

The Perception of Lexical Stress in German: Effects of Segmental Duration and Vowel Quality in Different Prosodic Patterns

Klaus J. Kohler

University of Kiel, Kiel, Germany

Abstract

Several decades of research, focusing on English, Dutch and German, have set up a hierarchy of acoustic properties for cueing lexical stress. It attributes the strongest cue to criterial-level f_0 change, followed by duration, but low weight to energy and to stressed-vowel spectra. This paper re-examines the established view with new data from German. In the natural productions of the German word pair *Kaffee* 'coffee' – *Café* 'locality' (with initial vs. final stress in a North German pronunciation), vowel duration was manipulated in a complementary fashion across the two syllables in five steps, spanning the continuum from initial to final stress on each word. The two base words provided different vowel qualities as the second variable, the intervocalic fricative was varied in two values, long and short, taken from *Café* and *Kaffee*, and the generated test words were inserted in a low f_0 tail and in a high f_0 hat-pattern plateau, which both eliminated f_0 change as a cue to lexical stress. The sentence stimuli were judged in two listening experiments by 16 listeners in each as to whether the first or the second syllable of the test word was stressed. The results show highly significant effects of vowel duration, vowel quality and fricative duration. The combined vowel-quality and fricative variable can outweigh vowel duration as a cue to lexical stress. The effect of the prosodic frame is only marginal, especially related to a rhythmic factor. The paper concludes that there is no general hierarchy with a fixed ranking of the variables traditionally adduced to signal lexical stress. Every prosodic embedding of segmental sequences defines the hierarchy afresh.

Copyright © 2012 S. Karger AG, Basel

1. The Research Question

Several decades of experimental research, since the seminal papers by Fry [1955, 1958, 1965], have established that lexical stress is signalled by f_0 change above a critical threshold [Fry, 1958 for English; Isačenko and Schädlich, 1966 for German; Sluijter and van Heuven, 1995 for Dutch], by syllable duration, specifically in the vowel nucleus [Fry, 1955 for English; Sluijter et al., 1997 for Dutch], but also in consonants,

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
0031–8388/12/0692–0068
\$38.00/0
Accessible online at:
www.karger.com/pho

Prof. Dr. Klaus J. Kohler
Holm 4
DE–24113 Molfsee
(Germany)
Tel. +49 43 165 1453, E-Mail kjk@ipds.uni-kiel.de

particularly initial ones [Huggins, 1972 for English], by spectral expansion in vowels [Fry, 1965 for English; van Bergem, 1993; van Heuven and de Jonge, 2011 for Dutch], and by greater acoustic energy [Fry, 1955 for English; Sluijter et al., 1997 for Dutch]. F0 change has been found to carry the greatest weight when it simultaneously signals sentence accent on the highlighted word [Sluijter and van Heuven, 1996], which in languages with lexical stress is normally manifested on the stressed syllable. Next in the factor hierarchy ranges duration. Acoustic energy has been allocated a minor role, and its influence has not been investigated thoroughly. Spectral expansion has been shown to have an influence, albeit much weaker, especially in languages that have strong centralization towards schwa in unstressed syllables, such as English, though other stress languages also have more peripheral vowels in stressed than in unstressed syllables, e.g. Dutch [Koopmans-van Beinum, 1980] or German [Mosshammer and Geng, 2008].

Given the multiparametric manifestation, spanning suprasegmental stretches of speech, lexical stress needs to be investigated as a dynamic pattern across at least two syllables, e.g. stressed-unstressed or unstressed-stressed disyllable words, or monosyllables followed or preceded by an unstressed syllable in another word, such as, for example, a function word. In such a sequence, the whole production mechanism is geared towards giving one part of the dynamic pattern, beginning or ending, more perceptual prominence over the other. This entails more versus less extensive opening and closing vocal-tract movements into and out of the syllable nuclei in the two parts of the dynamic pattern, resulting in more versus less peripheral and longer versus shorter vowels, but also affecting, to a greater or lesser extent, the degree and length of the surrounding consonantal strictures that are linked to one or the other part of the dynamic pattern. Thus in German '*umlagern*' 'shift to another place', '*umfahren*' 'run over' versus *um*'*lagern*' 'surround', *um*'*fahren*' 'drive around' both the vowel /ʊ/ and the consonant /m/ are longer in the stressed syllable, creating a different weighting of /ʊm/ versus /la:/ or /fa:/ [Kohler, 1991, p. 165].

But it is not only the vocal-tract articulation that is strengthened or weakened in such a dynamic pattern, the laryngeal source is equally adjusted, leading to greater versus smaller amplitude and frequency of the glottal vibration, especially when coupled with sentence accentuation, and to greater versus smaller airflow with an open glottis in voiceless obstruents. This accounts for the influence of f0 and acoustic energy in stressed syllables, and also for the effect of energy in fricative noise in conjunction with degree of stricture. The manifestation of voiceless obstruents in phrasally accented or unaccented syllables was investigated by Butcher [1977] in the German utterances *das schaff* [f] *ich doch nicht*, *das lass* [s] *ich doch nicht*, *das mach* [x] *ich doch nicht*, *das schnapp* [p] *ich doch nicht*, *das hatt* [t] *ich doch nicht*, *das pack* [k] *ich doch nicht*, where accent was either placed on the verb containing the obstruent or on the pronoun *ich* or on *doch*. The obstruents were shortest and showed the greatest tendency to becoming voiced when the accented syllable was farthest removed. Thus, the manifestation of an intervocalic obstruent depends on its integration in an accented syllable, where it is strong, or an unaccented one, where it is weak.

As for the exponency of lexical stress, intervocalic fricatives linked to unstressed syllables before a stressed one may also become short and weak enough for the supra/subglottal pressure differential to allow voicing to continue across them. Thus, in German, *Hannover* [ha'no:ʁə] with [f] following the stressed vowel is differentiated from *Hannoveraner* [hanovə'ra:nə] with [v], which is embedded in unstressed

syllables on either side [Duden Aussprachewörterbuch, 1974; WDA, 1969]. This is also the essence of Verner's Law [Verner, 1876], which states that in the Germanic consonant shift of intervocalic voiceless plosives to fricatives, voicelessness is kept when it is linked to the position immediately following the lexically stressed vowel, otherwise voicing occurs.

The dynamic syllable patterns for the signalling of lexical stress are inserted in larger prosodic frames determined by f_0 , energy and segmental timing profiles across utterances. This means that the manifestation of lexical stress is adjusted to its prosodic frame, which in turn depends on the communicative function of speech interaction in specific situations. Any of the parameters that have the potential of coding lexical stress may be weighted at the expense of some or all of the others. Thus, in focal pitch accentuation of words, f_0 movement carries the greatest weight, duration and spectrum of segments in the dynamic pattern become secondary. In post-nuclear low tails, f_0 movement is levelled and segmental attributes take over, their durations in the first place, but also their spectral make-up. The same can be expected in the top declination of hat patterns. In emphatic intensification, stressed-syllable initial consonants are strengthened in duration, energy and spectrum in addition to extensive f_0 movement [Niebuhr, 2010]. Such a dynamic perspective on lexical stress therefore precludes the search for a constant hierarchy of lexical stress parameters across all prosodic environments. At the same time, it makes it mandatory to base experimental investigations into lexical stress on contrastive dynamic syllable sequences in prosodic embedding of plausible natural utterances.

Kohler [1990] investigated the relative contribution of f_0 pattern and syllable duration to prefix or stem stress perception in the German lexical items '*umlagern* versus *um'lagern*, in the sentence context '*Er wird's wohl ____.*' ('He will probably shift/surround it.') with a late f_0 peak accent on '*umlagern* and an early f_0 peak accent on *um'lagern*. In the natural productions of the two sentences, used for stimulus generation in a perception experiment, the durations of the stressed and unstressed prefixes were 222 and 135 ms, respectively, while the stressed and unstressed *la* syllables measured 268 and 258 ms; the f_0 peak maxima occurred at practically the same position just before the onset of the stem [l], but the shapes of the peak contours were different, a more gradual descent in the early-peak stem accent. From these base stimuli two f_0 peak series were generated: (1) by shifting the late f_0 peak contour of the prefix-stress sentence to the right in six and to the left in five 30-ms steps, (2) by shifting the early peak contour of the stem-stress sentence to the left in eight 30-ms steps. Subjects' judgements reported one or the other lexical meaning. In (1) there is a clear change from initial to stem stress as the peak position moves from left to right, in spite of the duration of *um* pointing to the former. f_0 can thus override duration. In (2) there is no change between the stress categories; the answers remain predominantly in favour of stem stress. In this case, f_0 cannot override the duration cue because *um* is too short in relation to *la* to signal initial stress.

These data show that an f_0 peak contour can cue stress position by itself provided the syllable duration is not too short in relation to the neighbouring syllable. This in turn means that f_0 and duration do not occupy fixed positions in a stress cue hierarchy, and the same will apply to the relationship between duration and sound spectrum, once the strong f_0 cue has been eliminated. The latter effects are thus best analysed in f_0 patterns that exhibit very little movement, such as in the final low tail after a tonic fall, or in the gradually descending plateau of a hat pattern, which both exclude the cue of

f₀ change. If in these f₀ frames the acoustic energy of vowels is also equalized in word pairs that are only differentiated by stress position, segmental duration and vowel spectrum will become the main acoustic cues.

In a recent study of spectral and durational cue strength in Dutch, van Heuven and de Jonge [2011] excluded f₀ change from the test word by putting it in the post-accentual low-pitch tail. By systematically varying vowel duration and spectral composition (expansion/reduction) in a dual-parameter 7 × 7 design of a single minimal stress pair *canon* /'ka:nɔn/ 'canon, round song' ~ *kanon* /ka'nɔn/ 'cannon', they confirmed that temporal organisation is a strong cue to stress perception when the target word is without a pitch accent, while spectral expansion/reduction is very weak. Its effect was noticeable only when temporal structure was ambiguous. The authors conclude that spectral expansion/reduction is the weakest of the four traditional stress cues, at least in the West Germanic stress-accent languages Dutch and English.

This paper takes the Dutch results of van Heuven and de Jonge [2011] as its point of departure and applies their research question to German. It deals with a contrastive stressed-unstressed and unstressed-stressed German dynamic syllable pair of otherwise identical segmental types – *Kaffee* 'coffee' and *Ca'fé* (the locality), and places it in two prosodic frames, a low tail and a high hat-pattern plateau. The words are incorporated in two sentence slots:

(i) *Wir treffen uns 'regelmäßig beim 'Kaffee/Ca'fé dort an der 'Ecke.'* ('We regularly meet for coffee/at the café on the corner.') with focal accent on *regelmäßig* and partial deaccentuation [Kohler, 2009] of *Kaffee/Ca'fé*, *Ecke* in the low-tail condition; (ii) *Wir treffen uns 'regelmäßig beim 'Kaffee/Ca'fé dort an der 'Ecke.'* ('We regularly meet for coffee/at the café on the corner.') with evenly distributed accents across a high f₀ plateau spanning a hat pattern over *regelmäßig beim 'Kaffee/Ca'fé dort an der* followed by the nuclear fall on *'Ecke*. In the naturally produced utterances, the vowels in the stressed syllables are longer and more peripheral than the corresponding vowels in the unstressed syllables, and the single intervocalic consonant is longer and stronger before than after the stressed vowel. The ensuing investigation tests the following hypotheses:

Hypothesis 1: Effect of Vowel Duration. As German disyllables with initial or final stress in the same phonological segment sequence are produced with different vowel duration structures across the two stress patterns, a stepwise complementary disyllabic change of vowel duration within the range of the naturally produced stress pair effects a change in lexical stress perception.

Hypothesis 2: Effect of Vowel Quality. As different vowel qualities are produced in stressed and unstressed positions of such stress pairs, they have a further effect on lexical stress perception. Thus the complementary disyllabic vowel duration changes along the same scale on both items of the stress pair create different stress perception profiles.

Hypothesis 3: Effect of Intervocalic C Duration. As the duration of an intervocalic consonant in such stress pairs differs, depending on whether the unstressed or the stressed vowel follows, this also has an effect on lexical stress perception. Thus the complementary disyllabic vowel duration changes in items with a long or a short consonant create different stress perception profiles.

Hypothesis 4: Local Effects Are Additive. The effects of the disyllabic vowel duration and vowel quality patterns, and of the intervocalic consonant duration are additive, i.e. across the vowel duration scale, initial-stress quality + short consonant will yield the highest number of initial stress responses, final-stress quality + long consonant the

highest number of final stress responses; initial-stress quality + long consonant will reduce the number of initial stress responses and final-stress quality + short consonant will reduce the number of final stress responses.

Hypothesis 5: Global Prosody Effects. Although focal prosody on the test word pair is excluded by insertion in low tails and high plateaux, the two different prosodic frames may still trigger different vowel qualities and durations in production, and thus also have an effect on lexical stress perception.

Varying vowel duration systematically from equal to long-short_degree1, long-short_degree2 and short-long_degree1, short-long_degree2 in both original disyllables and in both prosodic patterns, creates four 5-point dyadic vowel duration series, one from original *Kaffee* and one from original *Café* in each of the two prosodic frames. This experimental design can test whether there are different profiles of initial and final stress perception for the duration series imposed on the two disyllables, which, due to their different dynamic patterns, in addition have differing vowel qualities. Supplementing the four series by another four in which each original [f] is replaced by its counterpart in the other word tests the influence of the consonant following or preceding the stressed vowel. So, the aim of the paper is to quantify the interdependence between duration and vowel quality as well as intervocalic fricative duration in two different prosodic patterns, which both eliminate macroprosodic f0 movements across the disyllables. Two experiments were run. The results of experiment 1 suggested an additional rhythmic effect, which was tested in experiment 2 with a new set of stimuli in the same overall paradigm under hypothesis 5a.

Hypothesis 5a: Rhythmic Effect. Different rhythmicities across the test word and the following frame section bias stimuli one way or the other. Audio illustrations of test stimuli, linked to graphic displays, are made available in wav files as online supplements at www.karger.com/doi/10.1159/000342126. They are indexed as *Audio n* + the test parameters.

2. Experiment 1: Testing for Local Segmental and Global f0 Effects

2.1. Stimulus Generation

The author recorded the 4 sentences several times. One token of each type was selected such that the f0 fluctuations in the low tail and the high plateau were smallest. The words *Kaffee* and *Café* were excerpted from the frames. As the vowels in *Café* had equal energy maxima in both vowels, the lower energy in the second vowel of *Kaffee* was raised by 4 dB in the tail and by 2 dB in the high plateau frame to equalize. The f0 courses were then levelled to auditory pitch equality in the two syllables by using PSOLA manipulation in Praat.

Figure 1a, b shows the adjusted high-plateau *Kaffee* and *Café* excerpts reinserted in the original high-plateau *Kaffee* frame. In the complete utterance we see the declining plateau of the hat pattern preceded by the high rise on *regelmäßig* and followed by the low fall on *Ecke*. Across the two syllables of both words there is a levelled f0 pattern around 129 Hz, and the energy contours are also comparable.

In parallel, figure 2a, b shows the adjusted low-tail *Kaffee* and *Café* excerpts reinserted in the original low-tail *Kaffee* frame. In the remaining utterance after the focused *regelmäßig*, there is a low f0 tail that fluctuates around 95 Hz, the fluctuation being in

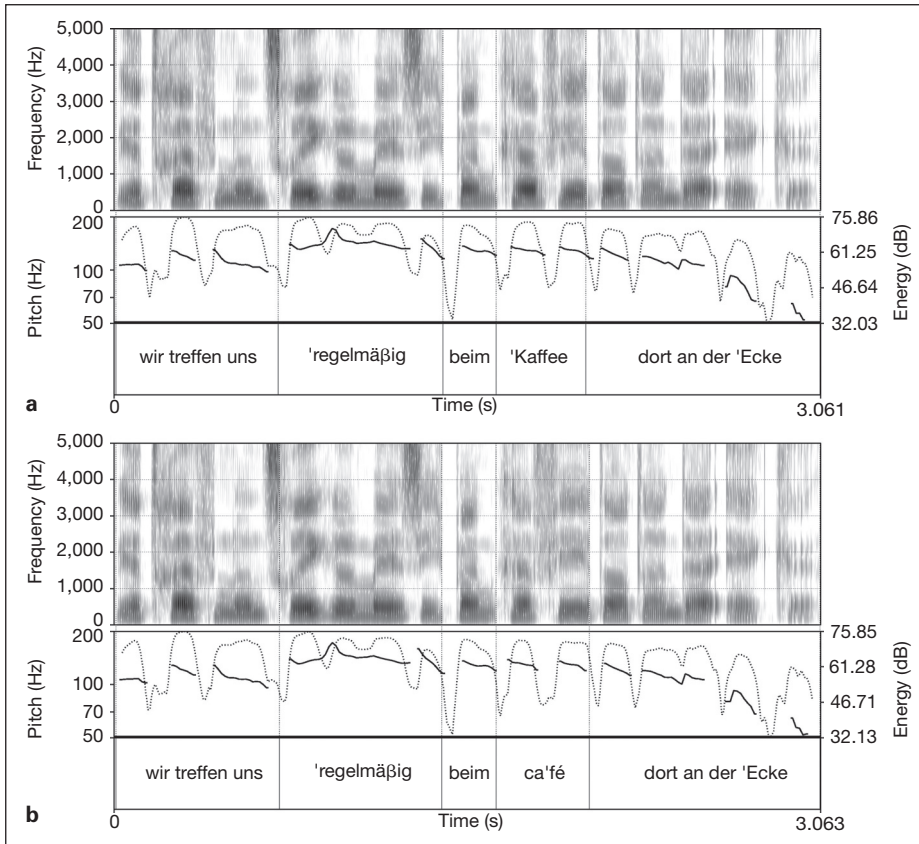


Fig. 1. Spectrogram, f_0 (plain; Hz along logarithmic scale) and energy (dotted; dB) contours across the original high-plateau *Kaffee* utterance, with the modified words *Kaffee* (**a**) and *Café* (**b**) spliced into *Wir treffen uns 'regelmäßig beim ___ dort an der 'Ecke*. Sections orthographically labelled in separate window. (*Audio 1* high-plateau frame, *Kaffee*; *Audio 2* high-plateau frame, *Café*.)

the main due to microprosodic disturbances, and to the final drop on the last syllable. The f_0 patterns across the two syllables of both words are within these microprosodic margins. The energy contours are also comparable.

Table 1 lists the durations of the vowels and the fricative, as well as the formant frequencies of /a/ and /e/ at mid-vowel position, in the two words in each original utterance. The /k/ release + aspiration has very similar durations in all 4 words (*Kaffee* low 53 ms, *Café* low 54 ms, *Kaffee* high 62 ms, *Café* high 65 ms), but their spectra are different, so it is the whole syllable /ka/ that differs across both words and both sentence frames.

The stressed vowels are longer than the unstressed ones in each word. The difference is greater for /a/ than for /e/, i.e. the pre-stress position leads to greater reduction than the post-stress position. The vowel durations in *Kaffee* are almost equal. In *Café*, the difference is about 20 ms in the low-tail frame, 40 ms in the high-plateau frame. From approximately equal vowel durations in the two syllables of *Kaffee* in the low-tail frame and in the high-plateau frame, the durations in *Café* are down by 22 ms or

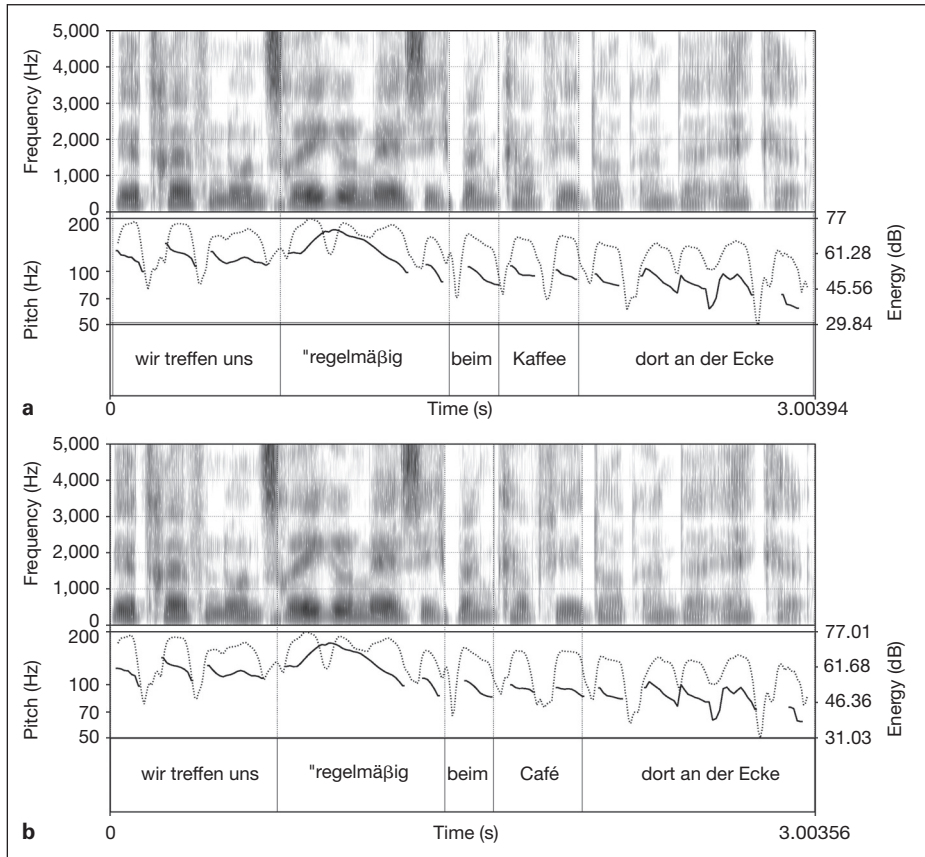


Fig. 2. Spectrogram, f_0 (plain; Hz along logarithmic scale) and energy (dotted; dB) contours across the original low-tail *Kaffee* utterance, with the modified words *Kaffee* (**a**) and *Café* (**b**) spliced into *Wir treffen uns* "regelmäßig beim ___ dort an der Ecke". Sections orthographically labelled in separate window. (*Audio 3* low-tail frame, *Kaffee*; *Audio 4* low-tail frame, *Café*.)

Table 1. Durations (in milliseconds) and formant frequencies F1, F2, F3 (in Hertz, mid-vowel position) of vowels, and durations of fricatives, in *Kaffee* and *Café* of the low-tail and high-plateau frames

		Low tail				High plateau			
		Kaffee		Café		Kaffee		Café	
		ms	Hz	ms	Hz	ms	Hz	ms	Hz
/a/	dur	110		88		124		96	
	F1		545		482		510		495
	F2		1,504		1,682		1,427		1,528
	F3		3,121		3,316		2,974		3,110
/e/	dur	98		112		126		135	
	F1		383		352		409		357
	F2		1,795		1,920		1,955		1,941
	F3		3,251		3,336		3,270		3,335
/f/	dur	76		104		72		100	

Table 2. Complementary first-second vowel durations long-short_degree2 (L2S), long-short_degree1 (L1S), equal (EQU), short-long_degree1 (S1L), short-long_degree2 (S2L) for both *Kaffee* and *Café* in either the low-tail or the high-plateau frame

	L2S	L1S	EQU	S1L	S2L
Low-tail	120-80	110-90	100-100	90-110	80-120
High-plateau	135-95	125-105	115-115	105-125	95-135

28 ms in unstressed position and up by 14 or 9 ms in stressed position in the low-tail or the high-plateau frame. The fricative is 28 ms longer preceding the stressed vowel in both utterance frames.

In both utterance frames, formants F1 and F2/F3 are further apart in the second vowel of *Café* than of *Kaffee*, indicating a closer front vowel quality in the stressed position, but in the low-tail frame, the formants of unstressed /e/ deviate more from a close vowel. Formants F1 and F2/F3 are closer together in the first vowel of *Kaffee* versus *Café* pointing to greater opening in the stressed position. In the high-plateau frame, /a/ in *Café* is more centralized than in *Kaffee*, and than in *Café* of the low-tail frame; it sounds more like [ə] and has a formant structure accordingly.

It may be argued that this interaction between prosodic frame and vowel quality reduction is unexpected, and so far only relates to a single token by a single speaker. It is true that these data have not been confirmed as a systematic phenomenon of speech production in German, and they may be idiosyncratic. But they no longer come as a surprise when speech analysis gives up the dichotomous approach to sounds and prosodies and looks at their constant interaction in speech communication. High-plateau and low-tail f0 can easily go together with vowel-fronting/raising and backing/lowering, respectively, as was demonstrated for the two forms of the German greeting *tschüss* ‘bye’ [tʃy:s] on a high f0 level and [tʃʏs] on a low one, or by fronted [ɥ] on rising f0 in *ja gut* versus back [u] on low falling f0 in *na gut* in the Kiel Corpus of Spontaneous Speech [Kohler, 2008a]. This prosody-segment interplay needs further detailed investigation, see also Niebuhr [2009a] and Li and Kong [2012]. But even if the stimulus production for this experiment were idiosyncratic, the present research question is whether listeners can judge the stimuli systematically with regard to lexical stress, which does not presuppose a systematic production study of these (potentially idiosyncratic) acoustic properties.

The duration patterns in the original 2 × 2 target word productions suggest that a systematic variation of vowel duration for shifting stress from the first to the second vowel needs to involve both syllables simultaneously in a complementary fashion, and that these changes span a duration range at a higher level in the high-plateau frame. Therefore, in each word in both prosodic patterns disyllabic vowel duration was manipulated in Praat along a 5-step scale of decreasing first and increasing second vowel duration: long-short_degree2 (L2S), long-short_degree1 (L1S), equal (EQU) short-long_degree1 (S1L), short-long_degree2 (S2L). This created four duration series, one from original *Kaffee* and one from original *Café* in each of the two prosodic frames. Table 2 gives the resulting vowel duration scales for the word pair in either prosodic frame. The 5 × 4 duration and f0 manipulations of the word excerpts from the original utterances were then PSOLA resynthesized in Praat.

In each of the 20 resynthesized sound files, the /f/ section was removed and replaced by the /f/ section from the original *Kaffee* and from the original *Café*, in the low and the high frame, respectively, thus generating $2 \times 5 \times 4 = 40$ sound files. The 20 low-pitched and the 20 high-pitched word files were then spliced into the low-tail and the high-plateau, respectively. The sentence frame was in either case taken from the original utterance containing *Kaffee*. The resulting stimuli sounded fluent and completely natural without any technical artefact.

2.2. Test Files

The stimuli were indexed according to four parameters:

- **f0 Frame** *high, low* (**Fh, Fl**).
- **Vowel quality dyad** *Word Kaffee, Café* (**Wk, Wc**).
- **Vowel duration dyad** *long-short degree2, long-short degree1, equal, short-long degree1, short-long degree2* (**L2S, L1S, EQU, S1L, S2L**).
- **Fricative Consonant** *long, short* (**Cl, Cs**).

In the stimulus name, the parameter labels are given in the above order, separated by `_`, e.g. `Fh_Wc_EQU_Cl`. The labels L2S, L1S, EQU, S1L, S2L reflect the concomitant descending and ascending scales of first and second vowel durations, going through equal.

Two test files were created for two perception tests in one test session with the Kiel reaction measuring instrumentation for recording responses and reaction times. All the generated stimuli were replicated 5 times, and formatted with beeps and pauses. The length of the answering window was 4 s. The 5×20 high-plateau stimuli were randomized in test Fh, the 5×20 low-tail stimuli in test Fl. The sequence of test Fh and test Fl was changed for different subject groups to get a balanced test paradigm.

2.3. Subjects and Tests

Sixteen students of linguistics and languages (3 male, 13 female, 1 male 66 years, 1 female 33 years, the rest 21–26 years) did the listening test in four groups of (a) 5, (b) 4, (c) 5 and (d) 2, respectively. Groups (a) and (d) did the listening session in the order Fh-Fl, groups (b) and (c) in the reverse order. The subjects were given simultaneous written and oral instructions. These instructions provided illustrations of distinguishing words by different stress placement. The word pair *Kaffee* versus *Café* was introduced as a minimal pair with first-syllable and second-syllable stress, respectively, to differentiate the drink from the locality in the northern variety of German. Since it could not be ruled out altogether that subjects might also use final stress for the drink, subjects were told that they should decide whether they heard the word in the utterance with stress on the first or the second syllable. They were to give their answers by pressing either button 1 or button 2 of the response boxes in front of them, and to react as quickly as possible following perception. They were told that there might be uncertain cases, but that they should still come to a quick decision. The instructions were followed by a practice run of responding to 10 stimuli, 5 from Fl, then 5 from Fh. The test session took 15 min for each test and about 10 min for the instructions, i.e. a total of approximately 45 min.

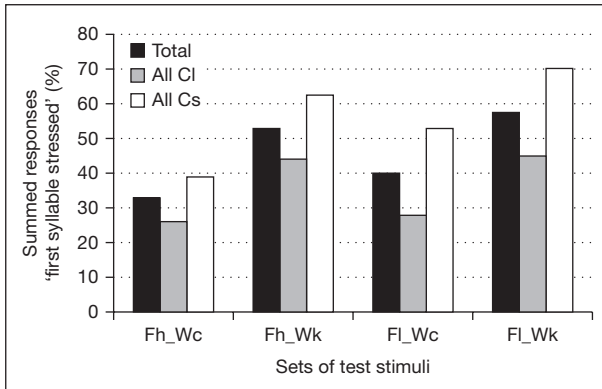


Fig. 3. Relative frequencies of ‘first syllable stressed’ responses in the 4 duration sets Fh_Wc, Fh_Wk, Fl_Wc, Fl_Wk, grand total and totals for long and short /f/; n = 800 in the total set and n = 400 in the Cl and Cs sets.

2.4. Results

There were 17 missing responses in the high-plateau and 6 in the low-tail frame, in each case out of 16×100 , i.e. $\leq 1\%$. Missed responses either mean that subjects reacted too fast, before the opening of the response window of the reaction hardware and ahead of the target stimulus, or too slowly, after the closing of the 4-second window and thus well outside a spontaneous response span. The low occurrence of this means that subjects had no problem with the task. Since the misses are spread randomly across stimuli and subjects, they are treated as non-1 responses. The RT measurements were also analysed but are not presented in this paper because they do not add any further information beyond the category judgements.

2.4.1. Averaged Responses in the Fh and Fl Tests

Figure 3 provides the relative frequencies of ‘first syllable stressed’ responses in the 4 duration sets Fh_Wc, Fh_Wk, Fl_Wc, Fl_Wk, in each case giving the grand total as well as the totals for long and short /f/. The emerging picture is clear. In both the high-plateau and the low-tail frame, there are fewer ‘first syllable stressed’ responses for the Wc as against the Wk stimuli – base *Café* versus base *Kaffee*, and for the Cl as against the Cs stimuli – long versus short /f/. There are also fewer ‘first syllable stressed’ responses in the high-plateau compared with the low-tail for the grand totals of the Wc and the Wk set; the difference is small in the Cs totals and absent from the Cl totals. This shows strong effects of vowel quality and /f/ duration, and a weak effect of prosodic frame for Wc_Cs and Wk_Cs stimuli.

2.4.2. Responses to the Test Stimuli in the Fh and the Fl Test

Figure 4 breaks the response data down into the two test frames Fh and Fl, and the two vowel quality dyads Wc and Wk in each, along the 5-point vowel duration scale, separating the fricative durations Cl and Cs at each point. The graph shows that the two vowel quality dyads form distinct sets in both frames, point by point along the scale for corresponding fricative durations. Figure 5 breaks the data down further into the 4 combinations of vowel quality and fricative duration Wk_Cs, Wk_Cl, Wc_Cs, Wc_Cl. The following observations can be derived from this graph:

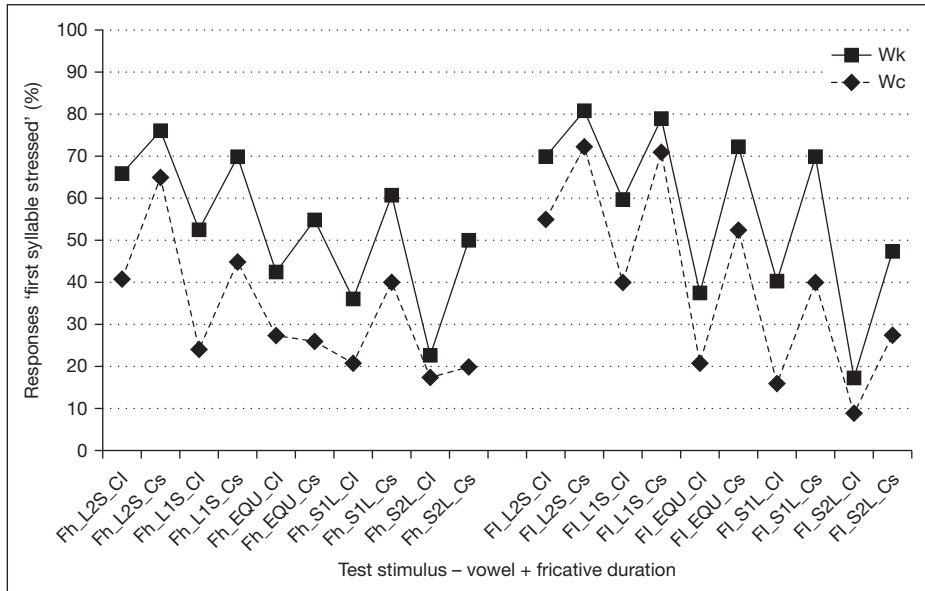


Fig. 4. Relative frequencies of ‘first syllable stressed’ responses to each of the 20 test stimuli in the high-plateau (Fh, left) and the low-tail (Fl, right). Plain/dotted lines connect responses to Wk/Wc for Cl and Cs along the 5-point duration scale L2S to S2L. Each mean is based on 80 responses.

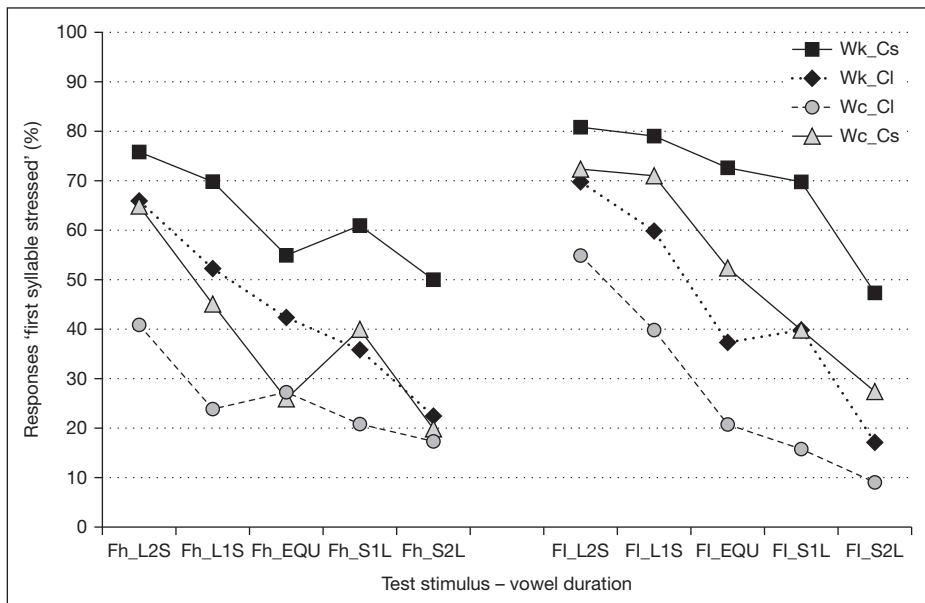


Fig. 5. Relative frequencies of ‘first syllable stressed’ responses to each of the 20 test stimuli in the high-plateau (Fh, left) and the low-tail (Fl, right), along the 5-point duration scale L2S to S2L in the 4 sets Wk_Cs, Wk_Cl, Wc_Cs, Wc_Cl (Fh: *Audios 5-8*; Fl: *Audios 9-12*). Each mean is based on 80 responses.

(1.1) In both frames, the stimuli Wk_Cs (original *Kaffee* with short /f/) show decreasing initial-stress responses along the duration scale from L2S to S2L, but the relative frequency for S2L stays around 50%, i.e. there is no change of lexical stress category from L2S to S2L.

(1.2) For all the corresponding stimuli with the other vowel dyad quality (original *Café* with short /f/: Wc_Cs), there are fewer initial-stress judgements in both frames, now resulting in a lexical stress category change from L2S to S2L.

(2.1) In both frames, the responses to stimuli Wc_Cl (original *Café* with long /f/) along the duration scale stay below or around 50% 'initial stress', i.e. there is no category change in the opposite direction from S2L to L2S.

(2.2) For all the corresponding stimuli with the other vowel dyad quality (original *Kaffee* with long /f/: Wk_Cl), there are more initial-stress judgements in both frames, now resulting in a lexical stress category change from S2L to L2S.

Comparing corresponding stimuli in the low-tail and the high-plateau frame, the following regularities of initial-stress responses are found:

(3.1) The relative frequencies for both Fl_Wk_Cs and Fh_Wk_Cs range from just under 50% to around 80%, i.e. when the vowel qualities (Wk) and the fricative (Cs) point towards initial stress, there is a general bias to initial-stress responses across the whole duration scale in both frames.

(3.2) The relative frequencies for Wk_Cl range from around 20% to around 70%, irrespective of the prosodic frame.

(3.3) The relative frequencies for Fl_Wc_Cs are substantially higher than for Fh_Wc_Cs at vowel durations L1S and EQU, i.e. when the vowel qualities (Wc) and the fricative (Cs) work against each other the low-tail frame boosts initial stress responses at the less decisive vowel durations.

(3.4) At the left end of the duration scale, the relative frequencies are higher for the 2 Fl_Wc_Cl than for the 2 Fh_Wc_Cl stimuli, i.e. when the vowel qualities (Wc) and the fricative (Cl) point towards final stress, but the vowel durations point towards initial stress, there is a bias to initial-stress responses in the low-tail frame.

These descriptive statistics suggest that complementary vowel duration patterns are not sufficient to differentiate the German minimal stress pair '*Kaffee/Ca'fé*' in levelled f0. Vowel quality in the two syllables and the intervocalic fricative do not just determine judgement at vowel durations that are indecisive as to one or the other category, but even at the polar ends of the duration scale. Their cue value is thus on a par with duration. Frame has a minor effect on stress perception. Differences between the two sets concern the boosting of initial-stress responses in the low-tail frame on two conditions: (a) when the contradictory vowel-quality/fricative cue to final stress (Wc_Cs) is combined with a less decisive duration cue to initial-stress (L1S, EQU), (b) when the strongest vowel-quality/fricative cue to final stress (Wc_Cl) clashes with a strong duration cue for initial stress (L2S, L1S).

2.4.3. Inferential Statistics

To test the significance of these observations, a repeated measures (RM) ANOVA was applied to the response data with the 4 factors Frame (2 levels high/low, Fh/Fl), Word (2 levels (original) *Café/Kaffee*, Wc/Wk), Vowel Duration (5 levels, L2S–S2L), and Fricative (2 levels long/schort, Cl/Cs). This type of inferential statistics presupposes interval-scaled stress judgements that have normal distribution across the 16 subjects. Although the subjects gave binary responses they also judged each

stimulus 5 times, thus yielding a score ranging from 0 to 5 for one stress position per stimulus and speaker. These scores can be treated as interval-scaled measures. Normal distribution was tested by the Shapiro-Wilk test. With test value $T' = 0.978 > \text{critical value } T(16; 0.05) = 0.887$, compatibility with a normal distribution cannot be rejected.

The within-subject factors Word, Vowel Duration and Fricative yield highly significant main effects: Word [$F(1, 15) = 36.290, p < 0.001, \eta_p^2 = 0.708$]; Vowel Duration [$F(4, 60) = 48.312, p < 0.001, \eta_p^2 = 0.763$]; Fricative [$F(1, 15) = 55.886, p < 0.001, \eta_p^2 = 0.788$]. These significances are linked to considerable and very similar effect sizes in terms of partial eta-squared, i.e. dyadic vowel quality and duration of intervocalic /f/ have equally significant effects as dyadic vowel duration. Frame misses significance: [$F(1, 15) = 4.350, p = 0.054$], and also has a much lower effect size [$\eta_p^2 = 0.225$]. Of the interactions only 3 reach significance with even smaller effect sizes: Frame \times Vowel_Duration [$F(4, 60) = 2.719, p = 0.038, \eta_p^2 = 0.153$]; Frame \times Vowel_Duration \times Fricative [$F(4, 60) = 2.950, p = 0.027, \eta_p^2 = 0.164$]; Word \times Vowel_Duration \times Fricative [$F(4, 60) = 2.975, p = 0.026, \eta_p^2 = 0.165$].

Since ANOVA is a linear statistical model based on additive variances, significant interactions indicate that strict linearity does not hold and that significant main effects cannot be assumed to be additive *prima facie* but need to be re-evaluated. The low effect sizes of the interactions, involving Word and Fricative with Vowel Duration, compared with the high and very similar main-effect sizes, point to a strong additive tendency of the factors Vowel Duration, Word and Fricative. To assess their relative contributions a 2-factor analysis was run on the data, with Vowel Duration and WordFric as variables. Separately for each of the two frames, WordFric combines the variables Word and Fricative to form a property scale from extreme *Kaffee*-ness in Wk-Ws to extreme *Café*-ness in Wc-Cl. Within-WordFric contrasts are then calculated on the following hypotheses:

- In Wk-Cs versus Wc-Cl, where both the vowel quality and the fricative converge either for initial or final stress, the strongest contrast effect is expected.
- In Wk-Cl versus Wc-Cs, where vowel quality and fricative work against each other in opposite directions, no contrast effect is expected.
- In Wk-Cs versus Wk-Cl and Wc-Cl versus Wc-Cs, where vowel quality and fricative converge for initial or final stress in one data set and diverge in the other, intermediate contrast effects are expected.

Table 3 lists the results of the WordFric \times Vowel Duration RM-ANOVAs and the within-subjects contrasts for the high-plateau (Fh) and the low-tail (Fl) data subsets. In both the Fh and the Fl frame the factors WordFric and Vowel are highly significant with large effect sizes, whereas their interactions are not significant and have very small effect sizes. The within-WordFric contrasts along the scale Wk-Cs, Wk-Cl, Wc-Cs, Wc-Cl are highly significant for levels 1_4, 1_2 and 3_4, but are not significant for levels 2_3, and the effect sizes decrease from 1_4 to 2_3 and then rise to 3_4.

A further 2-factor RM-ANOVA was run on the data, splitting both the Word and the Fricative factor into their two levels. It tests simple effects for Frame and Vowel Duration in Wc-Cl, Wc-Cs, Wk-Cs, Wk-Cl to assess whether the prosodic frame influences the response profiles along the duration scale, and whether such an influence differs when vowel quality and fricative work in unison for initial or final stress, or against

Table 3. Results of the WordFric × Vowel Duration RM-ANOVAs for the high-plateau (Fh) and the low-tail (Fl) subsets

Data subset/contrast	Factor/interaction	df1, df2	F	p	Partial η^2
Fh: WordFric_VowDur	WordFric	3, 45	20.748	<0.001	0.580
	VowDur	4, 60	23.155	<0.001	0.607
	WordFric × VowDur	12, 180	1.364	=0.187	0.083
Wk-Cs vs. Wc-Cl	WordFric	1, 15	29.466	<0.001	0.663
Wk-Cs vs. Wk-Cl		1, 15	17.718	=0.001	0.542
Wk-Cl vs. Wc-Cs		1, 15	1.835	=0.196	0.109
Wc-Cs vs. Wc-Cl		1, 15	12.488	=0.003	0.454
Fl: WordFric_VowDur	WordFric	3, 45	28.213	<0.001	0.653
	VowDur	4, 60	31.378	<0.001	0.677
	WordFric × VowDur	12, 180	1.507	=0.125	0.091
Wk-Cs vs. Wc-Cl	WordFric	1, 15	47.543	<0.001	0.760
Wk-Cs vs. Wk-Cl		1, 15	40.584	<0.001	0.730
Wk-Cl vs. Wc-Cs		1, 15	3.184	=0.095	0.175
Wc-Cs vs. Wc-Cl		1, 15	30.861	<0.001	0.673

Dependent is frequency of initial-stress responses.

Table 4. Results of the Frame × Vowel Duration RM-ANOVAs for the Wc-Cl, Wc-Cs, Wk-Cs, and Wk-Cl subsets

Data subset	Factor/interaction	df1, df2	F	p	Partial η^2
Wc-Cl: Fh/l_V	Frame	1, 15	0.184	=0.674	0.012
	VowDur	4, 60	13.398	<0.001	0.472
	Frame × VowDur	4, 60	4.587	=0.003	0.234
Wc-Cs: Fh/l_V	Frame	1, 15	6.908	=0.019	0.315
	VowDur	4, 60	20.998	<0.001	0.583
	Frame × VowDur	4, 60	2.821	=0.033	0.158
Wk-Cs: Fh/l_V	Frame	1, 15	2.655	=0.124	0.150
	VowDur	4, 60	10.951	<0.001	0.422
	Frame × VowDur	4, 60	1.040	=0.394	0.065
Wk-Cl: Fh/l_V	Frame	1, 15	0.071	=0.793	0.005
	VowDur	4, 60	19.025	<0.001	0.559
	Frame × VowDur	4, 60	0.455	=0.768	0.029

Dependent is frequency of initial-stress responses.

each other. Table 4 lists the results of the Frame × Vowel Duration RM-ANOVAs at the split levels Wc-Cl, Wc-Cs, Wk-Cs, and Wk-Cl.

The factor Frame and its interaction with the factor Vowel Duration are not significant for Wk-Cs and Wk-Cl. For Wc-Cs both are significant with quite substantial effect sizes in relation to the effect size of the factor Vowel Duration. Post-hoc two-tailed paired t tests yield significant differences for the central durations L1S [$t(15) = 2.730$, $p = 0.015$], and EQU [$t(15) = 3.158$, $p = 0.006$]. For Wc-Cl the main factor Frame is not significant but its interaction with Vowel Duration is, and a post-hoc two-tailed paired t test yields significance for L1S [$t(15) = 2.784$, $p = 0.0138$].

2.5. Discussion

The results of the RM-ANOVAs show that vowel quality and intervocalic fricative duration have equally large effects as vowel durations on initial or final stress perception in disyllables when f_0 is smoothed on a high plateau or in a low tail. Furthermore, their effects on the response profiles along the vowel duration scales are additive, with Wk-Cs – Wk-Cl – Wc-Cs – Wc-Cl yielding a descending order for initial and an ascending order for final stress. The inferential statistics thus buttresses the descriptive observations (1.1)–(1.2) and (2.1)–(2.2) in 2.4.2.

There is a small effect of frame (a) when the contradictory vowel-quality/fricative cue to final stress (Wc-Cs) is combined with a less decisive duration cue to initial stress, (b) when the strongest vowel-quality/fricative cue to final stress (Wc-Cl) clashes with a strong duration cue for initial stress. In both cases, there are fewer initial-stress responses in the high-plateau frame. This is reflected in the interactions of Frame with Vowel Duration and Fricative in the overall 4-factor ANOVA and in the interactions Frame \times Vowel Duration for Wc-Cs and Wc-Cl in the 2-factor ANOVAs. The descriptive observations (3.3)–(3.4) in 2.4.2 are thus inferentially validated. The reason for such a differential frame effect may be sought in an interplay between the different formant patterns for the word pair *Kaffee/Café* and their different high-plateau versus low-tail prosody embedding. In particular, the [ə]-like exponent of /a/ of base stimulus *Café* in the high-plateau frame would boost the *Café*-ness of the stimulus and favour fewer initial-stress responses in Wc-Cs and Wc-Cl, against their counterparts in the low-tail frame, when this vowel quality is combined with less decisive vowel durations for initial stress.

However, closer inspection of the test stimuli on the strength of the results of experiment 1 revealed that the low-tail beyond the test word did not have even prominence but that a slight f_0 rise-fall on *dort an* made *dort* a little more salient, which did not happen in the high-plateau frame. This was further enhanced by greater articulatory weight in the consonants (combined with similar vowel durations of 109 and 105 ms). In the low-tail frame, occlusion and release + aspiration were 53 ms and 17 ms for /d/, 40 ms and 43 ms for /t/, and there was glottalization in the following vowel. This compares with 43 ms and 16 ms for /d/, 38 ms and 23 ms for /t/ and smooth transition into the following vowel in the high-plateau frame. These differences in the two frames are clearly perceptible and add to the differences between the two prosodic patterns, potentially influencing the rhythmical sequencing in test word + *dort an* and favouring the perception of a trochaic pattern across the test word, i.e. initial stress *Kaffee*.

Thus, since the final sections of the two frames differ in their rhythmic structuring there may also be a rhythmic factor that would boost *Kaffee* perception in the low-tail frame. In the high-plateau frame, the rhythmic beats are as indicated by ': *Wir treffen uns 'regelmäßig beim 'Kaffee/Ca'fé dort an der 'Ecke*. The rhythmic regularity in the stimuli is guaranteed, irrespective of the number of syllables in between beat syllables: 4–4 with *Kaffee*, 5–3 with *Café*, because of compression, which reduces /a/ in *Café* to a [ə]-like vowel. In the low-tail frame, however, this rhythmic flow is broken by *dort*, which follows the test word, receiving an additional beat through extra prominence; *dort an* sets a heavy-light bar to which a heavy-light bar *Kaffee* may be judged a better rhythmic precursor than *Café* [Niebuhr, 2009b; Huggins, 1972], when vowel quality points to *Café* but vowel duration does not or is indecisive for either stress category. It is impossible to decide on the data of this test design which explanation is valid, since

the Fh and Fl sets differ as to vowel reduction in the test word and rhythmic structuring in the tail after the test word. To shed some further light on this issue a second experiment was run under hypothesis 5a.

3. Experiment 2: Testing for an Additional Rhythmic Effect

3.1 Test Design

The potential rhythmic effect in the low-tail versus the high-plateau frame prompted a follow-up experiment in which two sets of low-frame stimuli are generated that differ in the prominence of *dort* in the tail, supporting or not supporting a trochaic rhythm. The entire F1 set of experiment 1 was used again, and labelled Fl₁ (low-tail frame *with* prominence on *dort*). From this set a second set was created by splicing the low-tail *dort an der Ecke* from the original low-tail *Café* utterance to *Wir treffen uns regelmäßig beim Café/Kaffee* in each stimulus of the Fl₁ set, resulting in the Fl₀ set (low-tail frame *without* prominence on *dort*). In the original low-tail *Café* utterance, the durations in *dort* are: /d/ 43 ms occlusion and 16 ms release + aspiration, /t/ 44 ms occlusion and 15 ms release + aspiration, vowel 109 ms. They are the result of an iambic rhythm pattern following final-stress *Café* in production, and are comparable to the ones found in the original high-plateau utterance. As in the high-plateau frame, the transition into the vowel is without a glottalization break although the whole tail is breathy rather than modal-voiced. Thus, the set Fl₁ is identical with the set Fl in experiment 1, and set Fl₀ corresponds to Fh as far as the rhythmic structure in the tail is concerned, but the intonation difference is removed, and the test items in Fl₁ and Fl₀ are identical.

The two sets of 20 stimuli each were combined in one test file with 5 repetitions of each stimulus and one overall randomization. Otherwise the test procedure was the same as in experiment 1. A different sample of 16 students of linguistics and languages (6 male, 10 female, 1 female 50 years, 1 male 31 years, the rest 20–27 years) did the listening test in five groups.

3.2. Results

There were 6 missing responses in Fl₀ and 8 in Fl₁, in each case out of 16 × 100, i.e. <1%, which means that again subjects had no problem with the task. Since the misses are spread randomly across stimuli and subjects, they are treated as non-1 responses. Figure 6 provides the relative frequencies of ‘first syllable stressed’ responses in the 4 duration sets Fl₀_Wc, Fl₀_Wk, Fl₁_Wc, Fl₁_Wk, in each case giving the grand total as well as the totals for long and short /f/. The global data structure is fully comparable to the one in experiment 1.

Figures 7 and 8 break down the response data for experiment 2 in the same way as figures 4 and 5 do for experiment 1. Comparisons within figure 8, and of figure 8 with figure 5, provide a contrastive overview of the different response profiles for frames (a) Fl₀ (low-tail without prominence on *dort* in experiment 2), (b) Fl₁ (low-tail with prominence on *dort* in experiment 2), (c) Fh (high-plateau without prominence on *dort* in experiment 1), and (d) Fl (low-tail with prominence on *dort*

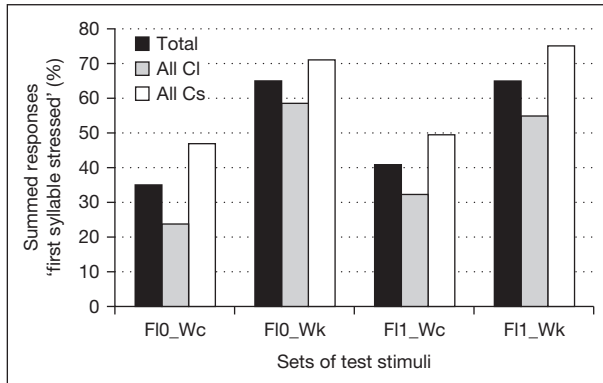


Fig. 6. Relative frequencies of 'first syllable stressed' responses in the 4 duration sets F10_Wc, F10_Wk, F11_Wc, F11_Wk, grand total and totals for long and short /f/; n = 800 in the total set and n = 400 in the CI and Cs sets.

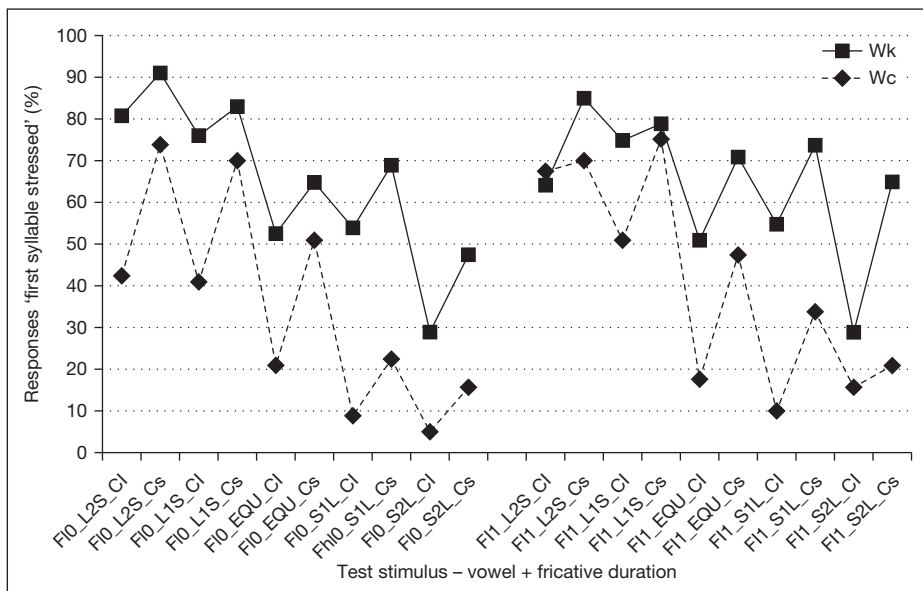


Fig. 7. Relative frequencies of 'first syllable stressed' responses to each of the 20 test stimuli in the low-tail frame without prominence on *dort* (F1₀, left) and in the low-tail frame with prominence on *dort* (F1₁, right). Plain/dotted lines connect responses to Wk/Wc for CI and Cs along the 5-point duration scale L2S to S2L. Each mean is based on 80 responses.

in experiment 1, identical stimulus set as for F1₁). Pairing (a) and (b) versus (c) and (d) leads to the following observations, which follow the same numbering as in experiment 1:

(1.1') For Wk_Cs, there is again no change of lexical stress category from L2S to S2L in either frame.

(1.2') For Wc_Cs, there are again fewer initial-stress judgements in both frames, now resulting in a lexical stress category change from L2S to S2L.

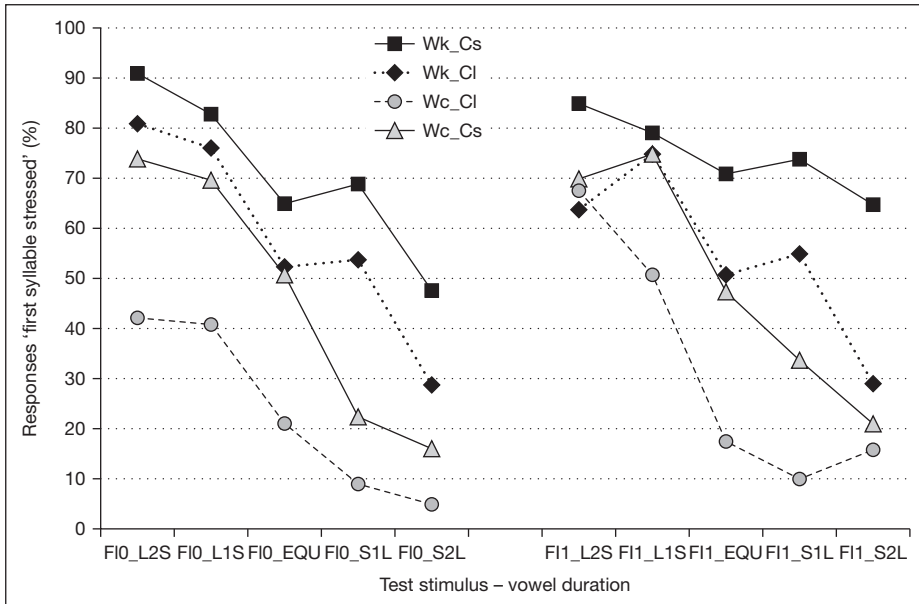


Fig. 8. Relative frequencies of 'first syllable stressed' responses to each of the 20 test stimuli in the low-tail frame without prominence on *dort* (Fl₀, left), and in the low-tail frame with prominence on *dort* (Fl₁, right), along the 5-point duration scale L2S to S2L in the 4 sets Wk_Cs, Wk_Cl, Wc-Cs, Wc_Cl (Fl₀: Audios 13-16). Each mean is based on 80 responses.

(2.1') For Wc_Cl, there is again no category change in the opposite direction from S2L to L2S in frame Fl₀, but the change does occur in L2S of Fl₁.

(2.2') For Wk_Cl, there are again more initial-stress judgements in both frames, now resulting in a lexical stress category change from S2L to L2S.

(3.1') In both frames, response profiles for Wk_Cs stay above 50% for initial stress. Initial-stress responses are boosted from L2S to S1L in Fl₀, compared with Fh, and are well above 50% at S2L in Fl₁, compared with Fl.

(3.2') For Wk_Cl, the relative frequencies range from 30% to around 80%, irrespective of the prosodic frame. In both frames, initial-stress responses are increased along the entire profile, compared with Fh and Fl.

(3.3') For Wc-Cs, Fl₀ and Fl₁ have very similar profiles with relative frequencies ranging from around 20% to around 70%, thus differing from Fh, which has fewer initial-stress responses than Fl at L1S and EQU.

(3.4') For Wc_Cl, there is an increase in initial stress responses at the left end of the duration scale in Fl₁ versus Fl₀, as in Fl versus Fh, but it is more pronounced, i.e. when the vowel qualities (Wc) and the fricative (Cl) point towards final stress, but the vowel durations point towards initial stress, there is a bias to initial-stress responses in the low-tail frame with prominence on *dort*.

The same inferential statistics were applied to the data. According to the Shapiro-Wilk test, compatibility with a normal distribution cannot be rejected with test value $T' = 0.950 > \text{critical value } T(16; 0.05) = 0.887$. The global RM-ANOVA with the 4

Table 5. Results of the WordFric \times Vowel Duration RM-ANOVAs for the non-prominent low-tail frame (F₀) and the prominent low-tail frame (F₁) subsets

Data subset/contrast	Factor/interaction	df1, df2	F	p	Partial η^2	
F ₀ : WordFric_Vowel	WordFric	3, 45	37.199	<0.001	0.713	
	VowDur	4, 60	33.543	<0.001	0.691	
	WordFric \times VowDur	12, 180	2.019	=0.025	0.119	
	Wk-Cs vs. Wc-Cl	WordFric	1, 15	82.596	<0.001	0.846
	Wk-Cs vs. Wk-Cl		1, 15	10.769	=0.005	0.418
Wk-Cl vs. Wc-Cs		1, 15	7.268	=0.017	0.326	
Wc-Cs vs. Wc-Cl		1, 15	15.172	=0.001	0.503	
F ₁ : WordFric_Vowel	WordFric	3, 45	29.711	<0.001	0.665	
	VowDur	4, 60	24.703	<0.001	0.622	
	WordFric \times VowDur	12, 180	5.346	<0.001	0.263	
	Wk-Cs vs. Wc-Cl	WordFric	1, 15	99.193	<0.001	0.869
	Wk-Cs vs. Wk-Cl		1, 15	28.274	<0.001	0.653
	Wk-Cl vs. Wc-Cs		1, 15	0.870	=0.366	0.055
	Wc-Cs vs. Wc-Cl		1, 15	17.000	=0.001	0.531

Dependent is frequency of initial-stress responses.

factors Frame (2 levels without/with *dort* prominence, F₀/F₁), Word (2 levels (original) *Café/Kaffee*, Wc/Wk), Vowel Duration (5 levels, L2S–S2L), and Fricative (2 levels long/schort, Cl/Cs) yields highly significant main effects: Word [F(1, 15) = 77.743, $p < 0.001$, $\eta_p^2 = 0.838$]; Vowel Duration [F(4, 60) = 35.310, $p < 0.001$, $\eta_p^2 = 0.702$]; Fricative [F(1, 15) = 28.061, $p < 0.001$, $\eta_p^2 = 0.652$]. As in experiment 1, these significances are based on considerable and very similar effect sizes in terms of partial eta-squared, i.e. dyadic vowel quality and duration of intervocalic /f/ have equally significant effects as dyadic vowel duration. Frame again misses significance: [F(1, 15) = 3.582, $p = 0.078$, $\eta_p^2 = 0.193$], but also has a much lower effect size. Of the interactions only 4 reach significance with much smaller sizes than the main effects: Word \times Vowel Duration [F(4, 60) = 6.411, $p < 0.001$, $\eta_p^2 = 0.299$]; Frame \times Word \times Vowel Duration [F(4, 60) = 3.342, $p = 0.015$, $\eta_p^2 = 0.182$]; Frame \times Word \times Fricative [F(4, 60) = 8.422, $p = 0.011$, $\eta_p^2 = 0.360$]; Word \times Vowel Duration \times Fricative [F(4, 60) = 5.141, $p = 0.001$, $\eta_p^2 = 0.255$]. Again the interactions involve Frame and Fricative, and Word and Fricative, with Vowel Duration. The same additional RM-ANOVAs were performed as in experiment 1. Table 5 lists the results of the WordFric \times Vowel Duration RM-ANOVAs and the within-subjects contrasts for the non-prominent low-tail frame (F₀) and the prominent low-tail frame (F₁) subsets.

In both the F₀ and the F₁ frame the factors WordFric and Vowel Duration are highly significant with large effect sizes. Their interactions are also significant, but with much smaller effect sizes. As in experiment 1, the within-WordFric contrasts along the scale Wk-Cs, Wk-Cl, Wc-Cs, Wc-Cl are highly significant for levels 1_4, 1_2 and 3_4, but are more weakly or not at all significant for levels 2_3, and the effect sizes decrease from 1_4 to 2_3 and then rise to 3_4.

Table 6 lists the results of the Frame \times Vowel Duration RM-ANOVAs for the Wc-Cl, Wc-Cs, Wk-Cs, and Wk-Cl data subsets.

Again the factor Frame is not significant in the global RM-ANOVA, but interacts significantly with Vowel Duration, Word and Fricative. In the 2-factor Frame \times Vowel

Table 6. Results of the Frame \times Vowel Duration RM-ANOVAs for the Wc-Cl, Wc-Cs, Wk-Cs, and Wk-Cl data subsets

Data subset	Factor	df	F	Significance	Partial η^2
Wc-Cl: Fl ₀ /I ₁ _V	Frame	1, 15	9.380	=0.008	0.385
	VowDur	4, 60	28.875	<0.001	0.658
	Frame \times VowDur	4, 60	3.734	=0.009	0.199
Wc-Cs: Fl ₀ /I ₁ _V	Frame	1, 15	1.268	=0.278	0.078
	VowDur	4, 60	36.899	<0.001	0.711
	Frame \times VowDur	4, 60	1.179	=0.329	0.073
Wk-Cs: Fl ₀ /I ₁ _V	Frame	1, 15	2.561	=0.130	0.146
	VowDur	4, 60	7.090	<0.001	0.321
	Frame \times VowDur	4, 60	3.595	=0.011	0.193
Wk-Cl: Fl ₀ /I ₁ _V	Frame	1, 15	1.136	=0.303	0.070
	Vowel	4, 60	13.531	<0.001	0.474
	Frame \times VowDur	4, 60	1.228	=0.309	0.076

Dependent is frequency of initial-stress responses.

Duration RM-ANOVAs, the factor Frame is not significant for Wc-Cs, Wk-Cs and Wk-Cl, nor is its interaction with the factor Vowel Duration for Wc-Cs and Wk-Cl. For Wc-Cl, Frame is significant with quite a substantial effect size in relation to the effect size of the factor Vowel Duration. The interactions Frame \times Vowel Duration are also significant for Wc-Cl and Wk-Cs. Post-hoc two-tailed paired t tests yield significant differences for the extreme vowel durations L2S in Wc-Cl [$t(15) = 3.872$, $p = 0.001$] and S2L in Wk-Cs [$t(15) = 3.217$, $p = 0.005$].

3.3. Discussion

The results show again that Vowel quality and intervocalic fricative duration have as strong effects as vowel duration on initial or final stress perception in disyllables when f_0 is smoothed in a low tail, and their effects on the response profiles along vowel duration are additive on a scale from *Kaffee* to *Café* manifestation. The general relationships between the response profiles for the different vowel-fricative combinations on this additive scale stay the same across the frames in both experiments. There is, however, also a small effect of Frame in both experiments. But whereas in experiment 1, the high-plateau frame triggers fewer initial-stress judgements for Wc-Cs stimuli at less decisive vowel durations than the low-tail frame with prominence, this does not apply to Fl₀ versus Fl₁ in experiment 2. On the one hand, *Café*-ness may be boosted by the more fronted/raised vowel qualities of high-plateau Wc-Cs, but on the other hand, *Kaffee*-ness may be boosted by the rhythmic patterning of low-tail Wc-Cs. So, the Fh_Wc-Cs and Fl_Wc-Cs profiles get separated at indecisive durations, resulting in a significant Frame effect in the 2-factor Frame \times Vowel Duration RM-ANOVAs of experiment 1. In experiment 2, on the other hand, only the rhythmic factor can be operative because the Wc-Cs stimuli are identical in the two low-tail frames. Thus, the Fl₁_Wc-Cs and Fl₀_Wc-Cs profiles can be expected to be less distinct. The main effect Frame is indeed not significant, and post-hoc two-tailed t tests for independent

samples with different variances yield significant differences for the vowel durations L1S and EQU in Fh_Wc-Cs versus F1₀_Wc-Cs [$t(28) = 2.298, p = 0.029$] and [$t(29) = 3.352, p = 0.002$], giving further support to this data interpretation.

Contrariwise, in experiment 2, the significant Frame effect shows up in stimuli where Vowel Quality and Fricative are clear cues to *Café*, but Vowel Duration is a clear cue to the opposite, i.e. in Wc_L2S_CI. Again only the rhythmic difference in the tail can have an influence, due to stimulus identity in the two low-tail frames. In such a constellation, a trochaic pattern in the rhythmic bar *dort an der* of F1₁ is most likely to bias the perception of the test word to initial stress to fit a trochaic frame. Since the same items, both with and without tail prominence, occurred in one test in experiment 2, the differential rhythmic effect can be expected to be heightened, compared with experiment 1, where the low-tails with prominence and the high-plateaux without prominence occurred in separate tests. The same argument also applies to the increase in initial-stress responses in S2L of F1₁/F1_Wk-Cs vs. F1₀_Wk-Cs.

4. Conclusion

This paper has investigated the contribution to stress perception of vowel quality and intervocalic fricative duration besides complementary vowel duration changes in the German disyllables *Kaffee* and *Café*, excluding critical f0 changes on the test words by placing them either in a low-tail or in a high-plateau hat pattern of a sentence frame. The analysis of the response data has confirmed hypotheses 1–3 for both experiments in as much as vowel duration, vowel quality and fricative duration have proved to be strong significant main effects in shaping the stress response profiles in German *Kaffee* versus *Café* along the duration scale.

All 4 combinations of vowel quality and fricative duration create different profiles, but there is no category change in the Word and Fricative combination Wc-CI, which has the strongest cue for final stress and does not change to initial stress even at the appropriate vowel durations, nor in Wk-Cs, which has the strongest cue for initial stress and does not change to final stress even at the appropriate vowel durations. In the weaker Word and Fricative combinations Wc-Cs and Wk-CI, where vowel quality and fricative duration provide opposing cues to initial or final stress, vowel duration becomes a dominant cue and the category change occurs. This means that the vowel quality and fricative duration values are additive to create a basic cue power for initial or final stress, onto which vowel duration is grafted. This applies to both prosodic frames in both experiments. Hypothesis 4 has thus also been confirmed.

As regards hypotheses 5 and 5a, there are marginal effects of the prosodic frame on stress perception. In experiment 1, the base stimulus *Café* in the high-plateau frame Fh was produced with greater centralization of /a/, approaching [ə], than in the low-tail frame F1. This can boost the final-stress cue of Wc-Cs in Fh but is absent from Wc-Cs in F1. In the case of Wc_CI, a rhythmic effect can operate in the opposite direction, increasing initial-stress responses in F1 and F1₁ because the low-tail prominence frames of both experiments contained an additional rhythmic beat on *dort an*, setting a trochaic pattern in which initial stress is a better fit. In Wc_L2S_Wc-CI, the strong Word and Fricative cues for final stress clash with the strong vowel duration cue for initial stress, which in the prosodic frame F1₀ results in indecision between the categories, but which in F1₁ can be resolved towards initial stress in the trochaic rhythmic pattern, thus

producing a category change. Similarly, in Wk_S2L-Cs there is a clash between the strong Word and Fricative cue for initial stress and the strong vowel duration cue for final stress, which results in indecision in the prosodic frame Fl_0 , but can be resolved towards initial stress in the trochaic rhythmic pattern of Fl_1 , thus bringing the response pattern along the whole duration scale well above 50%.

This data interpretation ties in with Niebuhr's [2009b] finding that the creation of a rhythmic pattern by f_0 in a context frame can influence lexical stress perception. The rhythmic-context effect not only occurred when the lexical target was at the end of the test sentences but also when it was at the beginning, and the perceptual effect was then regressive, as it would be in the present experiment, where both f_0 and duration properties create an extra rhythmic beat following the test word. The progressive or regressive direction of the effect is probably less relevant than the extension of the perceptual processing window to include syllable strings in which the test syllables are embedded. Huggins [1972], for example, mentions a duration effect on accent perception: in *has three sheets missing*, the primary accent shifts from *sheets* to *missing* in the same sentence intonation when the initial fricative is changed from long to short, which also involves assessment of segment durations in the entire phrase.

As f_0 on test items is levelled in a low-tail or a high-plateau hat pattern, vowel quality can become an important cue in lexical stress perception, further enhanced by the duration/strength of an intervocalic consonant. The combined cue force of these two parameters can outweigh the effect of vowel duration. Van Heuven and de Jonge [2011] also showed the effect of vowel quality for the Dutch minimal initial-final stress pair *canon* and *kanon*, but the effect is much smaller.

This is probably due to the different experimental design and to the lack of a proper minimal vowel quantity pair, since *canon* has a long stressed vowel, *kanon* a short unstressed one in the first syllable. */'kanɔn/ with a short stressed vowel in the first syllable would be a possible phonotactic structure but does not exist as a word. Therefore, changing duration along a 7-point scale that spans the long stressed and the short unstressed vowel in the Dutch word pair needs a wider range than the stressed-unstressed vowels in the German word pair, and a longer first-syllable vowel for *canon* identification. The manipulation of the first-syllable vowel covered a range of 120 ms, that of the second syllable 40 ms (at a 85/15 ratio for vowel/final /n/). This compares with just 40 ms in the vowels of both syllables of the German pair. Furthermore, the duration manipulation goes through */'kanɔn/, which does not have a straightforward association with *canon* in a word identification test and consequently with initial stress. The most reduced spectral steps do not go together with the longest duration steps in the production of /'ka:nɔn/, and listeners may therefore find it difficult to allocate them to either 'kanon' or 'canon'. Similarly, the most expanded spectral steps do not go together with the shortest duration steps in production, and may point to */'kanɔn/, which is not a possible response. Thus, duration is bound to have greater cue strength over vowel quality. The stimulus synthesis method may also have interfered with the naturalness of the sound output.

In a similar experiment to the German one, Ortega-Llebaria et al. [2010] manipulated duration in seven steps across the two syllables of Catalan *mama* ['mamə] and *mamà* [mə'ma], ranging from 87–45 ms to 45–87 ms, i.e. vowel durations are closer to the German than to the Dutch data, and range over 40 ms as in the former. Vowel quality has a strong effect: ['mamə] cannot be changed to final stress by short-long vowels, and vice versa for [mə'ma]. The effect is even stronger than the authors realize

because they applied an unusual and questionable test design, instructing listeners to respond only on hearing *mamá*, which must have increased false alarms and created a bias towards final stress. Nevertheless, the results parallel the ones for German in spite of the methodological flaw.

This is a field that needs further cross-linguistic investigation with more words, including different vowel and consonant types. Other test examples for German are: *Das ist bestimmt für den 'August/Au'gust nicht typisch.* 'That is certainly not typical for Augustus/the month of August.' or *Bei uns hat immer die 'Mama/Ma'ma/der 'Papa/Pa'pa das letzte Wort.* 'In our house, mum/dad always has the final word.', with different stylistic forms for mum/dad. The same stimulus generation with the variables Vowel Quality and Vowel Duration in the two syllables, Intervocalic Consonant, Low-tail or High-plateau frames, and the same test design would be applied.

From the limited data we have we can already conclude that investigations into the exponents of lexical stress and into their effects on stress perception need to open a broad analysis window that not only spans at least a stressed and an unstressed syllable, but includes all segmental and prosodic variables in it, whose weights for signalling stress are ranked differentially according to language and speech conditions.

The general hierarchy of the variables traditionally adduced to signal stress has been questioned in several investigations, e.g. by Berinstein [1979] and by Potisuk et al. [1996]. However, in these studies the phenomenon of 'stress' is approached from a variety of definitions which do not coincide with the category of 'lexical stress' applied in this paper. As Kohler [2008b] argued, it may refer (a) to relative syllable salience in an utterance, i.e. in the general sense of 'prominence', comprising all the factors that make syllables stand out to various degrees, which is syllable-string, not word-oriented, (b) to syllable weight at the level of the phonology of the word, i.e. in the most restricted sense of 'lexical stress', or (c) to word focus in utterances for various aspects of propositional and expressive meaning in the less restricted usage of '(sentence) accent'.

Sense (a) belongs to the study of speech rhythm [Kohler, 2009], sense (b) to lexical phonology, where lexical stress marks a position in a word for an accent to hook onto, sense (c) to sentence prosody. Of course, (b) only applies if the lexical items of the language have such a phonologically marked place holder; it does for example, not occur in French, where the word is a flexible unit in a 'mot phonétique' and a phrasal accent marks the final non-schwa syllable of such a phrasal unit [Kohler, 2010]. If words are phonologically marked for lexical stress position in a language, the default coincidence of the exponents of different types of accent with this position is crucial for intelligibility, no matter whether lexical stress is free (as in Russian) or marginally morphologically determined (as in Dutch, English, German), thus having some additional functional load in the differentiation of lexical meaning, or whether the position is fixed (initial in Czech, Finnish, Hungarian, final in K'ekci, penultimate in Polish). Lexical stress also determines the system of distinctive segmental units, especially vowels, that can fill the marked position against the unmarked ones. In German, for example, tense and lax vowels are only contrasted in lexically stressed syllables, in unstressed syllables of simplex words, tense vowels occur in open, lax vowels in closed syllables, in addition to [ə], e.g. *De'zember* [e], '*Kaffee* [a _ e:], '*Sep'tember* [ɛ]; '*Papa* [a _ a], '*Ca'fé* [a _ e:], '*Japan* [a: _ a], '*Papi* [a _ i:], '*Patin* [a: _ ɪ], '*Pappe* [a _ ə]; '*No'vember* [o], '*Auto* [o:], '*Ok'tober* [ɔ].

Senses (b) and (c) are often conflated in investigations, as both are word-oriented, although within different domains, i.e. the lexical item by itself versus the word in

an utterance. In this paper, 'stress' has been applied in the strict sense of (b), and the acoustic-perceptual properties of two words containing the same disyllabic segmental string with either initial or final stress position have been examined in two sentence accent patterns, determined by a high-plateau or a low-tail frame. Berinstein [1979], on the other hand, analysed the perception by English, K'ekci, and Spanish listeners of salient prominence in a synthesized series of four [bɪ] syllables, varying vowel duration in one of them in each of the four positions. This is a cross-language study of rhythmic prominence perception in logatomes [Kohler, 2009], and therefore represents sense (a) of 'stress'. So, it does not relate to the perception of real words that differ in property bundles for the cueing of 'lexical stress' position in different accentual patterns in specific languages. Potisuk et al. [1996] examined the acoustic correlates duration and f0 of 'stress' in Thai, but again they were not concerned with 'lexical stress' in the narrow sense of (b), but with accentual prosody patterns.

Henderson's [1949] detailed description of prosodies in Thai shows that Thai does not have the category of 'lexical stress' in its phonology of the word. The suprasegmental contrastive features operating at the word level are tone and vowel quantity. They are adjusted as they are embedded in phrase and sentence prosodies, resulting in accentual patterns, which Potisuk et al. [1996] refer to as 'stress'. Their investigation deals with the distinction between a two-syllable strong-strong noun-verb construction versus a two-syllable weak-strong noun compound. The accentual tone and quantity reorganization in compounds in Thai is a tone language parallel to lexical stress pattern readjustment above the elementary lexical level in English compounds and constructions, e.g. 'buttercup (flower), 'butter ,cup (butter dish), 'butter 'icing, where primary and secondary lexical stress need to be distinguished [Kingdon, 1963]. Restructuring of lexical stress patterns is also found in English words with double stress, e.g. numerals in '-teen or words like 'prin'cess, 'Picca'dilly. In the context of another lexical stress in an immediately preceding or following word, forming a close syntagma with the double-stress word, stress is phonologically reorganized into a new double-stress pattern at the initial and the final stress position of the syntagma, no doubt in order to create regular two-beat rhythms: 'just eigh'teen, 'eighteen 'years, or a 'young prin'cess, 'Princess 'Anne. The stressed syllables need not be contiguous for this to apply: a 'royal prin'cess, 'Princess Vict'oria, 'close to Picca'dilly, 'Piccadilly 'Circus.

Although Berinstein [1979] and Potisuk et al. [1996] did not do their analyses on the narrow level of 'lexical stress' they drew their conclusions regarding the relationship between 'stress' and acoustic parameters exclusively on the lexical phonological level and hypothesized that this relationship varies across languages as a function of their phonological structure. This functional load hypothesis states that 'phonetic resources are available for signalling stress if and only if they are not used up in signalling lexical contrast' [Potisuk et al., 1996, p. 201]. The generalization was contradicted straight away by their results, because, although f0, in keeping with the hypothesis, turned out to have a subsidiary role as a 'stress' correlate in distinguishing the two Thai constructions, duration had a primary role against the hypothesis. To maintain their general functional-load claim the authors modified the hypothesis post hoc: 'The question is not simply whether an acoustic parameter is implicated in lexical contrast, but rather what the degree of involvement is at the lexical or morphological level. In Thai . . . virtually every word or syllable carries a contrastive tone. Minimal pairs for vowel length, on the other hand, are comparatively rare. Under a revised version of the hypothesis, the magnitude of f0's involvement in lexical contrasts far exceeds that of

duration. Of the prosodic cues to stress, duration is available in Thai to signal stress, without significant disturbance of lexical contrasts. Therefore, the present findings are compatible with the idea that the relative importance of acoustic cues of stress varies depending on the extent of their use at the lexical or morphological level' [Potisuk et al., 1996, p. 211].

What is immediately disturbing in this argumentation is its circularity: the original hypothesis based on Berinsein's [1979] investigation into prominence perception was adjusted to fit the completely different Thai data domain, and then the Thai data are seen as the outcome of this new hypothesis. There is no doubt that languages organize the relationship between the three aspects of 'stress' in different ways, thus refuting a generally valid cue hierarchy, but the functional load hypothesis is not the answer. What we need are fine-phonetic cross-linguistic analyses of the acoustic parameters cueing 'stress' that at the same time keep the different referents of the term clearly separate. Moreover, we also need to add a speech dimension to the phonological argument. This paper has attempted such a dimension in the argumentation against a general hierarchy with a fixed ranking of the variables traditionally adduced to signal lexical stress. The conclusion is that variables of f_0 , vowel duration, vowel quality, consonant duration, acoustic energy operate together at all times to transmit 'lexical stress' information in a language, like German, that has the category. They do this not just in the stressed syllable but across at least a stressed/unstressed or unstressed/stressed syllable sequence, and the weight of each individual parameter is adapted to the contextual and situational demands of speech communication. Thus, every new prosodic embedding of syllable sequences defines the hierarchy afresh, and artificial separation of variables in experimental designs should always pay attention to naturalness and communicative appropriateness of the stimuli.

Acknowledgement

The idea for this paper originated from the submission to *Phonetica* of 'Spectral and temporal reduction as stress cues in Dutch' by Vincent van Heuven and Mirjam de Jonge (published in 2011). I am very grateful to the authors and to Vincent van Heuven in particular, with whom I corresponded about their study, and who gave insightful comments and made very helpful suggestions for revision of my own study. My thanks also go to Ernst Dombrowski and Oliver Niebuhr for fruitful discussions, and last but not least, to the many informants who participated in the listening tests without remuneration and thus made the investigation possible.

References

- Bergem, D. van: Acoustic vowel reduction as a function of sentence accent, word stress, and word class on the quality of vowels. *Speech Commun.* 12: 1–23 (1993).
- Berinsein, A.E.: A cross-linguistic study on the perception and production of stress. *UCLA Working Papers Phonet.* 47 (1979).
- Butcher, A.: Coarticulation in intervocalic voiceless plosives and fricatives in connected speech. *Arbeitsber. Inst. Phonetik Univ. Kiel (AIPUK)* 8: 154–213 (1977).
- Duden Aussprachewörterbuch; 2nd ed. (Bibliographisches Institut, Mannheim 1974).
- Fry, D.B.: Duration and intensity as physical correlates of linguistic stress. *J. acoust. Soc. Am.* 27: 765–768 (1955).
- Fry, D.B.: Experiments in the perception of stress. *Lang. Speech* 1: 126–152 (1958).
- Fry, D.B.: The dependence of stress judgments on vowel formant structure; in Zwirner, Bethge, Proc. 6th int. Congr. Phonet. Sci., pp. 306–311 (Karger, Basel 1965).

- Henderson, E.J.A.: Prosodies in Siamese: a study in synthesis. *Asia Major (New Series) 1*: 189–215 (1949). Reprinted in Palmer, *Prosodic Analysis*, pp. 27–53 (Oxford University Press, London 1970).
- Heuven, V.J. van; de Jonge, M.: Spectral and temporal reduction as stress cues in Dutch. *Phonetica 68*: 120–132 (2011).
- Huggins, A.W.F.: Just noticeable differences for segment duration in natural speech. *J. acoust. Soc. Am. 51*: 1270–1278 (1972).
- Isačenko, A.V.; Schädlich, H.-J.: Untersuchungen über die deutsche Satzintonation. *Studia Grammatica VII*, pp. 7–67 (Akademie-Verlag, Berlin 1966).
- Kingdon, R.: *The groundwork of English stress*; 3rd ed. Longmans, London 1963.
- Kohler, K.J.: Macro and micro F0 in the synthesis of intonation; in Kingston, Beckmann, *Papers in Laboratory Phonology. I*, pp. 115–138 (Cambridge University Press, Cambridge 1990).
- Kohler, K.J.: Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsber. Inst. Phonet. dig. Sprachverarb. (AIPUK) 25*: 115–185 (1991).
- Kohler, K.J.: ‘Speech-smile’, ‘speech-laugh’, ‘laughter’ and their sequencing in dialogic interaction. *Phonetica 65*: 1–18 (2008a).
- Kohler, K.J.: The perception of prominence patterns. *Phonetica 65*: 257–269 (2008b).
- Kohler, K.J.: Rhythm in speech and language. *Phonetica 66*: 29–45 (2009).
- Kohler, K.J.: The transmission of meaning by prosodic phrasing: a comparison of French with English and German using no Ls and Hs. *Phonetica 67*: 100–124 (2010).
- Koopmans-van Beinum, F.J.: *Vowel contrast reduction* (Academische Pers, Amsterdam 1980).
- Li, Y.; Kong, J.: An electropalatographic and acoustic study of the tonal effects on vowel production in Standard Chinese: a pilot study. *Proc. 3rd Int. Symp. on Tonal Aspects of Language, Nanjing*, p. 44 (2012).
- Moshammer, C.; Geng, C.: Acoustic and articulatory manifestations of vowel reduction in German. *J. int. phonet. Ass. 38*: 117–136 (2008).
- Niebuhr, O.: Intonation segments and segmental intonations. *Proc. 10th Interspeech Conf., Brighton 2009a*, pp. 2435–2438.
- Niebuhr, O.: Fundamental frequency-based rhythm effects on the perception of local syllable prominence. *Phonetica 66*: 95–112 (2009b).
- Niebuhr, O.: On the phonetics of intensifying emphasis. *Phonetica 67*: 170–198 (2010).
- Ortega-Llebaria, M.; Vanrell, M.M.; Prieto, P.: Catalan speakers’ perception of word stress in unaccented contexts. *J. acoust. Soc. Am. 127*: 462–471 (2010).
- Potisuk, S.; Gandour, J.; Harper, M.P.: Acoustic correlates of stress in Thai. *Phonetica 53*: 200–220 (1996).
- Sluijter, A.M.C.; van Heuven, V.J.: Effects of focus distribution, pitch accent and lexical stress on the temporal organisation of syllables in Dutch. *Phonetica 52*: 71–89 (1995).
- Sluijter, A.M.C.; van Heuven, V.J.: Spectral balance as an acoustic correlate of linguistic stress. *J. acoust. Soc. Am. 100*: 2471–2485 (1996).
- Sluijter, A.M.C.; van Heuven, V.J.; Pacilly, J.J.A.: Spectral balance as a cue in the perception of linguistic stress. *J. acoust. Soc. Am. 101*: 503–513 (1997).
- Verner, K.: Eine Ausnahme der I. Lautverschiebung. *Z. f. vergl. Sprachforsch. 23*: 97–130 (1876).
- WDA (Wörterbuch der deutschen Aussprache) (Hueber, München 1969).