dass ein Stimulus zufällig nie an die Reihe kommt, weil innerhalb der zwölf Runden immer andere Stimuli den Vorzug bekommen haben.

In den nächsten vier Parametern werden die Folienbeschriftungen der Titel- und Endfolie bzw. der Antwort- und Pausenfolien spezifiziert:

```
startText = "Dies ist ein Testexperiment.
Es können Instruktionen für das Experiment folgen!
```

```
   Klicke in das Fenster, um zu starten."
```

In dem Textparameter `startText` wird der Text spezifiziert, der auf der Titelfolie angezeigt werden soll. Meist werden hier Instruktionen zum Experiment aufgeführt. Im Folgenden wird beschrieben, was formal für alle Textparameter gilt. Im Beispiel sieht man, dass alle Zeilen manuell umgebrochen werden müssen, da ansonsten Zeilen, die für die Bildschirmauflösung zu lang sind, nicht vollständig angezeigt werden können. Wenn man eine leere Zeile auf der Folie anzeigen möchte, muss auch im Textparameter eine leere Zeile einfügen (s.o.). Falls man doppelte Anführungszeichen auf seinen Folien benutzen möchte, muss man diese doppelt aufführen, da sie ja eigentlich Textparameter spezifizieren. Wenn man also "hallo" schreiben möchte, muss der Textparameter wie solgt aussehen: ""hallo"". Wenn Text kursiv angezeigt werden soll, muss er zwischen %% und % stehen. Wenn Text fett angezeigt werden soll, muss er zwischen ## und # stehen. Weiter unten finden sich Beispiele für solche Schriftsätze. Wie zusätzlich die Schriftgröße auf Antwortfeldern geändert werden kann, findet sich in Abschnitt 3.3.3. Die nächste Zeile gibt den Instruktionstext an, der als Überschrift auf jeder Antwortfolie angezeigt wird:

```
runText = "Bitte antworte mit <JA> oder <NEIN>.
   Der Sprecher klingt..."
```

Der nächste Parameter spezifiziert den Text, der auf einer Pausenfolie angezeigt werden kann. Dabei wird im Beispiel das Wort Pause *kursiv* gesetzt.

```
pauseText = "Mach mal %%Pause%!
   Klicken damit es weitergeht!"
```

Der darauffolgende Parameter gibt den Text, der Endfolie an, wobei das Wort Danke **fett** gedruckt wird:

```
endText = "##Danke# fürs Mitmachen!"
```

Alle hier angegebenen Textelemente wie Folien- und Buttonbeschriftungen werden automatisch positioniert. Im Folgenden Abschnitt wird erklärt, wie man Buttons und Antwortfelder gezielt auf Folien positioniert.

3.3.3 Buttons, Antworten, Skalen und ihre Positionierung

Die nächsten Parameter beschreiben die Buttons:

```
maximumNumberOfReplays = 2
replayButton =
   0.35 0.65 0.1 0.2 "Nochmal hören?" ""
okButton = 0.4 0.6 0.2 0.3 "OK" ""
oopsButton = 0 0 0 0 "" ""
```

Die erste Zeile gibt an, dass die Versuchsperson den Stimulus oder die Stimuli noch maximal zweimal zusätzlich anhören darf. Diesen Wert größer Null zu setzen macht nur Sinn, wenn der ReplayButton auch definiert wird. Wenn hier 0 steht, hat die Versuchsperson nicht die Möglichkeit, Stimuli noch einmal anzuhören. Die anderen drei Zeilen spezifizieren die Buttons ReplayButton, OKButton und OopsButton. Der Aufbau ist dabei immer gleich. Die ersten vier Integer geben die Koordinaten für die Positionierung der Buttons an (s.u.). Wenn hier wie bei obigem OopsButton vier Nullen stehen, wird der

Button nicht angezeigt. Der erste Textparameter steht für die Beschriftung des Buttons. Der zweite String kann ein Tastenkürzel angeben, das denselben Effekt wie das Drücken des Buttons hat. Zum Beispiel kann dort mit " " ein Leerzeichen eingefügt werden, sodass das Klicken des Leerzeichens auf der Tastatur den Button während des Experiments auslöst. In diesem Fall sollte man die Versuchsperson darauf hinweisen, dass dieses Mittel zur Verfügung steht.

Man kann sich die grafische Oberfläche wie ein Koordinatensystem mit x- und y-Achse vorstellen, wobei sich der Nullpunkt des Koordinatensystems in der linken unteren Ecke des Fensters befindet. Die Koordinaten werden als vier Werte zwischen 0 und 1 (mit Dezimal*punkt* und beliebiger Anzahl Nachkommastellen) angegeben, sodass Antwortfelder und Buttons im Prinzip überall[9] auf der gesamten grafischen Oberfläche positioniert werden können. Die ersten beiden Koordinaten beschreiben die x-Achse und somit die horizontale Ausdehnung. Die zweiten beiden Koordinaten beschreiben die y-Achse und somit die vertikale Ausdehnung. Mayer (2013: 224) hat dies in einem Schema verdeutlicht:

horizontale Ausdehnung		**vertikale Ausdehnung**	
linke Kante	rechte Kante	untere Kante	obere Kante
X_{links}	X_{rechts}	Y_{unten}	Y_{oben}

Abbildung 5. Koordinatenschema, entnommen aus Mayer (2013: 224).

So können rechteckige Buttons, aber ebenso Antwortfelder aufgespannt und angezeigt werden. Die Erstellung der Antwortfelder werden im späteren Verlauf beschrieben. Für eine eindeutige Darstellung des Koordinatensystems siehe Abbildung 6.

In der nächsten Zeile nach den Angaben zu den Buttons folgt die Frage, ob die Antworten Sounddateien sind:

```
responsesAreSounds? <no> "" "" "" "" 0 0 0
```

Wenn hier wie oben <no> angegeben wird, müssen die folgenden Parameter als "" oder 0 angegeben werden. In dem Fall sind die Antworten die an anderer Stelle angegebenen Stimulibeschriftungen. Falls hier <yes> steht, sind die Antworten Sounds, und es müssen die folgenden weiteren Parameter analog zu den Beschreibungen der Stimuli in Abschnitt 3.3.1 spezifiziert werden:

```
responseFileNameHead = "C:/Experimente/Sounds/"
responseFileNameTail = ".wav"
responseCarrierBefore = "testton"
responseCarrierAfter = ""
responseInitialSilenceDuration = 0
responseMedialSilenceDuration = 0.5
responseFinalSilenceDuration = 0
```

Im Anhang findet sich ein Beispiel für ein Experiment, in dem die Antworten Sounds

[9] Buttons und Antwortfelder sollten nicht zu hoch platziert werden, da sie sonst ggf. Folienüberschriften überdecken.

sind. Danach folgt die Spezifikation der Antwortfelder, wobei die erste Zeile wieder die Anzahl der Antwortfelder bzw. die Anzahl der nachfolgenden Zeilen bestimmt:

```
numberOfDifferentResponses = 2
   0.2 0.4 0.4 0.6 "Ja" 60 " " "j"
   0.6 0.8 0.4 0.6 "Nein" 40 "" "n"
```

Die beiden unteren Zeilen bestimmen die Position, Beschriftung, Schriftgröße, Tastenkürzel und den Eintrag in der Ergebnistabelle von zwei Antwortfeldern. Die ersten vier Parameter stehen wie bei den Buttons für Koordinaten und funktionieren auf dieselbe Weise wie oben erklärt. Der fünfte Parameter gibt die Beschriftung des Antwortfeldes an, der sechste Parameter ändert die Schriftgröße. Standardmäßig eignet sich 40 als Schriftgröße, allerdings kommt es natürlich auf die Länge des Beschriftungstextes und die Größe des Antwortfeldes an. Der siebente Parameter kann wie bei den Buttons für ein Tastenkürzel zur Bedienung des Antwortfeldes dienen. Der letzte Parameter bestimmt, was in die Ergebnistabelle eingetragen werden soll, um festzuhalten, für welche Antwort sich die Versuchsperson entschieden hat. Wenn dieser letzte Parameter mit "" leer bleibt, wird das Antwortfeld zwar angezeigt und beschriftet, ist aber ausgegraut und kann nicht angeklickt werden. Diesen Umstand kann man z.B. nutzen, wenn man wie beim *jasně*-Experiment pro Antwortfolie einen bestimmten Namen einer Kategorie anzeigen möchte, aber dieser nicht als Antwortmöglichkeit angezeigt werden soll. Falls die Antworten Sounds sind, gibt der letzte Parameter den Dateinamen des Stimulus oder der Stimuli an.

Im obigen Beispiel werden die Antwortfelder auf jeder Antwortfolie mit "Ja" bzw. "Nein" beschriftet. Wenn eine stimulusabhängige Beschriftung gewünscht wird, muss die Stimulusspezifikation, wie in 3.3.1 erläutert, geändert werden. Bei `numberOfDifferentResponses` muss dann dieselbe Anzahl stehen, wie es Trennstriche | gibt. Dann werden die verschiedenen Antwortfelder wie sonst auch angelegt. Allerdings muss der fünfte Textparameter dann leer bleiben, da diese Beschriftung ja pro Antwortfolie verschieden ist (vgl. *jasně*-Experiment im Anhang). Wenn die Antworten Sounds sind, werden sie als anklickbare Antwortfelder dargestellt, sodass das Abspielen beginnt, sobald auf das Feld geklickt wurde. In dem Fall ist ein zusätzlicher OK-Button nötig, um zur nächsten Folie zu gelangen, da das Experiment sonst nicht fortgesetzt werden kann. Es ist wichtig, dass bei "Antworten sind Sounds", der Wert von `maximumNumberOfReplays` nicht auf 0 ist, da die Stimuli sonst nie wiederholt und somit nicht verglichen werden können. In neueren Versionen von *ExperimentMFC* ist es unter Windows und MAC auch möglich, Bilder als Antwortmöglichkeiten anzuzeigen. Dafür muss an Stelle des fünften Parameters, der sonst die Antwortfeldbeschriftung angibt, der Dateipfad und Dateiname des Bildes mit Hilfe von \FI angegeben werden:

```
0.2 0.4 0.4 0.6 "\FI/Bilder/hoch.jpg" 40 "" "hoch"
```

Die Schriftgröße ist dabei egal. Die Bilder werden so skaliert, dass sie in das Antwortfeld passen[10]. Es sind dabei auch andere Bilddateitypen wie ".png" möglich.

Als Letztes fehlt noch die Spezifikation der Bewertungsskala. Wenn keine Bewertungsskala gebraucht wird, geht man wie folgt vor, und die Steuerungsdatei ist fertig:

```
numberOfGoodnessCategories = 0
```

[10] Achtung: Dafür werden die Bilder möglicherweise gestreckt oder gestaucht.

Wenn man eine Bewertungsskala haben möchte, gibt man den Bereich der Skala bei `numberOfGoodnessCategories` an und fügt so viele Zeilen wie die Range angibt hinzu:

```
numberOfDifferentResponses = 3
   0.2 0.4 0.1 0.2 "absolut sicher"
   0.4 0.6 0.1 0.2 "sicher"
   0.6 0.8 0.1 0.2 "unsicher"
```

Dabei geben die ersten vier numerischen Parameter wieder die Koordinaten an, an denen die Skala positioniert werden soll. Der Textparameter gibt jeweils die Beschriftung der Skala und die Bezeichnung der Antwort in der Ergebnistabelle an. Damit ist die Steuerungsdatei fertig. Beispielhafte Steuerungsdateien befinden sich im Anhang.

Die folgende Abbildung 6 mit Koordinatensystem aus Mayer (2013: 225) zeigt noch einmal zusammenfassend, wie die Antwortfelder und die Bewertungsskala beispielsweise positioniert werden können:

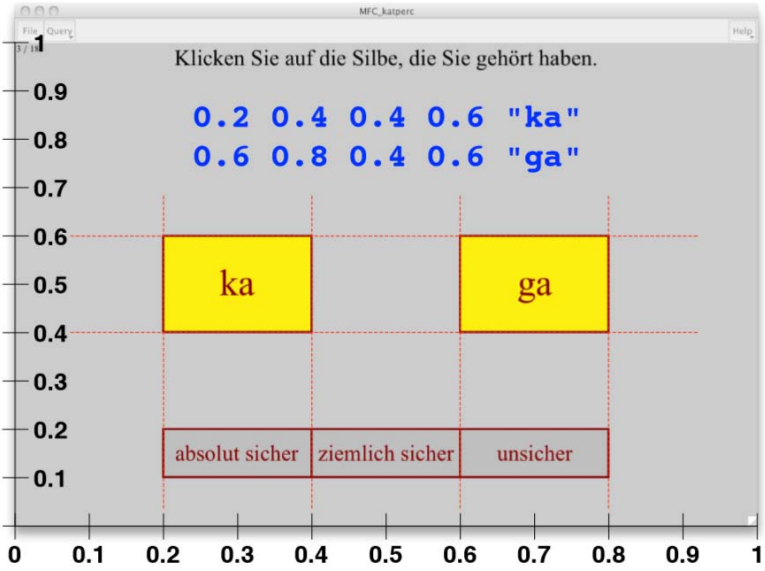


Abbildung 6. Koordinatensystem des ExperimentMFC-Fensters zur Platzierung grafischer Elemente, entnommen aus Mayer (2013: 225). Die schwarzen Zahlen und Linien zeigen dabei das Koordinatensystem an. Die blaue Schrift steht für die Angaben in der Steuerungsdatei (hier für Antwortfelder). Die entsprechenden Angaben für die Ratingskala wären:

0.2 0.4 0.1 0.2 "absolut sicher"
0.4 0.6 0.1 0.2 "sicher"
0.6 0.8 0.1 0.2 "unsicher".

Übrigens könnte die Steuerungsdatei beispielsweise auch wie folgt aussehen, da die Parameternamen nicht vorhanden und die Parameter auch nicht in einzelnen Zeilen sein müssen:

```
"ooTextFileExperimentMFC 6" <yes> <yes> "sounds/" ".wav" "" ""
0.5 0 0 6 "ga0" "" "ga1" "" "ga2" "" "ga3" "" "ga4" "" "ga5" "" 3 0
<PermuteBalancedNoDoublets> "Dies ist ein Wahrnehmungsexperiment.
Zum Starten klicken." "Klicken Sie auf die Silbe, die Sie gehört haben." ""
"Das Experiment ist beendet. Vielen Dank!" 0 0 0 0 0 "" "" 0 0 0 0 "" "" 0
0 0 0 "" "" <no> "" "" "" "" 0 0 0 2 0.3 0.49 0.4 0.6 "ka" 60 "" "k" 0.51 0.7
0.4 0.6 "ga" 60 "" "g" 0
```

Dieses Beispiel stammt aus Mayer (2013: 222) und ist zwar schlecht lesbar, aber funktionstüchtig, weil die Reihenfolge der Parameter korrekt ist.

3.4 Durchführung eines Experiments in *Praat*

Für die Durchführung des Experiments müssen die Stimuli an dem in der Steuerungsdatei angegebenen Verzeichnissen gespeichert sein und *Praat* installiert und geöffnet sein. Die Stimuli müssen nur auf dem Rechner vorhanden sein und nicht zusätzlich in *Praat* geladen werden. Dann wird die Steuerungsdatei mit (Open 〉 Read from file...) geöffnet und als *ExperimentMFC*-Objekt in der Objektliste angezeigt. Die Steuerungsdatei ist kein Skript und kann daher nicht mit (Praat 〉 Open Praat script...) geöffnet werden. Es gibt nun auf der rechten Seite zwei Auswahlmöglichkeiten. Zum einen kann mit (Run) die Steuerungsdatei ausgeführt und das Experiment auf diese Weise mit der Titelfolie gestartet werden. Zum anderen können die Ergebnisse mit (Extract results) bereitgestellt und gespeichert werden, was natürlich erst nach der Durchführung des Experiments sinnvoll ist. Während des Experiments werden Reaktionszeiten gemessen, die immer ab der Präsentation einer neuen Antwortfolie gemessen wird.

Wenn während der Testphase Änderungen an der Steuerungsdatei vorgenommen wurden, ist es sinnvoll, die alte Steuerungsdatei immer erst mit (Remove) aus der Objektliste zu entfernen, damit es nicht zu Verwechselungen mit der neuen Steuerungsdatei kommen kann. In der Objektliste können auch mehrere Teilexperimente gleichzeitig geöffnet sein.

4. Verarbeitung der Ergebnisse

Die Ergebnisse eines Perzeptionsexperiments sind typischerweise die (relative) Anzahl und Art der Antworten pro Kategorie, die Reaktionszeiten und falls vorhanden, die Ergebnisse der Bewertungsskalen. Mit statistischen Methoden können die Daten dann eventuell normalisiert oder weiter aufbereitet werden. Falls die Bewertungen z.B. mit (gut)/(schlecht) beschrieben wurden, müssen diese in numerische Werte wie (1)/(2) überführt werden. Die Auswertung und Aufbereitung von Experimenten erfolgt

typischerweise in Tabellenkalkulationsprogrammen wie Excel oder Calc und Statistikprogrammen wie SPSS, Rapidminer oder R (s. Referenzen für weitere Informationen). Für die Auswertung des *jasně*-Experiments wurde beispielsweise mit den Daten der 15 deutschen Versuchspersonen eine dreifaktorielle univariate Varianzanalyse mit SPSS durchgeführt. Die drei abhängigen Variablen waren dabei die Kategorie (JA=1, NEIN=0), die Bewertung (EHER SICHER=1, EHER UNSICHER=0) und die Reaktionszeit in ms. Die Faktoren waren die assoziierte und die angebotene Funktion.

Problematisch bleibt die Auswertung der Reaktionszeiten, auch wenn sie ungefähre Einblicke geben kann. Die Messung der Reaktionszeit erfolgt meist ab der Präsentation des Stimulus und muss daher oft angepasst werden, auch in Bezug auf die Zeit, die eine Versuchsperson eventuell zum Lesen eines Labels oder einer Instruktion braucht. Dies ist auch in *ExperimentMFC* der Fall. Andere Software soll akkuratere Reaktionszeitmessungen bieten, die z.B. auch zwischen Reaktionszeit und Antwortzeit unterscheidet und andere Faktoren wie die Schnelligkeit von Tastatur und Computermaus berücksichtigt (vgl. Mayer (2013: 203)).

4.1 Speichern und Verarbeiten in Praat

In *ExperimentMFC* wird das Experiment korrekt beendet, wenn das Fenster beim Anzeigen der Endfolie geschlossen wird. Dann muss in *Praat* (Extract results) angeklickt werden, um die Ergebnisse anzuzeigen und zu speichern. Die Ergebnisse werden beim Start eines neuen Experiments überschrieben, daher ist dieser Schritt besonders wichtig. Dabei wird ein *ResultsMFC*-Objekt ezeugt, das am besten direkt mit (Rename...) umbenannt wird und so z.B. einer Versuchsperson zugeordnet wird. Das Objekt kann dann mit (Save > Save as text file...) gespeichert werden. Für das *ResultsMFC*-Objekt gibt es auf der rechten Seite außerdem die Möglichkeit, die Anzahl und Art der Antworten mit (Query) einzeln anzuzeigen oder die Ergebnisse mit (To Categories (stimuli)) bzw. (To Categories (responses)) in ein anderes Format zu konvertieren. Am sinnvollsten ist es aber, alle Teilergebnisse zu markieren und mit (Collect to Table) das Objekt ‚Table allResults' zu erzeugen. Die Teilergebnisse werden dann alle zusammen in einer Tabelle gesammelt. Wenn das Tabellenobjekt markiert ist, kann die Tabelle dann u.a. mit (View & Edit) angezeigt werden. Die Tabelle kann dann wiederum mit (Save > Save as tab-seperated file...) bzw. (Save > Save as comma-seperated file...) gespeichert und für andere Programme verfügbar gemacht werden. Sie ist dann als Table-Datei in andere Programme importierbar.

4.2 Aufbereitung der Ergebnisse in LibreOffice Calc oder Excel

In dem frei verfügbaren Tabellenkalkulationsprogramm Calc wählt man über (Einfügen > Tabelle aus Datei...), die Tabellendatei aus und markiert im Importdialog bei den Trennoptionen (Getrennt) und (Tabulator), falls man sich zuvor für eine tab-seperated Tabellendatei entschieden hat. Im weiteren Verlauf muss man für sein Experiment passende Optionen auswählen und schließlich ist die Tabelle in Calc für Berechnungen wie ANOVAs verfügbar. In Excel geht man ähnlich vor, indem man (Daten > Externe Daten

Textdatei importieren...) wählt, die Tabellendatei auswählt und dann dem Importassistenten folgt. Dabei muss man auch wieder auf Optionen wie (Getrennt) und (Tabstopp) achten. In Excel muss eventuell auch Unicode UTF-8 als Dateiursprung angegeben werden, da es sonst zu Kompabilitätsproblemen kommen kann. Beide Programme bieten zudem frühzeitig die Möglichkeit, mit Hilfe einer Vorschau zu prüfen, ob die Spalten und Zeilen alle richtig erkannt wurden. Nun können die Ergebnisse aus ExperimentMFC weiter analysiert und aufbereitet werden.

5. Referenzen

G. MCGUIRE. A Brief Primer on Experimental Designs for Speech Perception Research. Department of Linguistics UC Santa Cruz. Draft: Summer 2010.

H. MCGURK, J. MACDONALD. Hearing lips and seeing voices. Nature, 1976, 746-748.

Praat:
http://www.fon.hum.uva.nl/praat/

Offizielle Anleitung von ExperimentMFC:
http://www.fon.hum.uva.nl/praat/manual/ExperimentMFC.html

Ausführliche Anleitung für Praat und ExperimentMFC von Jörg Mayer:

- J. MAYER. Phonetische Analysen mit Praat. Ein Handbuch für Ein- und Umsteiger. Handbuch-Version 2013/08.
- *http://praatpfanne.lingphon.net/downloads/praat_manual.pdf*

Weitere hilfreiche Software:

- Microsoft Excel Online: *https://office.live.com/start/Excel.aspx*
- LibreOffice Calc: *https://de.libreoffice.org/discover/calc/*
- SPSS: *http://www-01.ibm.com/software/de/analytics/spss/*
- Rapidminer: *https://rapidminer.com/*
- R: *http://www.r-project.org/*

6. Anhang

6.1 Steuerungsdatei (Mindestanforderungen)

Für die Ausführung dieser Steuerungsdatei wird nur ein Stimulus mit dem Dateinamen "testton.wav"benötigt, die im selben Verzeichnis wie die Steuerungsdatei abgespeichert ist.

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
stimulusFileNameHead = ""
stimulusFileNameTail = ".wav"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 0
stimulusMedialSilenceDuration = 0
stimulusFinalSilenceDuration = 0
numberOfDifferentStimuli = 1
        "testton" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 0
randomize = <CyclicNonRandom>
startText = "Starttext"
runText = "Instruktionstext"
pauseText = "Pausentext"
endText = "Endtext"
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0 0
numberOfDifferentResponses = 1
0.2 0.3 0.7 0.8 "Klick"40 "" "k"
numberOfGoodnessCategories = 0
```

6.2 Steuerungsdatei 1 des *jasně*-Experiments

Teilexperiment 1 des *jasně*-Experiments als Beispiel für ein vollständiges Identifikationsexperiment.

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
stimulusFileNameHead = "jasne-cut/"
stimulusFileNameTail = ".wav"
stimulusCarrierBefore = "Stimbeep"
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1.0
stimulusMedialSilenceDuration = 0
stimulusFinalSilenceDuration = 1.5
numberOfDifferentStimuli = 128
"H2a_2_FEJA_jasne" "|neutral|JA|NEIN"
"H2a_2_FEJA_jasne" eifrig|JA|NEIN"
"H2a_2_FEJA_jasne" "|abwürgend|JA|NEIN"
"H2a_2_FEJA_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_FEJA_jasne" "|erstaunt|JA|NEIN"
"H2a_2_FEJA_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_FEJA_jasne" "|resigniert|JA|NEIN"
"H2a_2_FEJA_jasne" "|beruhigend|JA|NEIN"
"H2a_2_JANA_jasne" "|neutral|JA|NEIN"
"H2a_2_JANA_jasne" "|eifrig|JA|NEIN"
"H2a_2_JANA_jasne" "|abwürgend|JA|NEIN"
"H2a_2_JANA_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_JANA_jasne" "|erstaunt|JA|NEIN"
"H2a_2_JANA_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_JANA_jasne" "|resigniert|JA|NEIN"
"H2a_2_JANA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_JARS_jasne" "|neutral|JA|NEIN"
"H1c_5_JARS_jasne" "|eifrig|JA|NEIN"
"H1c_5_JARS_jasne" "|abwürgend|JA|NEIN"
"H1c_5_JARS_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_JARS_jasne" "|erstaunt|JA|NEIN"
"H1c_5_JARS_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_JARS_jasne" "|resigniert|JA|NEIN"
"H1c_5_JARS_jasne" "|beruhigend|JA|NEIN"
"H2a_2_TROA_jasne" "|neutral|JA|NEIN"
"H2a_2_TROA_jasne" "|eifrig|JA|NEIN"
"H2a_2_TROA_jasne" "|abwürgend|JA|NEIN"
```

"H2a_2_TROA_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_TROA_jasne" "|erstaunt|JA|NEIN"
"H2a_2_TROA_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_TROA_jasne" "|resigniert|JA|NEIN"
"H2a_2_TROA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_BOZA_jasne" "|neutral|JA|NEIN"
"H1c_5_BOZA_jasne" "|eifrig|JA|NEIN"
"H1c_5_BOZA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_BOZA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_BOZA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_BOZA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_BOZA_jasne" "|resigniert|JA|NEIN"
"H1c_5_BOZA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_PCNA_jasne" "|neutral|JA|NEIN"
"H1c_5_PCNA_jasne" "|eifrig|JA|NEIN"
"H1c_5_PCNA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_PCNA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_PCNA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_PCNA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_PCNA_jasne" "|resigniert|JA|NEIN"
"H1c_5_PCNA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_GROA_jasne" "|neutral|JA|NEIN"
"H1c_5_GROA_jasne" "|eifrig|JA|NEIN"
"H1c_5_GROA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_GROA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_GROA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_GROA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_GROA_jasne" "|resigniert|JA|NEIN"
"H1c_5_GROA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_HEJA_jasne" "|neutral|JA|NEIN"
"H1c_5_HEJA_jasne" "|eifrig|JA|NEIN"
"H1c_5_HEJA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_HEJA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_HEJA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_HEJA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_HEJA_jasne" "|resigniert|JA|NEIN"
"H1c_5_HEJA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_HADA_jasne" "|neutral|JA|NEIN"
"H1c_5_HADA_jasne" "|eifrig|JA|NEIN"
"H1c_5_HADA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_HADA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_HADA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_HADA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_HADA_jasne" "|resigniert|JA|NEIN"

"H1c_5_HADA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_STOA_jasne" "|neutral|JA|NEIN"
"H1c_5_STOA_jasne" "|leifrig|JA|NEIN"
"H1c_5_STOA_jasne" "|abwürgend|JA|NEIN"
"H1c_5_STOA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_STOA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_STOA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_STOA_jasne" "|resigniert|JA|NEIN"
"H1c_5_STOA_jasne" "|beruhigend|JA|NEIN"
"H2a_2_HOBA_jasne" "|neutral|JA|NEIN"
"H2a_2_HOBA_jasne" "|leifrig|JA|NEIN"
"H2a_2_HOBA_jasne" "|abwürgend|JA|NEIN"
"H2a_2_HOBA_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_HOBA_jasne" "|erstaunt|JA|NEIN"
"H2a_2_HOBA_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_HOBA_jasne" "|resigniert|JA|NEIN"
"H2a_2_HOBA_jasne" "|beruhigend|JA|NEIN"
"H2a_2_OCEK_jasne" "|neutral|JA|NEIN"
"H2a_2_OCEK_jasne" "|leifrig|JA|NEIN"
"H2a_2_OCEK_jasne" "|abwürgend|JA|NEIN"
"H2a_2_OCEK_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_OCEK_jasne" "|erstaunt|JA|NEIN"
"H2a_2_OCEK_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_OCEK_jasne" "|resigniert|JA|NEIN"
"H2a_2_OCEK_jasne" "|beruhigend|JA|NEIN"
"H2a_2_SPIA_jasne" "|neutral|JA|NEIN"
"H2a_2_SPIA_jasne" "|leifrig|JA|NEIN"
"H2a_2_SPIA_jasne" "|abwürgend|JA|NEIN"
"H2a_2_SPIA_jasne" "|gleichgültig|JA|NEIN"
"H2a_2_SPIA_jasne" "|erstaunt|JA|NEIN"
"H2a_2_SPIA_jasne" "|erleuchtet|JA|NEIN"
"H2a_2_SPIA_jasne" "|resigniert|JA|NEIN"
"H2a_2_SPIA_jasne" "|beruhigend|JA|NEIN"
"H4a_4_SPIA_jasne" "|neutral|JA|NEIN"
"H4a_4_SPIA_jasne" "|leifrig|JA|NEIN"
"H4a_4_SPIA_jasne" "|abwürgend|JA|NEIN"
"H4a_4_SPIA_jasne" "|gleichgültig|JA|NEIN"
"H4a_4_SPIA_jasne" "|erstaunt|JA|NEIN"
"H4a_4_SPIA_jasne" "|erleuchtet|JA|NEIN"
"H4a_4_SPIA_jasne" "|resigniert|JA|NEIN"
"H4a_4_SPIA_jasne" "|beruhigend|JA|NEIN"
"H1c_5_TROA_jasne" "|neutral|JA|NEIN"
"H1c_5_TROA_jasne" "|leifrig|JA|NEIN"
"H1c_5_TROA_jasne" "|abwürgend|JA|NEIN"

"H1c_5_TROA_jasne" "|gleichgültig|JA|NEIN"
"H1c_5_TROA_jasne" "|erstaunt|JA|NEIN"
"H1c_5_TROA_jasne" "|erleuchtet|JA|NEIN"
"H1c_5_TROA_jasne" "|resigniert|JA|NEIN"
"H1c_5_TROA_jasne" "|beruhigend|JA|NEIN"
"H4c_2_STOA_jasne" "|neutral|JA|NEIN"
"H4c_2_STOA_jasne" "|eifrig|JA|NEIN"
"H4c_2_STOA_jasne" "|abwürgend|JA|NEIN"
"H4c_2_STOA_jasne" "|gleichgültig|JA|NEIN"
"H4c_2_STOA_jasne" "|erstaunt|JA|NEIN"
"H4c_2_STOA_jasne" "|erleuchtet|JA|NEIN"
"H4c_2_STOA_jasne" "|resigniert|JA|NEIN"
"H4c_2_STOA_jasne" "|beruhigend|JA|NEIN"
numberOfReplicationsPerStimulus = 3
breakAfterEvery = 64
randomize = <PermuteBalancedNoDoublets>
startText = "Dies ist ein Perzeptions-Experiment.
Du wirst dabei immer das tschechische Wort <jasne> hören,
was soviel wie <alles klar> bedeutet.

Das Wort wurde mit unterschiedlichen Stimmungen/Intentionen gesprochen.
D.h., der Sprecher klingt ...
<<neutral, eifrig, abwürgend, gleichgültig, erstaunt, erleuchtet, resigniert, beruhigend>>.

Deine Aufgabe ist es zu versuchen, diese Stimmungen/Intentionen zu erkennen, auch wenn es eine fremde Sprache ist. Dazu wirst Du immer ein Wort hören und eine der obigen 8 Kategorien sehen. Wenn Du meinst, dass der Sprecher in der Tat so klingt, wie es dort steht, dann antworte mit <JA>, sonst antworte mit <NEIN>. Antworte einfach spontan, ohne langes Nachdenken, denn es gibt keine falschen Antworten! Das Ganze dauert ca. 25 Minuten, mit Pausen dazwischen.

Bereit? Dann bitte einmal klicken!"
runText = "Bitte antworte mit <JA> oder <NEIN>. Der Sprecher klingt ..."
pauseText = "Mach mal Pause! Klicken damit es weitergeht!"
endText = "Danke fürs Mitmachen!"
maximumNumberOfReplays = 1
replayButton = 0.3 0.7 0.06 0.13 "Nochmal hören?"
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0 0
numberOfDifferentResponses = 3
0.2 0.8 0.75 0.85 "" 40 "" ""
0.2 0.45 0.4 0.65 "JA"40 "" "ja"

0.55 0.8 0.4 0.65 "NEIN"40 "" "nein"
numberOfGoodnessCategories = 2
0.3 0.5 0.2 0.3 "eher unsicher"
0.5 0.7 0.2 0.3 "eher sicher"

6.3 Steuerungsdatei (Beispiel für Antworten sind Sounds)

Anwendungsbeispiel für Antworten sind Sounds. Für die Ausführung werden Dateien mit den Namen "testton1.wav" und "testton2.wav" benötigt.

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <no> "" "" "" "" 0 0 0
numberOfDifferentStimuli = 2
        "ton1" "Welcher Sound ist deutlicher?"
        "ton2" "Welcher Sound ist lauter?"
numberOfReplicationsPerStimulus = 2
breakAfterEvery = 0
randomize = <PermuteAll>
startText = "Antworten sind Sounds. Test!"
runText = "Instruktionstext"
pauseText = "Pausentext"
endText = "Experiment beendet."
maximumNumberOfReplays = 5
replayButton = 0 0 0 0 "" ""
okButton = 0.9 1 0 0.1 "OK" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <yes>
responseFileNameHead = ""
responseFileNameTail = ".wav"
responseCarrierBefore = ""
responseCarrierAfter = ""
responseInitialSilenceDuration = 0
responseMedialSilenceDuration = 0.5
responseFinalSilenceDuration = 0
numberOfDifferentResponses = 2
        0.3 0.45 0.6 0.8 "Ein Ton" 20 "" "testton1"
        0.55 0.7 0.6 0.8 "Ein anderer
Ton" 20 "" "testton2"
numberOfGoodnessCategories = 2
        0.3 0.5 0.1 0.2 "sicher"
        0.5 0.7 0.1 0.2 "unsicher"
```

6.4 Typische Fehlermeldungen und ihre Behandlung

Keine Gewähr auf Vollständigkeit. Die Dateipfade, Dateinamen, Zeilennummern und Werte sind beispielhaft.

- Datei kann nicht geöffnet werden:
  - Cannot open file "C:\Experiment\sounds\beep.wav". ExperimentMFC "test" not started. Experiment window not created. Command "Run" not executed.
    → Die angegebene Datei beep.wav kann nicht geöffnet werden, da sie nicht existiert oder nicht im in der Steuerungsdatei test.txt angegebenen Verzeichnis abgespeichert ist. Überprüfe `stimulusFileNameHead` auf Richtigkeit, überprüfe auf Flüchtigkeitsfehler (Name richtig geschrieben?) und wo die einzelnen Stimuli (auch Rauschen oder Carrierphrasen) abgespeichert sind. Sind alle Stimuli im selben Ordner abgespeichert?
- Fehlende Angaben:
  - File "C:\Experiment\test.txt" not recognized.
    File "C:\Experiment\test.txt" not finished.
    → Zeile "ooTextFile" fehlt oder nicht in erster Zeile des Dokuments!
  - Found an enumerated value while looking for a string in text (line 2).
    Data not read from text file "C:\Experiment\test.txt".
    File "C:\Experiment\test.txt" not finished.
    → Zeile "ExperimentMFC 6" fehlt oder nicht in zweiter Zeile des Dokuments!
- Falscher Typ:
  - Found a number while looking for a string in text (line 14). String "name" not read. ExperimentMFC not read. Data not read from text file "C:\Experiment\test.txt". File "C:\Experiment\test.txt" not finished.
    → Es wurde in Zeile 14 der Steuerungsdatei eine Ziffer und kein String gefunden. In der entsprechenden Zeile muss der Parametertyp zu String geändert werden. Oder es wurde eine größere Anzahl bei `numberOfDifferentStimuli` angegeben als danach Stimulus-Zeilen folgen.
  - Found a string while looking for an enumerated value in text (line 4). ßtimuliAreSounds" not read. ExperimentMFC not read. Data not read from text file "C:\Experiment\test.txt". File "C:\Experiment\test.txt" not finished.
    → Es wurde in Zeile 4 der Steuerungsdatei ein String und kein Schlüsselwort gefunden. In der entsprechenden Zeile (hier sogar mit "stimuliAreSounds" angegeben) muss der Parametertyp zu Schlüsselwort geändert werden.
  - Signed integer not read from text file. "numberOfReplicationsPerStimulus" not read. ExperimentMFC not read. Data not read from text file

"C:\Experiment\test.txt". File "C:\Experiment\test.txt" not finished.
→ Der Wert von `numberOfReplicationsPerStimulus` konnte nicht gelesen werden, weil unter `numberOfDifferentStimuli` eine geringere Anzahl angegeben ist, als Stimuli-Zeilen folgen.

  - weitere Kombinationen analog...

- Abtastrate nicht gleich:
  - The sound in file "C:\Experiment\sounds\beep.wav"has different sampling frequency than some other sound. ExperimentMFC "test" not started. Experiment window not created. Command "Run"not executed.
    → Abtastraten der verschiedenen Stimuli sind nicht gleich und müssen angeglichen werden (vgl. 3.1).
- Dateiendungsprobleme:
  - Cannot open file "C:\Experiment\sounds\beep.mp3.wav". ExperimentMFC "test" not started. Experiment window not created. Command "Run"not executed.
    → Dateitypen der verschiedenen Stimuli sind verschiedenen und müssen ggf. konvertiert werden. Oder es wurden falsche Angaben innerhalb der Steuerungsdatei gemacht: Ähnlicher Fehler z.B. beep.wav.wav, weil irgendwo die Dateiendung des Stimulus angegeben wurde, obwohl die Dateiendung nur unter `stimulusFileNameTail` angegeben werden darf.
- Fatale Fehler und Laufzeitfehler:
  - Praat will crash. Notify the author (paul.boersma@uva.nl) with the following information: (Melder_malloc_f:) Can never allocate 0 bytes.
    Runtime Error! Program: C:\Users\Programs \Praat.exe
    This application has requested the Runtime to terminate it in an unusual way. Please contact the application's support team for more information.
    → Ein schweres Problem ist aufgetreten und Praat stürzt ab. Taucht z.B. auf, wenn `responsesAreSounds` und `stimuliAreSounds` widersprüchlich zueinander definiert sind. Es kann auch daran liegen, dass `responsesAreSounds` mit <yes> angegeben wurde und dann die folgenden Parameter fehlen. Oder dass `stimuliAreSounds` mit <no> angegeben wurde und die folgenden Parameter zu viel sind. Falls die Überprüfung dieser Werte nichts ergibt, am besten die aktuelle Steuerungsdatei verwerfen und eine andere, funktionierende Steuerungsdatei-Vorlage öffnen und mit dieser arbeiten.
- Sonstige Nachrichten:
  - There are zero trials in this experiment.
    → Bei `numberOfReplicationsPerStimulus` wurde 0 als Wert angegeben, weshalb es keine Antwortfolien gibt.

- Buttons und Felder:
  - Die Buttons bzw. Antwortfelder werden an der falschen Stelle, zu klein, zu groß oder gar nicht angezeigt.
    → Die Koordinaten der Buttons bzw. Antwortfelder wurden falsch angegeben. Eventuell kann auch die Schriftgröße der Beschriftung angepasst werden. Vergleiche mit Abschnitt 3.3.3.

ISSN: 2364-2459