

Institut für Phonetik	
Inv. Nr.	Sign. R
993,36	WIE 3 (H)

Eckart Weiher

LAUTWAHRNEHMUNG UND LAUTPRODUKTION
IM ENGLISCHUNTERRICHT FÜR DEUTSCHE

VORWORT

Ohne die Unterstützung zahlreicher Beteiligter wäre diese Arbeit nicht zustande gekommen. Zu danken ist: dem Direktor des Instituts für Phonetik der Universität Kiel, Prof. Dr. K. Kohler, der diese Untersuchung angeregt hat und mir zu ihrer Durchführung die Einrichtungen des Instituts zur Verfügung gestellt hat; Herrn Obermagistratsschulrat Maibohm vom Schulamt der Stadt Kiel, von dem ich die Erlaubnis zur Durchführung der Untersuchungen in zwei Kieler Schulen erhielt; Herrn Studiendirektor Klimant und Herrn Studienrat Menzel vom Gymnasium im Schulzentrum Mettenhof sowie dem Leiter der Ludwig-Richter-Schule, Herrn Rektor Thielmann, die mir ihre Klassen zur Verfügung stellten.

Dank gilt auch allen meinen Kollegen am Kieler Institut für Phonetik, die sich als Informanten bzw. Versuchspersonen zur Verfügung stellten und sich mit mir in die Aufgabe der Beurteilung der lautsprachlichen Produktion der Schülergruppen teilten; besonders hervorzuheben ist das Bemühen von W.J. Barry und A.R. Butcher. Bei der technischen Durchführung der Gruppen- und Einzeltests haben mich Herr Ing.(grad.) J. Petersen und Herr H. Janßen unterstützt.

Für Rat und Hilfe in statistischen Fragen schließlich habe ich Herrn Dipl.-Math. H. Nowak von der Abteilung Medizinische Statistik und Dokumentation der Rheinisch-Westfälischen Technischen Hochschule Aachen zu danken.

Diese Arbeit hat der Philosophischen Fakultät der Christian-Albrechts-Universität Kiel als Dissertation vorgelegen.

INHALT

1.0.0.	Einleitung	1
1.1.0.	Gegenstand der Untersuchung	1
1.2.0.	Kritik der kontrastiven Phonologie	5
1.3.0.	Kontrastive Analyse und der Gegenstands- bereich dieser Untersuchung	11
2.0.0.	Testkonstruktion, Testanordnung und Testdurchführung	14
2.1.0.	"Informant"	14
2.1.1.	Idiolekt	14
2.2.0.	Aufnahme	17
2.3.0.	Herstellung der Testbänder	17
2.4.0.	"Kontrollgruppe" und Versuchspersonen	18
2.4.1.	"Kontrollgruppe"	18
2.4.2.	Versuchspersonen	21
2.5.0.	Durchführung der Hörtests	23
2.5.1.	Kontrollgruppe	23
2.5.2.	Gruppe "LR"	24
2.5.3.	Gruppe "M"	25
2.5.4.	Test I: Perzeption (Hördiskrimination)	28
2.5.5.	Test II: Perzeption (Höridentifikation)	36
2.6.0.	Produktionstests	41
2.6.1.	Auswahl der Probanden	41
2.6.2.	"Informant"; Aufnahme; Herstellung der Testbänder	44
2.6.3.	Durchführung der Produktionstests	44
2.6.4.	Test III: Perzeption-Produktion: Nonsensewörter	46
2.6.5.	Test IV: Perzeption-Produktion: englische Einzelwörter	48
2.6.6.	Produktion-Perzeption-Produktion: englische Sätze	50
3.0.0.	"Testing the Test": Kritische Prüfung der Perzeptionstests	53
3.1.0.	Zufallsfreiheit der Entscheidungen	54
3.1.1.	Teststatistik	54

3.1.2.	Zufallsschranken für Test I und II	56
3.1.3.	Gruppenunterschiede (absolute Fehlerhäufigkeit)	57
3.2.0.	Unterschiede zwischen Test I und II	58
3.3.0.	Einfluß der Geschlechtszugehörigkeit	60
3.4.0.	Testverhalten in Abhängigkeit von der Gesamttestzeit	61
3.5.0.	"Positionseffekt"	64
3.5.1.	Ausgangshypothese	65
3.5.2.	Statistischer Nachweis	65
3.5.3.	Wertung	66
3.6.0.	Optimale Testdauer	67
3.6.1.	Statistischer Nachweis	67
3.6.2.	Wertung	69
3.7.0.	Trennung der Gruppen	70
4.0.0.	Beurteilung der Schülerproduktionen	72
4.1.0.	Auditive Beurteilung durch zwei geschulte Beobachter	72
4.1.1.	Auditive Beurteilung und Transkription: Beobachter 1	72
4.1.2.	Auditive Beurteilung: Beobachter 2	75
4.2.0.	Auditive Beurteilung und phonetische Parameter	76
4.2.1.	Phonetische Parameter: /θ,ð/	77
4.2.2.	Phonetische Parameter: /r/	78
4.2.3.	Phonetische Parameter: /i, e, æ/	80
4.2.4.	Phonetische Parameter: "Final voiced consonants"	80
4.3.0.	Beurteilung nach Akzeptabilität	83
4.3.1.	Beurteilungskriterien	83
4.3.2.	Normbezogenheit	84
4.4.0.	Nichtübereinstimmung der Akzeptabilitätsurteile	85
4.4.1.	Kriterienvergleich und Bestimmung der Urteilsabweichungen	86
4.4.2.	Nichtübereinstimmung in den unterschiedlichen Ausspracheproblembereichen	87
4.4.3.	Nichtübereinstimmung bei der Beurteilung individueller Leistungen	89

4.4. 4.	Entscheidung bezüglich der Nichtübereinstimmungen: Beobachter 3	89
4.5. 0.	Phonetische Fehleranalyse	91
4.6. 0.	Quantifizierung der Produktionsdaten	91
5.0. 0.	Lautwahrnehmung und Lautproduktion: Testergebnisse	94
5.1. 0.	TH-Test	94
5.1. 1.	Perzeption: /θ/	
5.1. 2.	Fehlervorgabe	95
5.1. 3.	Hördiskrimination	97
5.1. 4.	Höridentifikation	103
5.1. 5.	Perzeption-Produktion: Nonsensewörter	107
5.1. 6.	Perzeption-Produktion: englische Einzelwörter	115
5.1. 7.	Produktion-Perzeption-Produktion: englische Sätze	117
5.1. 8.	Beurteilung der /θ/-Realisationen nach Akzeptabilität	118
5.1. 9.	Vergleich der Perzeptions- und Produktionsleistungen (/θ/)	120
5.1.10.	Perzeption: /ð/	125
5.1.11.	Fehlervorgabe	125
5.1.12.	Hördiskrimination	125
5.1.13.	Höridentifikation	129
5.1.14.	Perzeption-Produktion: Nonsensewörter	132
5.1.15.	Perzeption-Produktion: englische Einzelwörter	136
5.1.16.	Produktion-Perzeption-Produktion: englische Sätze	138
5.1.17.	Beurteilung der /ð/-Realisationen nach Akzeptabilität	139
5.1.18.	Vergleich der Perzeptions- und Produktionsleistungen (/ð/)	139
5.2. 0.	R-Test	144
5.2. 1.	Perzeption: /r/	144
5.2. 2.	Fehlervorgabe	144
5.2. 3.	Hördiskrimination	146
5.2. 4.	Höridentifikation	150
5.2. 5.	Perzeption-Produktion: Nonsensewörter	153

5.2.6.	Perzeption-Produktion: englische Einzelwörter	155
5.2.7.	Produktion-Perzeption-Produktion: englische Sätze	157
5.2.8.	Beurteilung der /r/-Realisationen nach Akzeptabilität	158
5.2.9.	Vergleich der Perzeptions- und Produktionsleistungen	160
5.3.0.	V-Test	163
5.3.1.	Perzeption: /i,e,ə/	163
5.3.2.	Fehlervorgabe	163
5.3.3.	Hördiskrimination	164
5.3.4.	Höridentifikation	167
5.3.5.	Perzeption-Produktion: Nonsensewörter	170
5.3.6.	Perzeption-Produktion: englische Einzelwörter	177
5.3.7.	Produktion-Perzeption-Produktion: englische Sätze	181
5.3.8.	Beurteilung der Vokalrealisationen nach Akzeptabilität	187
5.3.9.	Vergleich der Perzeptions- und Produktionsleistungen	190
5.4.0.	FVC-Test	193
5.4.1.	Perzeption	193
5.4.2.	Fehlervorgabe	194
5.4.3.	Hördiskrimination	195
5.4.4.	Höridentifikation	197
5.4.5.	Perzeption-Produktion: Nonsensewörter	199
5.4.6.	Perzeption-Produktion: englische Einzelwörter	204
5.4.7.	Produktion-Perzeption-Produktion: englische Sätze	205
5.4.8.	Beurteilung nach Akzeptabilität	206
5.4.9.	Vergleich der Perzeptions- und Produktionsleistungen	210
6.0.0.	Hierarchie von Lernschwierigkeiten	213
6.1.0.	Perzeptionsschwierigkeiten	213
6.2.0.	Produktionsschwierigkeiten	214

6.3.0.	Perzeption und Produktion: Zusammen- hänge auf individueller Ebene	216
6.4.0.	Zusammenfassung und Schluß	220
	Literatur	222
Anhang:	Tab. I - VII	232
	Testwörter: Test III	
	Perzeption-Produktion (Nonsensewörter)	240
	Testwörter: Test IV	
	Perzeption-Produktion: englische Einzelwörter	241
	Testwörter: Test V	
	Produktion-Perzeption-Produktion: englische Sätze	242

1.0.0. Einleitung

Die vorliegende Arbeit untersucht das Verhältnis von lautsprachlicher Perzeption und Produktion bei zwei deutschen Schulklassen, die Englisch als erste Fremdsprache erlernen. Sie setzt voraus, daß, wie unterschiedlich auch immer die Gewichte gesetzt werden, der Erwerb einer weitgehend akzentfreien Aussprache Ziel jedes fremdsprachlichen Unterrichts ist, der letztthin darauf bedacht ist, den Schüler zu angemessenem sprachlichen Handeln in konkreten Kommunikationssituationen zu befähigen.¹ Ein solches Ziel verlangt nach Beseitigung solcher den reibungslosen Kommunikationsablauf gefährdenden Störungen, wie es das beharrliche Fortbestehen eines fremden Akzents beim Sprecher bedeutet.

1.1.0. Gegenstand der Untersuchung

Gegenstand der Untersuchung sind vier Ausspracheprobleme des Englischen, die sowohl in der einschlägigen Literatur als vor allem auch in der alltäglichen Praxis des Englischunterrichts für Deutsche reich dokumentiert sind, nämlich

1. die englischen interdentalen Frikative (stimmlos) /θ/ und (stimmhaft) /ð/;²
2. das englische /r/ mit seinen häufigsten Varianten;³

¹Wir teilen die Auffassung, daß das Prinzip der "Fertigkeitsspezifität" unbestrittenes Grundprinzip jeglichen Fremdsprachenunterrichts ist: siehe C. Heeschen u. M. Nehr, "Psycholinguistische Bemerkungen zur Stellung der generativen Transformationsgrammatik im Fremdsprachenunterricht", Praxis des neu sprachlichen Unterrichts, 21 (1974), 3-11.

²Zu /θ, ð/ siehe D. Jones, An Outline of English Phonetics, 9th ed. repr. Cambridge (Engl.) 1969, S. 182-184; A.C. Gimson, An Introduction to the Pronunciation of English, 2nd ed. London 1970, S. 183-185.

³Zu /r/ siehe Jones (1969), S. 193-200; Gimson (1970), S. 205-212.

3. die Reihe der englischen palatalen Kurzvokale /ɪ/, /e/, /æ/,⁴ und
4. die sog. "Stimmhaftigkeitsopposition" im Auslaut englischer Wörter, also der Gegensatz in Minimalpaaren wie /bit/ - /bid/, /sat/ - /səd/, /rent/ - /rend/.⁵

Die häufigsten Fehler deutscher Englischlerner bezüglich dieser Ausspracheprobleme sind:

1. die Realisation der englischen interdentalen Frikative /θ/ und /ð/ als [s] bzw. [z];⁶

⁴Jones' "English Vowels Nos. 2, 3, 4": siehe Jones (1969), S. 66-73; Gimson (1970), S. 101-107. - Bei der Wahl der Symbole für diese Vokalphoneme schließen wir uns dem Gebrauch von Gimson an.

⁵Im Unterschied zu den vorher genannten Ausspracheproblemen, die paradigmatische Relationen betreffen, sind bei diesem Gegensatz syntagmatische Einheiten betroffen: siehe im einzelnen weiter unten, 4.2.3. - Die einschlägigen Regeln für die Länge englischer Vokale und Konsonanten (Nasale und Liquide) finden sich bei Jones (1969), S. 232-237 (siehe besonders Rules II, III, V, VI); vgl. Gimson (1970), S. 94-95.

⁶Siehe etwa R. Arnold u. K. Hansen, Phonetik der englischen Sprache; Eine Einführung, Leipzig 1968, S. 85; W.J. Barry u. Chr. Gutknecht, Practical English Phonetics; A Programmed Course in English Pronunciation, 4. Aufl. Hamburg 1974, S. 38-39; H.L. Kufner, Kontrastive Phonologie Deutsch-Englisch, Stuttgart 1971, S. 40-44 (die von Kufner beschriebene Variante des Englischen ist das amerikanische Englisch); J.D. O'Connor, Better English Pronunciation, Cambridge (Engl.) 1967, S. 172-173; G. Scherer u. A. Wollmann, Englische Phonetik und Phonologie (=Grundlagen der Anglistik und Amerikanistik, Bd. 6), Berlin 1972, S. 94-95; C.M. Wise, Applied Phonetics, Englewood Cliffs 1957, S. 394-396 ("German Substitutions in English"). Die meisten dieser Werke enthalten sprachdidaktische Hinweise, die auf den Erwerb einer korrekten Aussprache der interdentalen Frikative zielen. - Kufner (1971) vermerkt, daß für den deutschen Schüler hauptsächlich /θ/ und /ð/ (selten /z/) als Ersatz für /ð/ in Frage kommen (S. 43). Wise (1957), S. 396, vertritt die Ansicht, daß [ð] "will usually be replaced by [d] (ähnlich [t] für [θ]); "a few Germans, usually the more educated ones, will substitute [z] for [ð]" (entsprechend [s] für [θ]). Auf die hohe auditive Ähnlichkeit von [θ, ð] und [f, v] verweisen Scherer u. Wollmann (1972), S. 95; K. Loggen,

2. der Ersatz englischer linguale r-Laute durch die nach regionaler Herkunft des Sprechers unterschiedlichen eigensprachlichen deutschen r-Laute;⁷
3. der häufig beobachtbare Zusammenfall (in der lautsprachlichen Realisation durch deutsche Sprecher) der englischen palatalen Kurzvokale /e/ und /æ/⁹ in einem Vokal von mittlerem Öffnungsgrad zwischen /e/ und /æ/, der qualitativ dem deutschen halboffenen kurzen /ɛ/¹⁰ ähnlich ist;

Didaktische Phonetik des Englischen im Studium des Sprachlehrers, Bad Heilbrunn 1972, S. 36-37, zieht daraus eine nach unserem Dafürhalten bedenkliche Folgerung für die Aussprachendidaktik ("Es verrät eine äußerst unglückliche Art der Lautschulung, wenn die Schüler ein /s/ für ein /θ/ sprechen, und der Lehrer nicht weiß, daß man zur Not auf /f/ ausweichen könnte").

⁷Siehe etwa Arnold u. Hansen (1968), S. 88-90; Barry u. Gutknecht (1974), S. 37-38; Kufner (1971), S. 49-51; O'Connor (1967), S. 173; Scherer u. Wollmann (1972), S. 102-106; Wise (1957), S. 395.

⁸Siehe etwa Arnold u. Hansen (1968), S. 57-58; Barry u. Gutknecht (1974), S. 46; Kufner (1971), S. 105 (siehe auch die Lehrvorschläge zu diesem Kontrastproblem, S. 110-111); O'Connor (1967), S. 173; Scherer u. Wollmann (1972), S. 136-137; Wise (1957), S. 396.

⁹Nach Gimson (1970), S. 104, liegt die Qualität von /e/ zwischen der der Kardinalvokale 2 ([e]) und 3 ([ɛ]), also /e/ = [e] oder [ɛ]; in der "Received Pronunciation" (RP) tendiert /e/ näher zu Kardinalvokal 2 als zu 3. - /æ/ ist qualitativ zwischen den Kardinalvokalen 3 ([ɛ]) und 4 ([a]) anzusiedeln, tendiert jedoch näher zu Kardinalvokal 3 als zu 4, also /æ/ = [ɛ]. - Zum System der Kardinalvokale ("Cardinal Vowels") siehe Jones (1969), S. 31-35; zur "Received Pronunciation" siehe Gimson (1970), S. 85-89; vgl. auch Jones' Einführung zu seinem English Pronouncing Dictionary, 13th ed. repr. London 1969, S. xvii-xviii.

¹⁰Von den drei gängigsten Aussprachewörterbüchern des Deutschen (Th. Siebs, Deutsche Aussprache; Reine und gemässigte Hochlautung mit Aussprachewörterbuch, 19. Aufl. 1983, v. H. de Boor, H. Moser u. Chr. Winkler, Berlin 1969; Duden Aussprachewörterbuch, bearbeitet von M. Mangold in Zusammenarbeit mit P. Grebe (= Der Große Duden, Bd. 6), Mannheim 1962;

4. die Nichtbeachtung der "Fortis-Lenis"-¹¹

Opposition im Auslaut englischer Wörter, die die Neutralisation einer für das Englische wichtigen phonologischen Opposition zur Folge hat.¹²

Die Untersuchung beschreitet den Weg der empirischen Erhebung von Befunden, wie ihn insbesondere die Arbeit von Nemser zur Kontaktsituation Ungarisch-Englisch gegangen ist.¹³ Sie emanzipiert sich damit weitgehend vom Verfahren der herkömmlichen kontrastiven Analyse, die in den letzten zehn Jahren das Feld phonetischer Forschung für den Fremdsprachenunterricht beherrscht hat. Es scheint angebracht, die Voraussetzungen und Methoden der kontrastiven Analyse (im Bereich der Phonologie) kurz darzustellen und kritisch zu prüfen.

Wörterbuch der deutschen Aussprache, hgg. v. H. Krech u.a., 2. Aufl. Leipzig u. München 1969) nimmt nur der Duden ausdrücklichen Bezug auf das System der Kardinalvokale (S. 22). Die Qualität des kurzen (halb-)offenen /ε/ wird von Mangold als zwischen Kardinalvokal 2 und 3, aber sehr nahe an 3 angegeben: vgl. das Vokalviereck für die Vokale der deutschen Hochlautung, Duden Aussprachewörterbuch, S. 29.

¹¹Zur hier verwendeten Terminologie siehe weiter unten, S. 80-83.

¹²Siehe Barry u. Gutknecht (1974), S. 35-36; Jones (1969), S. 243; Kufner (1971), S. 61-66; O'Connor (1967), S. 172-173; Scherer u. Wollmann (1972), S. 87-88, 113, 149-151; Wise (1957), S. 397. - Hinsichtlich der didaktischen Ratschläge, die zu einer Bewältigung dieser Schwierigkeit bei deutschen Englischlernenden führen soll, herrscht bei den genannten Autoren wenig Einheitlichkeit.

¹³W. Nemser, An Experimental Study of Phonological Interference in the English of Hungarians (=Indiana University Publications, Uralic and Altaic Series, Vol. 105), Bloomington u. Den Haag 1971.

1.2.0. Kritik der kontrastiven Phonologie¹⁴

Die Entwicklung des Phonembegriffs seit dem Ende des 19. Jahrhunderts weist wenig Einheitlichkeit auf und ist bis auf den heutigen Tag gekennzeichnet durch heftige Auseinandersetzungen bezüglich des Existenzmodus des Phonems und seines Status als einer Einheit linguistischer Beschreibung. Als fruchtbringend und nützlich hat sich das Konzept "Phonem" auf dem Gebiet der Verschriftung,¹⁵ der Entwicklung praktischer und ökonomischer Orthographien erwiesen, und in diesem praktischen Anliegen hat der Phonembegriff eine seiner wesentlichen Wurzeln. Mit Recht verweist Daniel Jones darauf, daß in Sweets Unterscheidung von "broad" und "narrow Romic" der spätere Phonembegriff, wie ihn die Londoner Schule erarbeitet hat, bereits impliziert ist.¹⁶ Noch ein halbes Jahrhundert nach Sweet sieht K.L. Pike eines der wichtigsten praktischen Anliegen der mittlerweile etablierten Phonologie in der Entwicklung praktischer Orthographien für unverschriftete Sprachen.¹⁷

Nachgerade unabhängig davon vollzieht sich die Entwicklung des Phonems als einer Einheit der Systemlinguistik, die in der Phonologie der Prager Schule einen ersten Höhepunkt findet.

¹⁴Zur kontrastiven Linguistik allgemein siehe die Beiträge in dem Sammelband G. Nickel (Hrsg.), Reader zur kontrastiven Linguistik, Frankfurt/Main 1972, insbesondere zu phonetischen Problemen den Beitrag von C. James, "Zur Rechtfertigung der kontrastiven Linguistik", S. 21-38. - Zur kontrastiven Linguistik Deutsch-Englisch siehe jetzt E. Burgschmidt u. D. Götz, Kontrastive Linguistik Deutsch-Englisch; Theorie und Anwendung, München 1974 (zur Phonologie bes. S. 196-216).

¹⁵Siehe K. Kohler, "Anwendungsorientierte Phonetik. Sinn und Unsinn der kontrastiven Phonologie für den Fremdsprachenunterricht", Ling. Berichte, 26 (1973), 70-78.

¹⁶D. Jones, The History and Meaning of the Term "Phoneme", Supplement zu Le Maître Phonétique, London 1957, S.4.

¹⁷Vgl. den Titel seines Werkes Phonemics; A Technique for Reducing Languages to Writing, repr. Ann Arbor 1971.

Die Kluft, die seit Trubetzkoy zwischen der Sprechaktlautlehre oder Phonetik und der Sprachgebildelautlehre oder Phonologie klafft, konnte bis auf den heutigen Tag nicht überbrückt werden. In der Londoner Schule des praktischen Phonetikers Daniel Jones wie auch in den meisten phonologischen Theorien des amerikanischen Strukturalismus war die Notwendigkeit genauer Substanzbeschreibungen betont worden, wie sie etwa aus der wiederholten Forderung Bernard Blochs herausklingt, allein die phonetischen Ereignisse könnten die für den Phonologen relevanten Daten abgeben.¹⁸ Bezeichnend ist dabei die Erwartung, die Bloch an die Entwicklung der modernen Experimentalphonetik, insbesondere der akustischen Phonetik knüpft. - Hingegen ist für Trubetzkoy das phonologische System einer Sprache nur noch das abstrakte Korrelat der in der betreffenden Sprache vorherrschenden phonologischen Oppositionen, und der Phonologe ist nicht angewiesen auf genaue Substanzbeschreibungen, die das Feld der messenden, physikalisch-physiologischen (Instrumental-)Phonetik definieren.

Die größere Nähe zur phonetischen Substanz, die etwa die phonologischen Theorien der Nach-Bloomfield-Ära kennzeichnet, zeigt sich bei Bloch insbesondere in der Beschränkung auf den Idiolekt¹⁹ als Ausgangspunkt der phonologischen Analyse. Mag diese Beschränkung auch von der Mehrzahl der Phonologen bejaht werden, so ist doch allzu häufig und leichtfertig von dem phonologischen System des Deutschen, des Englischen, des Russischen geredet worden, was gegen die oben genannte Einschränkung verstößt. - Es fehlt auch bis auf den heutigen Tag etwa für das Deutsche an ausführlichen phonologischen Beschreibungen einzelner Dialekte, wie sie etwa Kufner für die Münchner und Heike für die Stadtkölner Mund-

¹⁸B. Bloch, "Phonemic Overlapping", in V.B. Makkai (Hrsg.), Phonological Theory; Evolution and Current Practice, New York 1972, S. 66-70 (S. 69).

¹⁹Siehe B. Bloch, "A Set of Postulates for Phonemic Analysis", in Makkai (1972), S. 167-199 (S. 170). Der Terminus "Idiolekt" ist Blochs Prägung.

art vorgelegt haben.²⁰ Genaue phonologische Beschreibungen regionaler wie auch soziolokaler Varianten sind aber die Voraussetzung für spezifische Aussprachelehren (des Englischen), wie sie Koziol wenigstens für jedes größere, halbwegs einheitliche Dialektgebiet des Deutschen gefordert hat²¹ und wie auch wir sie für wünschenswert halten.²² Die allein auf Kategorisierung der phonologischen Gegensätze ausgehende phonologische Theorie der Prager Schule hingegen schien eher darauf angelegt zu sein, den Vergleich phonologischer Systeme verschiedener Art zu ermöglichen - dies allerdings auf einem abstrakt-typologischen Niveau, dessen Entfernung von phonetischen Substanzmerkmalen beträchtlich ist. Es muß betont werden, daß keine der bis auf den heutigen Tag vorgelegten phonologischen Theorien von vornherein für die pragmatische Zielsetzung des Fremdsprachenunterrichts konzipiert worden ist.²³

²⁰H.L. Kufner, Strukturelle Grammatik der Münchner Stadtmundart, München 1961; G. Heike, Zur Phonologie der Stadtkölner Mundart; Eine experimentelle Untersuchung der akustischen Unterscheidungsmerkmale (=Deutsche Dialektgeographie, Bd. 57), Marburg 1964.

²¹H. Koziol, Die Aussprache des Englischen, Wien 1959, S. 29. - Zum Problem "Muttersprache/Dialekte - Fremdsprache" siehe auch E. Burgschmidt u. D. Götz, "Kontrastive Phonologie Deutsch-Englisch und Mundartinterferenz", Linguistik und Didaktik, 11 (1972), 209-225.

²²W.J. Barry hat mit Erfolg versucht, die Ansätze der herkömmlichen kontrastiven Phonologie zu überschreiten und herkunftsmäßige (dialektale) Unterschiede in den kontrastiven Ansatz mit einzubeziehen: siehe W.J. Barry, Perzeption und Produktion im subphonemischen Bereich; Eine kontrastive Untersuchung an intersprachlichen Minimalpaaren des Deutschen und Englischen (Linguistische Arbeiten, Bd. 15), Tübingen 1974. Die Ergebnisse dieser Arbeit liegen auch in Aufsatzform vor: W.J. Barry, "Language Background and the Perception of Foreign Accent", Journal of Phonetics, 2 (1974), 65-89.

²³Kohler (1973), 75. Siehe auch K. Kohler, "On the Adequacy of Phonological Theories for Contrastive Studies", in G. Nickel (Hrsg.), Papers in Contrastive Linguistics, Cambridge (Engl.) 1971, 83-88.

Für solche Zwecke wurden die bestehenden Theorien erst mit dem Heraufziehen der modernen Fremdsprachenunterrichts-Technologie expropriert - nicht ohne vorher gehörig verwässert worden zu sein. Hatte E. Haugen²⁴ noch auf den Charakter des Phonems als einer relationell definierten Einheit in einem spezifischen System hingewiesen, die das Phonem als eine Einheit bilingualer Beschreibung zumindest fragwürdig erscheinen lassen aufgrund des sui generis Charakters eines jeden Systems, das nur wenige perfekte formale Korrespondenzen mit anderen Systemen zeigt, so wurde insbesondere in der Nachfolge von R. Lado²⁵ die Methode des abstrakt-typologischen Sprachvergleichs auf Phonemebene institutionalisiert. Dies geschah mit durchaus unterschiedlicher phonetischer (allophonischer) Spezifikation der möglichen Gegensätze zweier in Kontakt miteinander tretender Systeme; fast nie wurden zudem zwei Systeme in ihrer Gesamtheit gegenübergestellt, sondern allenfalls Subsysteme, wobei das Kriterium von möglicher Interferenz zwischen Ausgangs- und Zielsprache oft arbiträr festgelegt wurde.

²⁴E. Haugen, "The Phoneme in Bilingual Description", Language Learning, 7 (1956-57), 3-4; 17-23.

²⁵R. Lado, Linguistics Across Cultures; Applied Linguistics for Language Teachers, Ann Arbor 1957, ist der erste Versuch, die Ergebnisse der strukturalistischen Linguistik für die Praxis des Fremdsprachenunterrichts zu verwerten. Die Methode ist vor allem in den vom "Center for Applied Linguistics" herausgegebenen Werken der Contrastive Structure Series weiter verfolgt worden, von denen sechs vorliegen (Englisch-Deutsch, Englisch-Spanisch, Englisch-Italienisch; je ein Band zur Phonologie und zur Grammatik).

Es soll nicht bestritten werden, daß die kontrastive (phonologische) Analyse die Heuristik für kontrastive Substanzuntersuchungen abgeben kann insbesondere dort, wo Substanzbeschreibungen von vornherein die phonologisch-kontrastive Analyse ergänzen.²⁶ Man darf aber nicht aus den Augen verlieren, daß zwischen den abstrakten phonologischen Beschreibungen auf der Ebene des Sprachgebildes oder -systems und der Substanz des konkreten Sprechaktes, wie er im fremdsprachlichen Unterricht gegeben ist, eine unüberbrückbare Kluft besteht. Das hohe Abstraktionsniveau phonologischer Beschreibungen; der statische Charakter von Phonemsystemen schlechthin; die Weigerung auch, die Vielfalt sprachlicher Schichtung regionaler wie sozialer Natur zur Kenntnis zu nehmen - alle diese Momente werden den spezifischen Modalitäten des Fremdspracherwerbs und der Dynamik von Lernprozessen wenig gerecht. Es mangelt empfindlich an kontrollierter Beobachtung tatsächlichen Lernverhaltens und experimentellen Untersuchungen der Art, wie Brière und Nemser²⁷ sie vorgenommen haben. Psychologische Aspekte von Lernverhalten, wie sie der konkrete Sprachgebrauch des Lernenden widerspiegeln, werden zumeist ebenso außer acht gelassen wie die spezifischen Modalitäten der Lernsituation; die Altersfunktion ist nahezu ebensowenig erforscht wie andere entwicklungspsychologische Variablen. Die Theorie des Fremdsprachenunterrichts bewegt sich allzu häufig im Fahrwasser anderer Disziplinen, deren spezifische Problemstellungen nicht mit ihren eigenen zur Deckung gebracht und deren Er-

²⁶Siene W.J. Barry u. F. Schindler, "Forschungstendenzen in der angewandten Phonetik", in G. Nickel (Hrsg.), Kongressbericht der 5. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) e.V. (= IRAL-Sonderband), Heidelberg 1974, S. 230-234.

²⁷E.J. Brière, A Psycholinguistic Study of Phonological Interference (= Janua Linguarum, Ser. minor, 66), Den Haag 1968; Nemser, An Experimental Study of Phonological Interference in the English of Hungarians; siehe auch W. Nemser, "The Predictability of Interference Phenomena in the English Speech of Native Speakers of Hungarian", in G. Nickel (Hrsg.), Papers in Contrastive Linguistics, S. 89-96.

gebnisse nur mittelbar in die eigene Theorie integriert werden können.²⁸

Darüber hinaus vernachlässigt die kontrastive Analyse solche Aspekte individuellen Lernverhaltens und seiner Variabilität, wie es nur eine Kontaktanalyse in dis z e n t i zutage fördern kann.²⁹ Auch die sog. Fehleranalyse³⁰ führt nur dort zu verlässlichen Ergebnissen, wo sie sich auf empirisch gesicherte Befunde, auf kontrollierte Beobachtung und psycholinguistische Experimente stützt.

Auf andere wesentliche Einschränkungen, denen die kontrastive phonologische Analyse unterliegt, soll hier nur kurz verwiesen werden: Die Beschränkung der meisten Phonologien auf das isolierte Wort hat zu einer Vernachlässigung des wichtigen Bereiches der satzphonetischen Erscheinungen (wie Elisionen, Assimilationen, schwache Formen) geführt, der noch weitgehend unerforscht ist,³¹ und nicht anders verhält es sich mit dem

²⁸Vgl. Heeschen u. Nehr (1974), 9.

²⁹W. Nemser u. T. Slama-Cazacu, "A Contribution to Contrastive Linguistics (A Psycholinguistic Approach: Contact Analysis)", Revue Roumaine de Linguistique, 15 (1970), 101-128 (118).

³⁰Zur Fehleranalyse siehe S.P. Corder, "The Significance of Learner's Errors", IRAL, 5 (1967), 161-170; dsb., "Idiosyncratic Dialects and Error Analysis", IRAL, 9 (1971), 147-159; dsb., "Zur Beschreibung der Sprache des Sprachlerner", in G. Nickel (Hrsg.), Reader zur kontrastiven Linguistik, S. 175-184; des weiteren die Beiträge in G. Nickel (Hrsg.), Fehlerkunde; Beiträge zur Fehleranalyse, Fehlerbewertung und Fehlertherapie, Berlin 1972. Siehe auch W. Nemser, "Approximative Systems of Foreign Language Learners", IRAL, 9 (1971), 115-123; L. Selinker, "Interlanguage", IRAL, 10 (1972), 209-231; dsb., "The Psychologically Relevant Data of Second-Language Learning", in P. Pimsleur u. T. Quinn (Hrsgg.), The Psychology of Second Language Learning, Cambridge (Engl.) 1971, S. 35-43.

³¹Siehe jedoch K. Kohler, "Contrastive Sentence Phonology", in Proc. III. AILA Congress (Copenhagen 1972), Vol. I, Heidelberg 1974, S. 129 ff.

weiten Feld der prosodischen Eigenschaften, das die kontrastive Analyse meist stillschweigend ausgespart hat. - Ein wesentliches Anliegen jeder phonologischen Kontrastanalyse ist die Vorhersage möglicher Fehler beim Lernenden und die Aufstellung einer Hierarchie der Schwierigkeiten, wie sie sich aus dem Systemvergleich der beiden miteinander in Konflikt tretenden Systeme ergibt.³² Gerade in dieser Hinsicht haben sich jedoch die Vorhersagen der kontrastiven Analyse häufig nicht erfüllt. Die jedem Phonetiker selbstverständliche Scheidung der Bereiche Perzeption und Produktion schließlich ist in keiner der herkömmlichen kontrastiven Analysen vollzogen worden, trotz mannigfaltiger Unterschiede in beiden Bereichen, die auch unterschiedliche Fehlervoraussagen notwendig zu machen scheinen.³³

In Ergänzung zu dieser Kritik der Voraussetzungen und Methoden der phonologisch-kontrastiven Analyse erscheint es angebracht, die Voraussagen der kontrastiven Analyse wenigstens in kurzen Zügen für diejenigen Probleme zu prüfen, die den Gegenstand dieser Untersuchung bilden.

1.3.0. Kontrastive Analyse und der Gegenstandsbereich dieser Untersuchung

Eine ausgearbeitete kontrastive Phonologie Deutsch-Englisch, die von strengen theoretischen Voraussetzungen ausgeht und detaillierte Fehlerprognosen macht, liegt nicht vor: Kufner (1971) ist stark von didaktischen Gesichtspunkten geprägt und verzeichnet nur (einen Teil der) bei deutschen Englischlernenden häufig beobachtete(n) Fehler, für deren Eintreten selten eine theoretische Rechtfertigung gegeben wird.

³²Siehe etwa R.P. Stockwell u. J.D. Bowen, The Sounds of English and Spanish (=Contrastive Structure Series), Chicago 1965, bes. S. 9-18 ("Sound Systems in Conflict: A Hierarchy of Difficulty").

³³Brière (1968), S. 72-74; Nemser, An Experimental Study of Phonological Interference, S. 141-142.

So konnte bislang keine plausible Begründung dafür angegeben werden, weshalb deutsche Englischlernende für /θ, ð / häufig [s, z] substituieren,³⁴ allenfalls noch [t, d], niemals (?) jedoch [f, v]. - Die bei Kufner (1971) wie auch schon bei Moulton³⁵ vorgenommene Unterscheidung von "phonemischen" und "phonetischen" Schwierigkeiten beim Erwerb einer Fremdsprache trägt wenig zur Erklärung der in der täglichen Praxis vorkommenden Aussprachefehler bei, suggeriert jedoch eine Hierarchie von phonemischen und (weniger gravierenden) phonetischen Fehlern, mit der wir uns nicht einverstanden erklären können. Daß "phonemische Korrektheit" als Ziel des Ausspracheunterrichts ein falsches und prinzipiell nicht realisierbares Konzept ist, hat Heike nachgewiesen, und wir teilen seinen Standpunkt, daß als Ziel grundsätzlich nur die "subphonemische Korrektheit" im Sinne einer Aneignung der Realisationsnormen der Fremdsprache in Frage kommen kann.³⁶ Daß die Substitution eigensprachlicher (d.i. deutscher) r-Laute für die englischsprachigen r-Artikulationen einen weniger gravierenden Aussprachefehler darstellen soll als die Substitution von (modifizierten) s-Artikulationen für /θ, ð/, ist nicht einzusehen;³⁷ in beiden Fällen handelt es sich um schwerwiegende Verstöße gegen englische Realisationsnormen, die Ab-

³⁴Siehe die Literaturangaben auf S. 2, Fußnote 6. - Es wird zumeist der Eindruck erweckt, als handle es sich bei diesen geläufigen Substitutionen um unmodifizierte s-Artikulationen; unsere eigenen Beobachtungen zeigen, daß in der Mehrzahl der Fälle der richtige Artikulationsort sowie die richtige Zungenkonfiguration ("ungrooved") wenigstens approximiert werden: siehe weiter unten, S. 107 ff.

³⁵W.G. Moulton, The Sounds of English and German (=Contrastive Structure Series), Chicago 1962; Kufners Buch besteht zu 60% aus - teilweise wörtlichen - Übertragungen oder Überarbeitungen von Moultons Text: siehe die Rezension von Kufner (1971) von W.O. Droyscher, Phonetica, 27 (1973), 111-114.

³⁶G. Heike, "Die Rolle der Phonetik im Ausspracheunterricht", in G. Nickel u. A. Raasch (Hrsgg.), Kongreßbericht der 3. Jahrestagung der Gesellschaft für Angewandte Linguistik GAL e.V. (IRAL-Sonderband), Heidelberg 1972, S. 1-13 (S. 4).

³⁷Burgschmidt u. Götz (1974), S. 201.

hilfe erfordern. - So ist es auch von geringer Bedeutung, daß im Bereich der englischen palatalen Kurzvokale eine zusätzliche phonologische Distinktion gegenüber dem System des Deutschen zu konstatieren ist, die gemeinhin das Argument für eine sorgfältige Einübung des /e/ - /æ/-Kontrastes liefern muß: Die Fachdidaktiker warnen vor Phonem-Zusammenfall und Verlust wichtiger phonologischer Distinktionen. Uns scheint dieses Argument von den wahren Problemen abzulenken, da der Verlust phonologischer Distinktionen in der tatsächlichen Kommunikation aufgrund der Redundanzeigenschaften natürlicher Sprachen praktisch nie eintritt, wo syntaktischer und situativer Kontext stets für die richtige Dekodierung beabsichtigter Phonemketten sorgen.³⁸ Wir vermuten auch, daß mit Schwierigkeiten bei der Aneignung akzeptabler Realisationsnormen nicht nur im Bereich der englischen Vokalphoneme /e/ und /æ/ (und erst recht nicht ausschließlich beim Erlernen des /æ/) zu rechnen ist, sondern die gesamte Reihe der englischen palatalen Kurzvokale betroffen ist; das Problem ist in unserer Untersuchung entsprechend ausgeweitet worden.

Die phonologische Kontrastanalyse legt schließlich nahe, daß für das letzte von uns untersuchte Ausspracheproblem - das der "final voiced consonants" - Lernschwierigkeiten nur für ein Glied solcher paradigmatischer Gegensätze, wie sie in Minimalpaaren wie /bid/ - /bit/, /seiv/ - /seif/ vorliegen, beobachtet werden; Wörter mit fortis-Endkonsonant(en) sollten für deutsche Englischlernende kein Problem darstellen.

Zur Klärung vieler der hier angeschnittenen Fragen sowie zum Zwecke der Erhebung empirischer Fehlerdaten für die genannten vier Ausspracheschwierigkeiten deutscher Englischlernender wurde eine Batterie von Perzeptions- und Produktionstests konstruiert; diese Tests, ihre Konstruktion, Anordnung und Durchführung werden im folgenden Kapitel dargestellt.

³⁸Vgl. Barry, Perzeption und Produktion, S. 6-8 sowie die dort angegebene Literatur.

2.0.0. Testkonstruktion, Testanordnung und Testdurchführung

2.1.0. "Informant"

Als Sprecher sämtlicher in den einzelnen Tests zur Beurteilung oder Nachahmung angebotenen Testwörter und -sätze diente ein englischer Informant süd-ostenglischer Herkunft, dessen Aussprache weitgehend frei von regionalen und soziolektalen Färbungen ist und uneingeschränkt als Norm "des (brit.) Englischen" schlechthin gelten kann. Der Informant ist in Ipswich, Suffolk, aufgewachsen und hat dort die ersten achtzehn Lebensjahre verbracht. Die Studienjahre (Studium der Linguistik und Phonetik sowie der romanischen Sprachwissenschaft) verbrachte er in Edinburgh (drei Jahre) und London (zwei Jahre); sie waren unterbrochen von einem einjährigen Frankreich-Aufenthalt. Der Informant verfügt über gute französische und deutsche Sprachkenntnisse.

2.1.1. Idiolekt

Nach eigenem Bekunden gehörten nur zwei der in den Testmaterialien geforderten Aussprachebesonderheiten nicht zu seinem Idiolekt: 1. die Realisation von intervokalischem /r/ als sog. "flapped r" ([ɾ]) - in unserem "R"-Test ohnehin eine marginale Erscheinung; 2. die Realisation des Halbvokals [j] in der Umgebung C_u (der Informant spricht normalerweise [nʲu:] anstatt [nɟu:], [stʲu:] anstatt [stɟu:]). Die korrekte Produktion der hier geforderten Varianten bereitete jedoch keinerlei Schwierigkeiten und wurde von anderen "native speakers" als uneingeschränkt akzeptabel bezeichnet.

Zwei Besonderheiten der Aussprache des Informanten, die nach eigenem Bekunden seine geographische Herkunft erkennen lassen, sind noch zu erwähnen: 1. Der Informant neigt dazu, /i/ in postvokalischer Stellung zu vokalisieren. Um unerwünschte Nebeneffekte auszuschalten, wurde er angehalten, postvokalisches /i/ vollkonsonantisch zu realisieren. - 2. Die Reihe der kurzen Palatalvokale /i, e, æ/, wie sie der Informant produziert, weist gegenüber der Reihe der entsprechenden RP-Vokale

eine geringfügige Verschiebung auf, die nicht ohne Bedeutung für die Ergebnisse der Hör- und Nachsprechttests gewesen sein mag, da die Reihe der englischen palatalen Kurzvokale eines der untersuchten Ausspracheprobleme darstellte. Aus diesem Grunde wurden anhand von Breitbandsonagrammen von 20 Einzelwörtern, die in den Produktionstests Verwendung fanden, Formantmessungen vorgenommen. Da die Vokale sowohl in initialer wie auch in medialer Position (zwischen Konsonanten) vorkamen und der konsonantische Kontext von Fall zu Fall variierte, war die Feststellung von einheitlichen Zielpositionen, denen sich die Vokalformanten nähern, nicht möglich;¹ alle Formantmessungen wurden deshalb willkürlich am Vokalmittelpunkt durchgeführt. Messungen wurden an den beiden ersten Formanten vorgenommen; die ermittelten Werte sind:²

- /i/ F₁ : 476 Hz (356); F₂ : 2034 Hz (2098)
- /e/ F₁ : 634 Hz (569); F₂ : 1805 Hz (1965)
- /æ/ F₁ : 792 Hz (748); F₂ : 1534 Hz (1746)

Die von uns gemessenen Werte für die ersten Formanten von /i, e, æ/ liegen sämtlich über den von Wells (1962) angegebenen Werten der RP-Vokale, während alle F₂-Werte niedriger als die von Wells angegeben sind. Bei dem Versuch, akustische Daten, wie sie in Sonograph-Aufzeichnungen vorliegen, artikulatorisch zu interpretieren, hat Joos³ einen Zusammenhang zwischen dem ersten Vokalformanten und der Höhe des artikulatorischen

¹Vgl. I. Lehiste u. G.E. Peterson, "Transitions, Glides, and Diphthongs", *JASA*, 33 (1961), 268-277.

²Wir geben in Klammern die von Wells gemessenen Werte der entsprechenden RP-Vokale an; vgl. J.C. Wells, *A Study of the Formants of the Pure Vowels of British English*, M.A. thesis, University of London, March 1962 (Die Arbeit hat mir nur in Auszügen vorgelegen).

³M. Joos, *Acoustic Phonetics* (= *Language Monograph No. 23*), Baltimore 1948, S. 49-59.

lierenden Zungenteils konstatiert, wobei höhere F_1 -Werte "größere Öffnung"⁴ bedeuten (d.h. Senkung der Zunge resultiert in höheren Werten für den ersten Formanten); für F_2 wurde eine Relation zur Bewegung der gesamten Zungenmasse im Mundraum (Verschiebung nach vorne oder hinten) festgestellt,⁵ wobei Erniedrigung des zweiten Formanten als durch Verschiebung der Zungenmasse in den hinteren Teil des Mundraums bewirkt gesehen wird. Bei vorsichtiger Interpretation unserer Messungen, wie sie allein aufgrund beträchtlicher Abweichungen zwischen einzelnen Werten für denselben Vokal geboten ist, kann gesagt werden, daß die drei palatalen Kurzvokale unseres Informanten geringfügig offener und zentraler realisiert werden, als dies in der RP-Norm, wie Wells sie angibt, der Fall ist.⁶ Mit dem auditiven Befund stimmt diese Interpretation gut überein. Auf die Feststellung, daß die Vokalrealisationen unseres Sprechers in keinem Fall als normverletzend oder auch nur "peripher" im Sinne einer Schulnorm des Englischen für Deutschlernende anzusehen sind, muß Wert gelegt werden.

⁴Delattre hat Joos' physiologische Interpretation leicht modifiziert, insofern er nicht die Zungenhöhe als allein entscheidend ansieht, sondern allgemeiner von "Öffnung" spricht ("meaning overall opening of the oral tract with definite relation to the width of the strictures at the main points of articulation, but not depending entirely on that"): siehe P. Delattre, "The Physiological Interpretation of Sound Spectrograms", *PMLA*, 66 (1959), 864-875.

⁵Vgl. auch Delattre (1959), 868-871; Delattre hat darüber hinaus die Abhängigkeit von F_2 vom artikulatorischen Parameter "Lippenrundung" festgestellt (größere Lippenrundung resultiert in einer Senkung der Werte für den zweiten Formanten). Dieser Parameter ist in unserem Fall nicht involviert.

⁶Vgl. jedoch die bei Barry, *Perzeption und Produktion*, S. 93 aufgeführten Formantwerte zweier RP-Sprecher für /i/, deren Durchschnittswerte ebenfalls stark von den bei Wells angegebenen Werten abweichen (Sprecher P: $F_1=491$ Hz, $F_2=1890$ Hz; Sprecher B: $F_1=511$ Hz, $F_2=2164$ Hz).

2.2.0. Aufnahme

Sämtliche Tonbandaufzeichnungen wurden in dem akustisch gedämpften Aufnahmestudio des sog. "Sechseckbaus" der Universität Kiel hergestellt. Als Aufnahmegerät diente eine NAGRA III der KUDELSKI NAGRA AG (Vollspuraufzeichnung bei 19 cm/sec., gelegentlich auch bei 38 cm/sec.). Als Mikrofon fand ein SENNHEISER Typ MD 421 N (Nierencharakteristik) mit Windschutz Verwendung, als Tonträger Studioband BASF LGR 30 mit günstigen Kopierdämpfungseigenschaften (-62 dB, gemessen auf REVOX A 77).

Der Informant war angehalten, alle Einzelwörter mit normal fallender Intonation zu sprechen und sich bei der Aussprache der Sätze der sog. "Normalintonation" zu bedienen.⁷ Hyperkorrekte Aussprache der Testwörter und -sätze wurde vermieden. Die Produktion des Informanten wurde vom Versuchsleiter ständig überwacht und protokolliert; einzelne Beispielwörter, die nach Ansicht des Versuchsleiters nicht einwandfrei der anzustrebenden Realisationsnorm genügten, wurden erneut aufgenommen und später an die Stelle der "fehlerhaften" gesetzt.

2.3.0. Herstellung der Testbänder

Die Originalaufnahme wurde vom Versuchsleiter nochmals über Kopfhörer abgehört und kritisch geprüft. Konnten keine weiteren Unzulänglichkeiten gefunden werden, wurden von den Originalbändern - z.T. nach umfangreichen Kopier- und Schneidarbeiten⁸ - Testbänder angefertigt (Wiedergabegeschwindigkeit einheitlich 19 cm/sec.).

⁷Siehe E. Sods, "Englische Satzintonation", *Phonetica*, 15 (1966), 129-215 (1967).

⁸Siehe unten, S. 30-31, 33.

2.4.0. "Kontrollgruppe" und Versuchspersonen

2.4.1. "Kontrollgruppe"

Die für die Perzeptionstests (Hördiskrimination und -identifikation: s.u., S.28ff.) hergestellten Testbänder wurden vor ihrer Verwendung in den Gruppentests mit deutschen Englischlernenden einer aus kompetenten Beobachtern und "native speakers" bestehenden Kontrollgruppe vorgespielt, die sich wie folgt zusammensetzte:

Untergruppe Ph (E): zwei englische Phonetiker mit Studienabschluß (Promotion) in Phonetik und Fremdsprachenkenntnissen in Deutsch bzw. Deutsch und Französisch;

Untergruppe Ph (D): zwei Phonetiker mit Deutsch als Muttersprache und sehr guten Englischkenntnissen (M.A. in Englisch als Haupt- bzw. Nebenfach) sowie Kenntnis des Französischen;

Untergruppe NSp: zwei "native speakers" ohne phonetische Ausbildung; es handelte sich um Studentinnen (Hauptfach: Geschichte bzw. Germanistik), die zu einem einjährigen Studienaufenthalt in Deutschland weilten und über gute deutsche Sprachkenntnisse verfügten.

Keine der sechs die Kontrollgruppe bildenden Versuchspersonen war mit der Problematik der Hörtests vertraut gemacht worden; es muß aber angenommen werden, daß die getesteten Ausspracheprobleme ihnen im Verlauf der sieben Einzeltests schnell durchschaubar wurden.

Die Durchführung der Hörtests mit den Angehörigen dieser Kontrollgruppe hatte zum Zweck, sicherzustellen, daß geschulte Beobachter und "native speakers" in der Lage sein würden, die ihnen gestellte Diskriminations- bzw. Identifikationsaufgabe

ohne Schwierigkeiten zu lösen. Es galt insbesondere, herauszufinden, inwieweit diese Gruppe sich fähig zeigte, Unterschiede zu diskriminieren, die als minimal nicht nur im phonologischen Sinne, sondern auch aufgrund allgemeiner perzeptorischer Gesetzmäßigkeiten anzusehen sind.⁹

Um ein gültiges Maß für die Einschätzung des Hör- und Diskriminationsvermögens der deutschen Englischlernenden zu gewinnen, wäre es angezeigt gewesen, englische Schüler und Schülerinnen der gleichen Altersgruppe derselben Testprozedur zu unterziehen. Die Möglichkeit zu solchen Paralleltests war leider nicht gegeben. Es kann jedoch angesichts des verwendeten Testformats sowie der Tatsache, daß die getesteten Distinktionen häufig von minimaler Art (im oben genannten Sinne) waren, kaum davon ausgegangen werden, daß eine unseren Probanden vergleichbare Gruppe von "native speakers" nur korrekte Response auf die dargebotenen Stimuli zeigen würde. Die Ergebnisse einer von Nemser zitierten Studie¹⁰ mit 150 Probanden - 75 einsprachige Amerikaner; 50 einsprachige und 25 zweisprachige Siamesen - zeigten, daß in Perzeptionstests des ABX-Formats Amerikaner und Siamesen eigensprachliche Unterschiede nur in 93 bzw. 83 % der Fälle richtig diskriminierten, fremdsprachliche Unterschiede hingegen in (immerhin) 81 bzw. 74% der Fälle. Die Fehlerraten der beiden "native speakers" unserer Kontrollgruppe, die ja ausschließlich eigensprachliche Unterschiede zu beurteilen hatten, müssen im Lichte dieser Zahlen gesehen werden.

⁹Siehe weiter unten, S.97 ff., bes. S. 100.

¹⁰J.T. Melamed, An Experiment in Sound Discrimination; Ph.D. dissertation, University of Indiana, 1961 (die Arbeit war mir nicht zugänglich); vgl. Nemser, An Experimental Study of Phonological Interference, S. 22-23.

Zu berücksichtigen ist schließlich auch, daß das in den Hörtests zur Beurteilung angebotene Material zu mehr als 70 % aus Nonsensewörtern (Logatomen) bestand,¹¹ die, wenngleich nach den Regeln englischer Phonetik gebildet, dennoch einen gewissen "Entfernungsgrad" von tatsächlichen englischen Wörtern aufwiesen. Untersuchungen von Greenberg und Jenkins¹² haben ergeben, daß Versuchspersonen englische Wörter in "ungewöhnlichen" Kombinationen oder Positionen als ebenso "fremd" beurteilen wie nichtenglische Laute, und in der Einschätzung der "psychologischen Distanz" von Nonsensewörtern der Art, wie sie in unseren eigenen Hörtests verwendet wurden, gegenüber englischen Wörtern zeigte sich ein deutlicher Schwelleneffekt für die englischen Wörter /stri:k/, /græs/, /spɛɪ/, /træk/ gegenüber solchen durch einfache Substitution in englische Lexeme überführbaren Logatomen wie /slæk/, /swit/, /kɪab/ und /bræd/: Die letzteren wurden von den Versuchspersonen bereits als in beträchtlicher Entfernung zum Englischen stehend eingestuft.¹³ - Die von der Kontrollgruppe abgegebenen Urteile in den Diskriminations- und Identifikationstests verdienen im Lichte dieser Untersuchungen eine kritische Überprüfung; vor allem ist das Verhalten der Untergruppe der "native speakers" genau zu prüfen, die im Gegensatz zu der Gruppe der geschulten Beobachter (Phonetiker) keinerlei Erfahrung mit Testwörtern der hier verwendeten Art besaß.

Die mit der aus geschulten Beobachtern und "native speakers" bestehenden Kontrollgruppe durchgeführten Tests hatten den weiteren Zweck, solche Testwortreihen auszuscheiden, bei denen die Mehrheit der Mitglieder der Gruppe (3 oder mehr) nicht

¹¹Siehe dazu im einzelnen weiter unten, S. 28-29.

¹²J.H. Greenberg u. J.J. Jenkins, "Studies in the Psychological Correlates of the Sound System of American English", *Word*, 20 (1964), 157-177.

¹³Greenberg u. Jenkins (1964), 161-164.

korrekt diskriminierte oder identifizierte. Die Festlegung eines solch willkürlichen Kriteriums geschah aufgrund von Überlegungen, die von vornherein nicht ausschließen wollten, daß der Perzeptionsfehler seinen Grund im Testmaterial selber hatte. Das Kriterium wurde nur ein einziges Mal erfüllt und führte zum Austausch der entsprechenden Testwortreihe.

Über das Abschneiden der Kontrollgruppe bei den einzelnen Hörtests wird weiter unten im einzelnen zu berichten sein.

2.4.2. Versuchspersonen

Als Vpp für die hier beschriebenen Untersuchungen dienten Schüler und Schülerinnen des fünften Schuljahres der Kieler Hauptschule "Ludwig-Richter-Schule" (Gruppe "LR") sowie Schüler und Schülerinnen des fünften Schuljahres des "Gymnasiums im Schulzentrum Mettenhof" in Kiel-Mettenhof (Gruppe "M"). Die Schüler(innen) beider Gruppen gehörten sämtlich den Geburtsjahrgängen 1963/64 an. Den einzelnen Hörtests unterzogen sich von der Gruppe "LR" bis zu 14, von der Gruppe "M" bis zu 39 Schüler(innen); vollständige Testunterlagen aus den sieben Einzel-(Hör-)Tests waren aber aufgrund ständiger Fluktuation bei der Gruppe "LR" und aufgrund von Krankheitsfällen bei der Gruppe "M" nur für 8 ("LR") bzw. 20 ("M") Vpp zu erhalten (es wurde davon abgesehen, versäumte Einzeltests nachzuholen).

Für die Produktionstests wurde unter den beiden Gruppen eine Auswahl getroffen, da eine vollständige Erhebung von Produktionsdaten bei allen Vpp einen unzumutbaren Zeit- und Arbeitsaufwand bedeutet hätte. Produktionsdaten wurden demnach nur von vier Angehörigen der Gruppe "LR" sowie von acht Angehörigen der Gruppe "M" erhoben.¹⁴

¹⁴zur Auswahl der Vpp für die Produktionstests siehe unten, S. 41ff.

Die Zusammensetzung der beiden Gruppen zeigt die folgende Aufstellung:

	Perzeption	Produktion
"LR"	8 (5 ♀ + 3 ♂)	4 (2 ♀ + 2 ♂)
"M"	20 (8 ♀ + 12 ♂)	8 (4 ♀ + 4 ♂)

Beide Gruppen weisen bezüglich der regionalen Herkunft ihrer Angehörigen ein Höchstmaß an Homogenität auf. Die Überwältigende Mehrheit der Schüler(innen) ist in Kiel oder der näheren Umgebung geboren und aufgewachsen; nur ein Schüler der Gruppe "M" ist am Niederrhein geboren und hat nach eigenem Bekunden dort auch die längste Zeit seines Lebens gewohnt. Keine(r) der Schüler(innen) verfügte über Kenntnisse in einer anderen Fremdsprache außer Englisch. Der Englischunterricht hatte für alle Vpp einheitlich mit dem Beginn des fünften Schuljahres begonnen; zum Zeitpunkt des Beginns dieser Untersuchung (Dezember 1973 bzw. Januar 1974) hatten die Schüler-(innen) vier bis fünf Monate englischen Unterricht gehabt mit einheitlich fünf Wochenstunden.

Der Einfluß des jeweiligen Fachlehrers auf die Leistungen beider Gruppen in den hier beschriebenen Hör- und Aussprachetests ist nicht untersucht worden, obwohl nicht zu bestreiten ist, daß ergänzende Untersuchungen zu diesem Aspekt dringend erforderlich wären; ihre Durchführung und genaue Kontrolle gestaltet sich jedoch meist sehr schwierig. - Es scheint aber notwendig, darauf hinzuweisen, daß der Englischunterricht der Hauptschüler (Gruppe "LR") nicht von einer Fachkraft wahrgenommen wurde, über die die Schule z.Zt. nicht verfügt. In keiner der beiden Gruppen von Vpp ist die Aussprache des Englischen jemals thematisiert worden; naturgemäß stehen der Erwerb der Lexik und elementarer grammatischer Strukturen im Vordergrund des Anfangsunterrichts. Nach eigenen Aussagen haben die Fachlehrer jedoch häufig die Aussprache ihrer Schüler korrigiert, ohne dabei jemals ausdrücklichen Bezug auf artikulatorisch/

motorische Bewegungsabläufe, Stellungen der Artikulationsorgane u.ä. zu nehmen: Korrekatives Eingreifen beschränkte sich stets auf Beispieligabe und Kontrolle der Schülernachahmungen.

Eine pauschale Einschätzung der englischen Aussprache der Schüler(innen) beider Gruppen nach wenigen Monaten Unterricht in der Fremdsprache kann sich nur auf die wenigen Daten stützen, die bei den zwölf Schülern gewonnen wurden, die zu Produktionstests mit sehr beschränktem Material herangezogen wurden; in den von den Probanden gesprochenen Testsätzen lag jedoch ein umfangreicheres Material vor als für die hier untersuchten Ausspracheprobleme zur Auswertung gelangte. Die Eindrücke lassen sich so zusammenfassen:

Das Bemühen der Schüler(innen) um Modifikationen der Artikulation in Richtung auf die fremdsprachliche Norm ist deutlich ausgeprägt, die Aneignung lautsprachlicher Erscheinungen - paradigmatischer wie syntagmatischer Natur - weitgehend vollzogen. Schwierigkeiten bereiten allerdings die zeitliche Koordination der motorischen (Artikulations-)Abläufe sowie die Integration von Äußerungen in akzeptable prosodische Konturen. Im übrigen ist ein hohes Maß an individueller Variation zu konstatieren, das jedes weitere generalisierende Urteil als unangebracht erscheinen läßt. Der Rahmen dieser Untersuchung ist zudem so abgesteckt, daß Aussagen, die über die hier untersuchten Ausspracheprobleme des Englischen im Unterricht für deutsche Anfänger hinausgehen, nicht erwartet werden sollten. Der Interpretation der Ergebnisse der einzelnen Hör- und Produktionstests bleibt jedoch vorbehalten, den Wert dieser Tests im Hinblick auf umfassendere Fragestellungen kritisch zu prüfen.

2.5.C. Durchführung der Hörtests

2.5.1. Kontrollgruppe

Die in sieben "Blöcke" aufgeteilten Hördiskriminations-

und Identifikationstests¹⁵ wurden mit den sechs Angehörigen der Kontrollgruppe individuell durchgeführt. Die Vpp erhielten kurze Anweisungen, das verwendete Testformat und ihre Aufgabe betreffend. Gegenstand der Untersuchung seien gewisse lautliche Unterschiede des Englischen; auf den konstruierten Charakter des in den Tests verwendeten Materials (Nonsense-Wörter) wurden die Vpp besonders hingewiesen. Die Untergruppe der geschulten Beobachter wurde darüber hinaus aufgefordert, Kritik an dem hier angewendeten Verfahren zu üben und ihre Urteile zu kommentieren; solche Kritik sollte zu weiteren Verbesserungen, gegebenenfalls auch zu durchgreifenden Änderungen des Testverfahrens führen.

Die Testbänder wurden den Vpp einzeln über ein Wiedergabegerät vom Typ REVOX A 77 K bei 19 cm/sec Bandgeschwindigkeit über Kopfhörer vom Typ BEYER DT 100 abgespielt. Den Vpp wurde freigestellt, die Wiedergabelautstärke selber zu regeln. Die Antworten wurden auf den gleichen vorbereiteten Fragebögen festgehalten, die auch in den Gruppentests verwendet wurden.

2.5.2. Gruppe "LR"

Als Kriterium für die Auswahl der Testgruppen hatte ursprünglich gegolten, die Tests nur an solchen Schulen durchzuführen, die über eine einsatzbereite Sprachlehranlage (HS- oder HSA-Labor) verfügten. Es stellte sich heraus, daß von den Hauptschulen im Einzugsbereich von Kiel einzig die Ludwig-Richter-Schule wenn auch nicht über ein Sprachlabor, so doch über eine vergleichbare Anlage verfügte, die es erlaubte, den Vpp die Testbänder über Kopfhörer zuzuspielen. Bei dieser Anlage handelte es sich um einen "AUDICORD Sprach- und Gruppentrainer", bestehend aus einem modifizierten AEG Tonbandgerät Typ Magnetophon 86 (zwei Wiedergabegeschwindigkeiten: 9,5 und 19 cm/sec) mit Induktionsschleife und Hör/Sprechmöglichkeit

¹⁵Zu den Tests im einzelnen siehe unten, S. 28ff.

(über batterieversorgte Kopfhörer mit individueller Lautstärkeregelung und rechtsseitig angebrachtem Mikrofon) für bis zu 22 Schüler. Die ca. 25 m lange Induktionsschleife war im Biologieraum der Schule fest verlegt (rundum in ca. 2,5 m Höhe verlaufend). Die Kopfhörer wurden vor jedem Einzeltest vom Versuchsleiter einzeln auf Funktion (Ladungszustand der Batterien und Lautstärkeregelbarkeit) überprüft; des Weiteren wurden die nicht benötigten Mikrophone nach oben abgehoben.

Die Probanden wurden so über die zur Verfügung stehenden Plätze verteilt, daß sie sich nicht hinderlich waren und in ihren Urteilen von Nebenmännern unbeeinflusst blieben.

Die Abhörbedingungen für diese Gruppe können nicht optimal genannt werden, da über die Induktionsschleife auch Außengeräusche sowie Unruhe im Klassenraum selbst übertragen wurden; einige der Probanden neigten zudem dazu, einzelne Testwörter spontan nachzusprechen. Diese Störungen erreichten jedoch nie ein solches Ausmaß, daß ein Abbruch der Versuche angezeigt schien; die höhere Störanfälligkeit dieser Gruppe im Vergleich zu der anderen Gruppe von Versuchspersonen sollte allerdings nicht aus dem Auge verloren werden und will bei der Interpretation der Ergebnisse berücksichtigt werden.

2.5.3. Gruppe "M"

Das Schulzentrum in Kiel-Mettenhof, in dem die Versuche mit der zweiten (zahlenmäßig größeren) Gruppe von Versuchspersonen durchgeführt wurden, verfügt über eine RINK-Sprachlehranlage (HS-Labor mit 32 Plätzen in einem akustisch behandelten Raum) und erfüllt somit das oben genannte Kriterium. Zu Beginn der Datenerhebung waren nur 18 Plätze funktionsfähig, was die Aufteilung der Klasse in zwei Gruppen notwendig machte. Bei einem ersten Versuch mit der ersten Halbgruppe stellte sich der desparate Zustand der Anlage vollends heraus:

Nachdem die Vpp bereits über schlechte Hörmöglichkeiten geklagt hatten, ergab eine Analyse der Ergebnisse dieses Tests eine durchschnittliche Fehlerrate von 35 % (gegenüber nur 22 % beim ersten Versuch mit der Gruppe "LR"). Das schlechte Abschneiden der (Halb-)Gruppe "M" wurde unmittelbar auf die Unzulänglichkeiten der Sprachlehranlage zurückgeführt; der Versuchsleiter entschloß sich daraufhin in einem weiteren Vorversuch (dieses Mal mit der gesamten Gruppe) zur Wiedergabe über Außenlautsprecher (REVOX-Gerätelautsprecher sowie ein zusätzlicher Hi-Fi-Lautsprecher ISOPHON HSP 6001; Nennbelastbarkeit 40 W; Übertragungsbereich 35-20 000 Hz). Die Ergebnisse dieses zweiten Vorversuchs wiesen eine deutliche Verbesserung der Hörleistungen auf (Rückgang der durchschnittlichen Fehlerrate auf 15 %); für alle weiteren Einzeltests wurde diese Art der Durchführung beibehalten, zumal die gute Raumakustik einwandfreien Empfang der abgestrahlten Signale gewährleistete.

Auch bei der Gruppe "LR" wurde bei der Durchführung des Höridentifikationstests (s.u.) zu dieser Art der Wiedergabe übergegangen, so daß für diese Tests die äußeren Bedingungen für beide Gruppen annähernd gleich waren.

Die im folgenden detailliert beschriebenen Tests wurden nach den jeweils notwendigen Vorversuchen mit den beiden Gruppen zwischen Dezember 1973 und Mai 1974 durchgeführt. In der Regel fand nur ein Test pro Woche statt (Ausnahme: Block 2 und 3 des Höridentifikationstests bei der Gruppe "M", die an zwei aufeinanderfolgenden Tagen durchgeführt wurden). Testzeit war für die Gruppe "LR" regelmäßig der späte Vormittag (~ 12.15 - 12.45 Uhr), für die Gruppe "M" hingegen der frühe Morgen (~ 9.00-9.30 Uhr; auch diese Tatsache muß bei der Interpretation der unterschiedlichen Ergebnisse beider Gruppen berücksichtigt werden. Die Reihenfolge aller Hör- und Produktionstests war für alle Vpp (einschließlich der Mitglieder der Kontrollgruppe, mit der nur Hörtests veranstaltet wurden) identisch. Die Tat-

sache, daß die Serie von Perzeptionstests (Hördiskrimination und -identifikation) den Produktionstests vorausging, darf nicht als Präjudiz für einen wie auch immer gearteten Vorrang von Perzeption gegenüber Produktion gewertet werden. Diese Reihenfolge wurde allein durch praktische Überlegungen diktiert. Bei umgekehrter Anordnung der Testfolge hätten die zu Produktionstest herangezogenen Schüler(innen) einen Vorteil gegenüber den anderen Probanden gehabt, was ihre Vertrautheit mit dem Testmaterial anbetrifft (vor allem mit den ungewohnten Nonsensewörtern); hier galt es, für alle Vpp gleiche Voraussetzungen zu schaffen. Es muß allerdings zugestanden werden, daß die Hörtestergebnisse auf die Auswahl der verringerten Zahl von Probanden, mit denen auch Produktionstests veranstaltet wurden, von Einfluß war.¹⁶

Obwohl jeder einzelne Hörtest vom Versuchsleiter unmittelbar nach der Durchführung des Tests ausgewertet und die Ergebnisse in Form von Fehlertabellen aufbereitet wurden, wurde das allgemein bekundete Verlangen der Versuchspersonen beider Gruppen nach Mitteilung der Ergebnisse nicht befriedigt, um einem möglichen Leistungsabfall bei einzelnen Probanden vorzubeugen. Die Schüler(innen) wurden stets darauf verwiesen, daß ihnen erst nach Durchführung sämtlicher Einzeltests ihr persönliches Abschneiden mitgeteilt werden würde, was schließlich auch geschah. Bei dieser abschließenden Eröffnung wurde beiden Gruppen vom Versuchsleiter pauschal ein unerwartet gutes Abschneiden bescheinigt; Wert wurde auch auf die Feststellung gelegt, wie nahe die Ergebnisse aller Versuchspersonen beieinander lägen, die eigentlich sämtlich die Benotung "gut" verdient hätten.

Die beobachtete Leistungsver schlechterung mehrerer Vpp der Gruppe "LR" in den späteren Phasen der Hörtestdurchführung (s. im einzelnen weiter unten, S. 41ff) veranlaßte den Ver-

¹⁶Siehe weiter unten, S. 41-43.

suchsleiter, bei den mit einer gewissen zeitlichen Verzögerung stattfindenden Tests mit der Gruppe "M" einen zusätzlichen Leistungsanreiz zu geben und einen Preis für die beste Einzelleistung auszusetzen (dies geschah jedoch erst vor dem letzten Block von Test I). Tatsächlich wurden nach Beendigung der Serie von sieben Einzel-Hörtests drei kleine Buchpreise verliehen. Es kann jedoch kaum davon ausgegangen werden, daß dieser kleine Anreiz das Testverhalten der Gruppe "M" nachhaltig beeinflusst hat.

2.5.4. Test I: Perzeption (Hördiskrimination)

Das für den aus vier Blöcken bestehenden Hördiskriminationstest verwendete Testformat wurde aus der schon mehrfach genannten Untersuchung von Nemser (1971) übernommen. Jeder der vier Blöcke bestand aus 80 (Block 4: 79) Testwortreihen aus je vier Nonsensewörtern (Logatomen), von denen eines minimal unterschiedlich bezüglich der getesteten Distinktion war.

Auf die Problematik der Verwendung von Nonsensewörtern in unseren Tests ist weiter oben bereits kurz eingegangen worden; sie muß an dieser Stelle noch einmal aufgenommen werden. Von zusätzlichen Fremdheitsmerkmalen, die als unkontrollierte Variablen in die Tests eingegangen seien, kann insofern keine Rede sein, als diese als "englisch" ausgegebenen Stimuli nicht von englischen Sprechern/Hörern beurteilt wurden,¹⁷ sondern von deutschen Englischlernenden, die zum Zeitpunkt der Untersuchung über einen minimalen Wortschatz in der Fremdsprache geboten. Insofern konnte

¹⁷ Dies trifft natürlich nicht auf die Angehörigen der "Kontrollgruppe" zu, die in der Mehrzahl aus "native speakers" bestand. Tatsächlich äußerte eine der Angehörigen der Untergruppe NSp, daß ihr die abgeforderte Entscheidung in den Fällen leichter gefallen sei, wo die zu diskriminierenden bzw. identifizierenden Laute in tatsächlichen englischen Wörtern vorkamen. Es handelt sich hierbei um die Vp CJ, die von allen Mitgliedern der Kontrollgruppe die schlechtesten Hörleistungen aufweist.

auch nicht von Bedeutung sein, daß etwa 70% der in den Tests verwendeten Stimuli, deren phonetische Beurteilung das Ziel der Untersuchung war, Nonsensewörter waren, die restlichen 30% hingegen tatsächliche englische Wörter (bzw. flektierte Formen von solchen), von denen mit Sicherheit angenommen werden konnte, daß sie den Schüler(inne)n unbekannt waren. Bei der Zusammenstellung des Testmaterials wurde dennoch auf größtmögliche Homogenität geachtet insofern, als mit wenigen Ausnahmen solche Stimuli "generiert" wurden, die nicht nur als generierbar im Sinne phonotaktischer Regeln des Englischen gelten, sondern darüber hinaus durch eine einzige Lautsubstitution in tatsächliche englische Wörter überführt werden können. Vermieden wurden auch solche Testwörter, die bei den Probanden zu unkontrollierbaren lautlichen wie auch lexikalisch-semantischen Assoziationen Anlaß geben konnten.¹⁸

Die Verwendung von Nonsensewörtern in dieser Untersuchung legte sich nicht nur deshalb nahe, weil solches Material beliebig manipulierbar ist und die Aufmerksamkeit stärker auf die phonetischen Details lenkt, die das Untersuchungsziel abgeben, sondern vor allem aus dem Grunde, daß lautliche Minimalpaare zu den untersuchten Ausspracheproblemen des Englischen im Wortschatz der hier getesteten Anfänger nicht in ausreichendem Maße vorhanden waren. Bei Untersuchungen mit verschiedenen Gruppen ist man im übrigen darauf angewiesen, bereits durch die Auswahl des Testmaterials weitgehend identische Voraussetzungen zu schaffen, wie es das von uns benutzte Material gewährleistet.

¹⁸ Kontrastive phonologische Analysen haben sich oft allein auf das Lautliche beschränkt und die Bereiche Lexik und Semantik ausgeschlossen, die (wie selbstverständlich auch die Schrift) die sprachlautliche Perzeption und Produktion in hohem Maße beeinflussen können: siehe H. Platt, "A Comparative Study of the Phonetics of Australian English and German", *Phonetica*, 21 (1970), 1-30; 75-106. Siehe auch U. Weinreich, "On the Description of Ironic Interference", *Word*, 13 (1957), 1-11 (2-3).

Die in den Hördiskriminationstests verwendeten Testwortreihen waren von der folgenden Form:

- (TH-Test) /bifts/ - /bifts/ - /bifθs/ - /bifts/
- (R-Test) /wɛntʃ/ - /rɛntʃ/ - /wɛntʃ/ - /wɛntʃ/
- (V-Test) /ɪmps/ - /ɛmps/ - /ɛmps/ - /ɛmps/
- (FVC-Test)¹⁹ /klɛns/ - /klɛnz/ - /klɛns/ - /klɛns/

Die Pausen zwischen den einzelnen Testwörtern hatten eine durchschnittliche Länge von ~ 1.6 sec., jede aus vier Testwörtern bestehende Reihe eine durchschnittliche Dauer von ~ 7.6 sec. Jede neue Reihe wurde durch einen kurzen 300 Hz-Ton angekündigt. Als Dauer für jeden der vier Blöcke ergibt sich damit eine Zeit von ~ 20 min.

Einige Bemerkungen bezüglich der Identität der drei "identischen" Stimuli in jeder Testwortreihe (siehe die Beispiele oben) sind angebracht. Die Platzierung des "unterschiedlichen" Wortes jeder Stimulusreihe sowie die Bestimmung der Abfolge der Reihen selber waren bereits vor der Aufnahme mit dem Informanten vorgenommen worden, so daß vollständige Stimulusreihen mit den entsprechenden Pausen zwischen den Stimuli und Reihen aufgezeichnet werden konnten. Korrekturen waren dann nur noch hinsichtlich des von Fall zu Fall notwendigen Austausches kompletter Stimulusreihen und Vereinheitlichung der Pausendauern notwendig. Die Tatsache, daß die drei (phonologisch) identischen Stimuli jeder Reihe ein gewisses Maß an (phonetischer bzw. subphonemischer) Variabilität aufwiesen, insofern es sich bei ihnen um drei verschiedene Produktionen des gleichen "Wortes" handelte, wurde dabei zunächst in Kauf genommen. Variabilität ist in allen Formen lautsprachlicher Kommunikation gegenwärtig; ihr vollständiges Ausschneiden in einer ohnehin artifiziellen Testsituation schien dem Versuchsleiter nicht angezeigt. Vereinzelt laut werdende

¹⁹FVC = final voiced consonants.

Einwände von Seiten einiger Mitglieder der Kontrollgruppe²⁰ bezüglich der zweifellos vorhandenen Variabilität innerhalb der Stimulusreihen führten dann jedoch zu einer Revision des Verfahrens. Zwar wurden nach wie vor komplette Stimulusreihen mit drei verschiedenen Produktionen des "identischen" Stimulus aufgenommen. Die Aufnahme wurde sodann vom Versuchsleiter abgehört; die drei "identischen" Stimuli wurden auditiv mit dem vierten - unterschiedlichen - Stimuluswort verglichen vor allem hinsichtlich solcher möglichst konstant zu haltenden Eigenschaften wie Akzent- und Intonationsmuster und gemeinsamer lautlicher (segmenteller) Merkmale. Das in diesen Eigenschaften am besten mit dem unterschiedlichen Stimulus übereinstimmende Stimuluswort wurde sodann zweimal kopiert und nach Herausschneiden der beiden auszuscheidenden Stimuluswörter an deren Stelle eingefügt, so daß die resultierenden Reihen aus vier Stimuli tatsächlich nur in der zu testenden Distinktion voneinander abwichen und die drei identischen Stimuli ein und dieselbe Produktion desselben Wortes darstellten. Bei den dadurch notwendigen umfangreichen Kopier- und Schneidearbeiten mußte mit besonderer Sorgfalt verfahren werden, da die Bandschnitte nicht hörbar sein durften und die kopierten Stimuli bezüglich des Intensitätspegels den originalen Stimuluswörtern durch entsprechend höhere Aussteuerung angeglichen werden mußten.

Nur der erste Einzeltest dieser Serie (Block 1) enthielt die ursprünglichen Reihen mit phonetisch geringfügig variablen "identischen" Stimuluswörtern;²¹ die Testbänder der drei weiteren Tests (Block 2-4) hingegen wurden nach dem geschicktesten zeitraubenden Kopier- und Schnittverfahren erstellt. Auf die damit möglicherweise gegebenen Unterschiede zwischen Block 1 und den anderen drei Blöcken wird weiter unten zurückzukommen sein.

²⁰Bezeichnenderweise nur Mitglieder der Untergruppen Ph(E) und Ph(O), nicht aber Angehörige der Untergruppe NSp.

²¹Aus 2 litgründen, aber auch aus testtechnischen Gründen wurde dieser Einzeltest nicht wiederholt.

Der Grund, weshalb Wortreihen aus vier Stimuli der Vorzug vor Stimuluspaaren oder Dreierfolgen gegeben wurde, liegt auf der Hand: Die hohe theoretische Zufallswahrscheinlichkeit für richtige Antworten bei Tests mit Stimuluspaaren (50%) oder Dreierfolgen (33,3%) stellt einen unerwünschten experimentellen Effekt dar und verlangt nach statistischer Korrektur der Ergebnisse. Bei Tests mit Vierer- oder gar Fünferfolgen ist eine solche Korrektur normalerweise nicht angezeigt.²²

Daß andererseits Tests der hier verwendeten Art andere unerwünschte experimentelle Effekte zeitigen können, wird insbesondere deutlich, wenn man die beträchtliche zeitliche Ausdehnung unserer Stimulusreihen bedenkt, die gewisse Gedächtniseffekte erwarten läßt. Das verwendete Testformat wird in dieser Hinsicht kritisch zu prüfen sein.²³

Die Vpp hatten die Aufgabe, auf vorbereiteten Antwortbögen, auf denen die Wortreihen durch Kästchen symbolisiert waren (siehe Abb. 1), das als unterschiedlich gehörte Wort an-



Abb. 1: Darstellung der Testwortreihen auf den Antwortbögen der Vpp (Hördiskrimination)

zukreuzen. Sie wurden ausdrücklich darauf hingewiesen, daß nur eines der Wörter in einer Reihe unterschiedlich war und jede Reihe über ein solches unterschiedliches Wort verfügte. Der somit ausgeübte Entscheidungszwang wurde fast stets befolgt (> 99,9% aller Antworten); wo dies nicht der Fall war, ergab sich daraus für die Auswertung dennoch kein Problem, in der zwischen "nicht diskriminiert" und "falsch diskriminiert" nicht unterschieden zu werden brauchte.

²²Vgl. D.P. Harris, Testing English as a Second Language, New York 1969, S. 23; 128-129.

²³Siehe weiter unten, Kapitel 3, bes. 3.5. (S. 64-67).

Alle Anweisungen vor dem Test wurden mündlich gegeben (es wurden keine Anweisungen auf die Testbänder selbst gesprochen); auf diese Weise schien dem Versuchsleiter besser gewährleistet, daß die Vpp ihre Aufgabe verstanden und bestehende Unklarheiten durch sofortiges Fragen beseitigten. Insbesondere erhielten die Vpp folgende Hinweise:

- Ziel dieser Hörtests sei es festzustellen, inwieweit sie (die Probanden) nach vier- bzw. fünfmonatigem Englischunterricht in der Lage seien, lautliche Unterschiede in der "neuen" Sprache wahrzunehmen. An Minimalpaaren, die nicht den Gegenstand der Untersuchung bildeten, wurde den Vpp demonstriert, was "lautliche Unterschiede" bedeutete; auf "gröbere" und "feinere" Unterscheidungsmöglichkeiten wurde in diesen Beispielen ausdrücklich Bezug genommen.²⁴
- Bei den ihnen über Kopfhörer zugespielten Wortreihen (das Testformat konnte an dieser Stelle unter Bezugnahme auf die anschaulich gestalteten Antwortbögen erklärt werden) handele es sich um englische Wörter, von einem "richtigen" Engländer gesprochen, die ihnen jedoch unbekannt seien (vom konstruierten Charakter des Testmaterials wurde somit abgelenkt).
- Alle Entscheidungen seien spontan und ohne langes Überlegen zu treffen. Jede Wortreihe enthalte ein und nur ein unterschiedliches Wort; sollte in einer Reihe kein Unterschied wahrgenommen werden, sei gegebenenfalls die Position des unterschiedlichen Wortes zu raten.

Diesen Anweisungen folgte eine Reihe von Beispielen; ergaben sich nach Aufforderung keine weiteren Fragen, wurde den Probanden das Testband zugespielt, das dann nur noch kurz angehalten wurde, wenn eine neue Seite des Testformulars um-

²⁴"gröbere" Unterschiede wurden anhand vokalischer Unterschiede demonstriert (z.B. /rao/ - /rao/ - /rao/ - /rao/), "feinere" Unterschiede an Reihen wie /dʒəos/ - /dʒəos/ - /tʃəos/ - /tʃəos/, /vɔ:lt/ - /vɔ:lt/ - /vɔ:lt/ - /vɔ:lt/.

geschlagen wurde; dies geschah vor allem, um zu überprüfen, ob sich alle Probanden auf gleicher Testhöhe befanden.

Um den Probanden die Möglichkeit zu geben, sich bei jedem Einzeltest an Aufbau und Form des Tests zu gewöhnen, wurden die ersten fünf Testwortreihen an das Bandende kopiert und dort erneut zur Beurteilung angeboten; den Vpp blieb dieses Verfahren verborgen.²⁵

Problemspezifische Hörtests (mit nur einem der untersuchten Ausspracheprobleme als Testgegenstand) hätten die zusätzliche Einführung zahlreicher Kontrollwörter und somit einen weitaus größeren Zeitaufwand erfordert, als vertretbar erschien. Es wurde deshalb die Form integrierter Hörtests gewählt, die alle vier untersuchten Problemlaute bzw. -lautverbindungen zum Gegenstand hatten. Die Vielzahl der Testwörter machte dennoch eine Aufteilung des Tests in vier Blöcke notwendig. Eine Übersicht über die vorgenommene Blockbildung und den Anteil der vier untersuchten Ausspracheprobleme gibt Tabelle 1 (siehe S. 35); danach sind die Testanteile wie folgt verteilt:

- TH-Test: 97 Testwörter (= 32%)
- R-Test: 58 " (= 19%)
- V-Test: 84 " (= 28%)
- FVC-Test: 60 " (= 20%)

Die Prozentquoten (= Anteile der einzelnen Ausspracheprobleme) sind für jeden der vier Blöcke nahezu gleich, die einzelnen Blöcke somit von gleicher Schwierigkeit. Da bei der Länge der einzelnen Testwortreihen mit der Wirksamkeit von "Positionseffekten" gerechnet werden mußte, wurde dafür gesorgt, daß dieser mögliche Effekt die einzelnen Testlaute in gleicher Weise betraf. Dies sei an einem Beispiel kurz erläutert: Noch vor der Zuweisung in einen der vier Blöcke wurde für die 60 Testwortreihen des FVC-Tests die Position des "unterschied-

²⁵Die Notwendigkeit bzw. Nützlichkeit einer solchen Einübungsphase wird weiter unten, S.68-69 kritisch geprüft.

Test	Kontext ²⁷	Block 1	Block 2	Block 3	Block 4	Σ
TH: /a/	1# _v	5	4	5	4	18
	2# _r	1	1	1	1	4
	3 C ¹ _s#	-	1	-	1	2
	4 V_s#	-	-	2	2	4
	5 V_t#	2	2	-	-	4
	6 V_#	2	3	2	2	9
	7 C ² _1_#	3	3	3	3	12
TH: /ø/	1# _v	5	4	5	4	18
	2 V_d#	1	2	1	-	4
	3 V_z#	1	-	1	2	4
	4 V_#	5	4	5	4	18
V	1# _C ³ ₁	6	6	6	6	24
	2 C ³ ₁ _C ³ ₁	15	15	15	15	60
R	1# _v	6	5	5	5	21
	2# C ¹ _v	3	4	4	4	15
	3# C ² _v	1	2	1	2	6
	4 V_v	4	4	4	4	16
FVC	1 V̄C ³ ₁ #	7	7	9	9	32
	2 V:C ³ ₁ #	3	3	5	4	15
	3 V̄VC ³ ₁ #	5	5	1	2	13
Σ		75	75	75	74	299

Tab. 1: Test I (Hördiskrimination): Blockbildung

lichen "Leitwörter" systematisch variiert, so daß je 15 Wortreihen für unterschiedlichen Stimulus in Position 1, 2,

²⁷Phonetische Kontexte werden mit den in der generativen Phonologie üblichen Konventionen dargestellt: Vgl. S.A. Fisiene, Generative Phonology, Englewood Cliffs 1973.

3 und 4 enthielten. Da jeder einzelne Block auch bezüglich des (möglichen) Positionseffekts von gleicher Schwierigkeit sein sollte, wurden auch die Positionsquoten im vorhinein festgelegt (siehe Tabelle 2).

Block	Position des "unterschiedlichen" Stimulus				Σ
	1	2	3	4	
1	19	19	19	18	75
2	19	19	18	19	75
3	19	18	19	19	75
4	19	19	19	17	74
Σ	76	75	75	73	299

Tab. 2: Positionsquoten (Test I)

Erst nach Festlegung dieser beiden Quoten erfolgte die Randomisierung der Darbietungsfolge der einzelnen Testwortreihen. Durch die Integration der vier Ausspracheprobleme in einen Test schien gewährleistet, daß ein erklärter Lerneffekt ausgeschaltet war. Ein möglicher Positionseffekt betrifft jedes der vier Ausspracheprobleme und jeden der vier Blöcke in gleicher Weise. Im übrigen verweisen wir auf Kapitel 3, wo das hier verwendete Testformat in allen seinen Auswirkungen einer kritischen Prüfung unterzogen wird.

2.5.5. Test II: Perzeption (Höridentifikation)

Das in der zweiten, aus drei einzelnen Blöcken bestehenden Serie von Hörtests verwendete Testformat ist ebenfalls anderweitig bereits erprobt.²⁸ Auch hier wurde konstruiertes Material (Nonsensewörter) derselben Art verwendet, wie es in der ersten Serie von Perzeptionstests benutzt worden war. Die drei Blöcke setzten sich aus je 70 Testwortreihen zusammen;

²⁸ Siehe S.M. Sapon u. J.B. Carroll, "Discriminative Perception of Speech Sounds as a Function of Native Language", *General Linguistics*, 3 (1958), 62-72; die Wortreihen in dieser Untersuchung bestanden allerdings aus 1 + 4 Wörtern.

Block 1 und 2 enthielten einen "Vorspann" von 5 bzw. 4 Reihen von Testwörtern, die am Ende des Testbandes erneut zur Beurteilung angeboten und nur dort gewertet wurden; im letzten Block wurde auf einen solchen Vorspann verzichtet.

Die individuellen Testwortreihen waren von folgender Art: Ein sog. "Mottowort", das den jeweiligen Problemlaut enthielt, wurde nach einer Pause von - 3 sec. Dauer von drei Vergleichswörtern gefolgt, von denen eines mit dem Mottowort identisch war.²⁹ Die Aufgabe der Vpp kann als Identifikationsaufgabe bezeichnet werden (Wiedererkennen und "matching" identischer Stimuli). Typische Beispiele für in den Einzeltests tatsächlich vorkommende Testwortreihen sind die folgenden Stimulusreihen (:: bezeichnet die Pause zwischen Mottowort und den zum Vergleich angebotenen Testwörtern):

- (TH-Test) /ðeəz/ :: /deəz/ - /veəz/ - /ðeəz/
- (R-Test) /grend/ :: /gwend/ - /glend/ - /grend/
- (V-Test) /slɪŋz/ :: /slenz/ - /slɪŋz/ - /sləŋz/
- (PVC-Test) /zɒlvz/ :: /zɒlfs/ - /zɒlfs/ - /zɒlvz/

Die Pausen zwischen den einzelnen Testwörtern der Vergleichsreihe hatten eine Dauer von - 1.4 sec., die durchschnittliche Reihenlänge betrug - 8.2 sec., während die Pausen zwischen den einzelnen Reihen mit - 6.7 sec. konstant gehalten wurden. Als durchschnittliche Testdauer ergibt sich damit eine Zeit von - 17.5 min. Der Beginn einer neuen Reihe wurde durch einen 300 Hz - Ton von kurzer Dauer angekündigt; zwischen Mottowort und den drei Vergleichswörtern wurde ein weiterer Ton (Frequenz: 1000 Hz) aufgebracht, der den Beginn der Vergleichsreihe ankündigte.

²⁹ Die Identität war insofern gegeben, als das mit dem "Mottowort" identische Vergleichswort eine Kopie des ersteren darstellte; auch bei der Herstellung der Testbänder zu diesem Test waren somit umfangreiche Kopier- und Schneidearbeiten erforderlich.

Um den Vpp eine Überprüfung zu ermöglichen, ob sie sich auf gleicher Testhöhe befanden, wurden außerdem die Ordnungszahlen für jede zehnte Stimulusreihe auf das Testband aufgesprochen.

Die Vpp hatten die Aufgabe, auf vorbereiteten Antwortbögen, auf denen die Wortreihen wie aus Abb. 2 ersichtlich dargestellt waren, das mit dem "Mottowort" identische Wort der

14 /XXXXX/ :

--	--	--

Abb. 2: Darstellung der Testwortreihen auf den Antwortbögen der Vpp (Höridentifikation)

Vergleichsreihe anzukreuzen. Alle Anweisungen vor dem Test wurden auch in diesem Falle mündlich gegeben; die Vpp wurden insbesondere auf die unterschiedliche Zusammensetzung der Testwortreihen (gegenüber der ersten Serie von Hörtests) hingewiesen, die in diesem Test jeweils drei anstatt einer einzigen Distinktion enthielten.³⁰ Besonderer Wert wurde darauf gelegt, daß die Probanden über jede Reihe ihr Urteil abgaben; die Betonung des Entscheidungszwanges war notwendig, da nur Identifikationsfehler eine eindeutige Auswertung gestatteten, bei Identifikationsmängeln ohne Urteilsabgabe hingegen keine eindeutige Feststellung perzeptorischer Verwechslungen möglich war. Der Entscheidungszwang wurde in diesem Test von allen Vpp ausnahmslos befolgt.

Nach einer Reihe von Beispielen, die nicht Gegenstand der hier untersuchten Problematik waren, und einer letzten Aufforderung zu Fragen wurde das Testband den Vpp zugespielt.

Auch in dieser Serie von Tests wurde auf problemspezifische Einzeltests, die die Einführung zahlreicher zusätzlicher Kontrollwörter notwendig gemacht hätte, verzichtet und stattdessen die Form integrierter Hörtests gewählt, die alle vier den Gegenstand der Untersuchung bildenden Problemlaute

³⁰Eine Ausnahme stellen die Wörter des FVC-Tests dar: siehe weiter unten, 5.4.4., und das Beispiel oben, S. 37.

bzw. -lautverbindungen zum Gegenstand hatten. Eine Übersicht über die Zusammensetzung der einzelnen Blöcke und die Anteile der vier untersuchten Ausspracheprobleme gibt Tabelle 4 (siehe S. 40). Danach sind die Anteile wie folgt verteilt:

TH-Test:	79 Testwörter	(= 39%)
R-Test:	36 "	(= 18%)
V-Test:	48 "	(= 24%)
FVC-Test:	38 "	(= 19%)

Die Zuweisung der Stimuli zu den einzelnen Blöcken und die Festlegung der Position des mit dem Mottowort identischen Wortes der Vergleichswortreihe erfolgte wie auch die Randomisierung der Darbietungsfolge wie in Test I (Hördiskrimination). Die Homogenität der drei Blöcke bezüglich der Anteile der vier Ausspracheprobleme und der Positionsvariablen war somit gewährleistet (siehe Tabelle 3).

Block	Position des identischen Stimulus			Σ
	1	2	3	
1	21	22	22	65
2	22	23	21	66
3	23	24	23	70
Σ	66	69	66	201

Tab. 3: Positionsquoten (Test II)

In einer Hinsicht jedoch ist die Forderung nach Homogenität dieser Testserie verletzt worden. Während die Vergleichswortreihen für den TH-, R- und V-Test aus je drei Stimuli bestanden, die sämtlich in einem entscheidenden phonetischen Merkmal voneinander unterschieden waren, enthielten die Vergleichswortreihen des FVC-Tests nur die den Gegenstand dieses Tests bildende Fortis/Lenis-Distinktion, so daß je zwei Stimuli jeder Vergleichsreihe miteinander identisch waren.³¹

³¹Siehe die Beispielsreihe weiter oben, S. 37. Identität

Test	Kontext	Block 1	Block 2	Block 3	Σ
TH: /θ/	1 # __v	5	5	5	15
	2 # __r	-	1	1	2
	3 v __t#	1	-	1	2
	4 v __#	3	3	3	9
	5 c __#	3	4	3	10
TH: /ð/	1 # __v	6	6	7	19
	2 v __d#	2	1	1	4
	3 v __z#	1	2	1	4
	4 v __#	5	5	4	14
V	1 # __c ₁ ³	4	4	4	12
	2 c ₁ ³ __c ₁ ³	12	12	12	36
R	1 # __v	6	6	6	18
	2 # c ₁ ¹ __v	4	4	4	12
	3 # c ₂ ² __v	-	1	1	2
	4 v __v	-	-	4	4
FVC	1 v̇c ₁ ³ #	7	6	5	18
	2 v:c ₁ ³ #	1	4	3	8
	3 v̇vc ₁ ³ #	5	2	5	12
Σ		65	66	70	201

Tab. 4: Test II (Höridentifikation): Blockbildung

Für die Vpp stellte dies zweifellos eine Erleichterung der Identifikationsaufgabe dar. Es wurde aus diesem Grunde zeitweilig erwogen, eine zweite Distinktion in diese Vergleichswortreihen einzufügen (z.B. durch Austausch eines Segmentes).

ist auch hier im strengen Sinne zu verstehen, d.h. bei der identischen Stimuli der Vergleichswortreihe handelte es sich jeweils um Kopien eines der beiden Stimuli.

Der Versuchsleiter gelangte jedoch zu der Einsicht, daß dies zu einer unzulässigen Vermengung der syntagmatischen und paradigmatischen Ebenen geführt hätte; auf der syntagmatischen Ebene der hier den Testgegenstand bildenden Distinktion stand jedoch keine weitere gleichgeartete Distinktion zur Verfügung, die sich problemlos in die Vergleichswortreihe hätte einbauen lassen können. Trotz gewisser Unterschiede dieser Testanteile gegenüber den anderen dreien wurde von der Form der integrierten Tests nicht abgegangen, da sie sonst gesondert hätten aufbereitet werden müssen; es gilt aber, den genannten Unterschied bei der Interpretation der Ergebnisse zu berücksichtigen. - Im Übrigen sei hier auf Kapitel 3 verwiesen, wo das hier verwendete Testformat und die möglichenfalls dadurch verursachten experimentellen Effekte einer kritischen Prüfung unterzogen werden.

2.6.0. Produktionstests

Eine Serie von Produktionstests verfolgte das Ziel, systematische Fehlerdaten bezüglich der hier untersuchten vier Ausspracheprobleme des Englischen, wie sie zuvor für den Perzeptionsbereich erhoben worden waren, auch für den Bereich der lautsprachlichen Produktion zu gewinnen. Es galt, unter den Angehörigen der beiden Gruppen "LR" und "M", die sich allen oben beschriebenen Gruppen-Einzeltests (Hördiskrimination und -identifikation) unterzogen hatten, eine Auswahl in der Weise zu treffen, daß die für die ausgewählten Probanden ermittelten Ergebnisse als hinreichend repräsentativ für die gesamte Gruppe angesehen werden können.

2.6.1. Auswahl der Probanden

Zwei Kriterien bestimmten die Auswahl der Versuchspersonen für die im folgenden beschriebene Serie von individuellen Produktionstests: 1. Da frühere Untersuchungen zur Diskriminationsfähigkeit und anderen sprachlichen Fähigkeiten bei Kindern häufig geschlechtsge- und unter-

schiede zutage gefördert haben,³² sollten aus jeder Gruppe gleiche Anzahlen von Jungen und Mädchen herangezogen werden. 2. Da das Verhältnis von Perzeption und Produktion für den Bereich der hier untersuchten Aussprache-probleme des Englischen untersucht werden sollte, lag es nahe, die Ergebnisse der Perzeptionstests als Kriterium heranzuziehen und eine entsprechend repräsentative Auswahl zu treffen. Zu diesem Zwecke wurden die absoluten Fehlerzahlen der einzelnen Probanden aus beiden Hörtests (also insgesamt sieben Einzeltests) addiert und den Versuchspersonen ("LR": 8; "M": 20) entsprechende Rangziffern zugeordnet (siehe Tab. 5, S. 43). Es wurde entschieden, daß die jeweils "besten" und "schlechtesten" Hörer beider Gruppen als Kandidaten für die Produktionstests nicht in Frage kamen, da ihre Hörtestergebnisse sich z.T. signifikant von denen der anderen Gruppenmitglieder unterschieden;³³ die weitere Auswahl wurde weitgehend mechanisch vorgenommen und war darüber hinaus von äußeren Gegebenheiten (Verfügbarkeit der Vpp zur anberaumten Testzeit; Ausfälle wegen Krankheit o.ä.) abhängig. So erklärt sich die Tatsache, daß aus der Gruppe "M" mehr Vpp mit niedrigen Rangziffern (d.h. relativ guten Hörtestergebnissen) zu den Produktionstests herangezogen wurden, während für die Gruppe "LR" ein leichtes Übergewicht der "schlechten" Hörer besteht.

³²Siehe etwa R.L. Politzer, "Developmental Aspects of Auditory Discrimination", in A. Valdman (Hrsg.), *Studies in Linguistics and Phonetics to the Memory of Pierre Delattre* (= *Janua Linguarum, Ser. maior*, 54), Den Haag 1972, S. 425-432 (dort auch weitere Literatur).

³³Dieses Verfahren legte sich auch deshalb nahe, weil die aus den Perzeptionstests mit dem besten Ergebnis hervorgegangene Vp der Gruppe "LR" über einen leichten Sprachfehler (Asigmatismus) verfügte und insofern für Produktionstests nicht in Frage kam; entsprechend war dann auch mit der Gruppe "M" zu verfahren.

Gruppe	Vp	♀♂	Fehler (Perz. tests)	Rang	Teilnahme Prod. tests
"LR"	PA	♀	79	1	
	WB	♂	89	2	X
	OK	♀	107	3	
	SS ₂	♀	117	4	X
	SS ₁	♂	125	5	
	SS ₃	♀	127	6	X
	SA ₁	♂	134	7	X
	SA ₂	♀	166	8	
"M"	WV	♀	26	1	
	RR	♂	40	2	X
	HU	♀	43	3	X
	GK	♀	45	4	
	BM	♀	46	5.5	X
	JK	♀	46	5.5	X
	WF ₂	♂	47	7	
	DL	♂	52	8.5	X
	KB	♀	52	8.5	
	SM	♂	57	10	X
	WJ	♂	62	11	
	BS	♂	63	12.5	
	KS	♂	63	12.5	
	DK	♀	64	14.5	
	HH	♀	64	14.5	X
	ST	♂	71	16	
	FT	♂	93	17	
	WF ₁	♂	98	18	
	DM	♂	100	19	X
	BT	♂	103	20	

Tab. 5: Perzeptionstests: Rangziffern

2.6.2. "Informant"; Aufnahme; Herstellung der Testbänder

Als Sprecher der in den einzelnen Produktionstests zur Nachahmung angebotenen Stimuluswörter und -sätze diente derselbe Informant wie in den Perzeptionstests. Die Aufzeichnung des Testmaterials geschah in der weiter oben (siehe S. 17) bereits geschilderten Weise. Identische Stimuluspaare (Test III) wurden durch Kopieren und Bandschneiden hergestellt.

2.6.3. Durchführung der Produktionstests

Sämtliche der im folgenden eingehend beschriebenen Produktionstests wurden individuell durchgeführt. Die Reihenfolge der Einzeltests war für alle Probanden die gleiche; Testzeit war für die Angehörigen der Gruppe "M" jeweils der frühe Vormittag (zwischen 9.00 und 10.30 Uhr), für die der Gruppe "LR" die Zeit zwischen 12.00 und 13.00 Uhr. Test III (Perzeption-Produktion: englische Nonsensewörter) und V (Produktion-Perzeption-Produktion: englische Sätze) wurden jeweils unmittelbar aufeinanderfolgend (mit ca. 5 Minuten Pause zwischen beiden Einheiten) durchgeführt, Test IV (Perzeption-Produktion: englische Einzelwörter) fand an einem anderen Tag statt. Stimuluswörter in Test III und IV wurden den Vpp über Kopfhörer vom Typ BEYER DT 100 zugespielt; als Wiedergabegerät wurde ein HiFi-Tonbandgerät REVOX A 77 K verwendet. Im Falle von Test V hörten die Vpp die Stimulussätze über die Außenlautsprecher des Wiedergabegerätes. Alle Schülerproduktionen wurden über ein hochwertiges Mikrophon (Typ: SENNHEISER MD 421 N mit Windschutz) bei 9.5 cm/s Aufnahmegeschwindigkeit auf Tonband (Studioband BASF LGR 30) monaural aufgezeichnet; als Aufnahmegerät fand eine weitere Tonbandmaschine REVOX A 77 K Verwendung. Da die Räumlichkeiten, in denen die Aufzeichnungen stattfanden, in keiner Weise akustisch gedämpft waren und ungünstige Halleigenschaften aufwiesen, war darauf zu achten, daß die Probanden in möglichst geringer Entfernung vom

Mikrophon (~ 10 cm) sprachen; der richtige Aussteuerungspegel wurde vor Beginn des Tests ermittelt. Da der Versuchsleiter die Probanden bei ihren Äußerungen ständig kontrollieren und ihre Produktion protokollieren mußte, oblag die Kontrolle und Überwachung der Tonbandaufzeichnung einem Assistenten, der die Aufnahmen über Kopfhörer hinter Band mithörte und ggf. Korrekturen an der Aussteuerung vornahm.³⁴ Die Qualität der Tonbandaufzeichnungen kann trotz der relativ schlechten Bedingungen, unter denen sie zustande kamen, als zufriedenstellend bezeichnet werden; für eine spätere auditive Beurteilung der Schülerproduktionen waren damit alle Voraussetzungen geschaffen.

Für sämtliche Einzeltests wurden für jeden einzelnen der Probanden im voraus Protokollunterlagen vorbereitet, die den Versuchsleiter bei der später vorgenommenen auditiven Beurteilung der Schüleräußerungen unterstützen sollten. Die Protokolle sollten eine Entscheidungshilfe vor allem dort liefern, wo eine rein ohrenphonetische Beurteilung häufig der Gefahr von Irrtümern und Verwechslungen unterliegt (besonders im Bereich der Beurteilung von TH-Realisationen bzw. möglicher Annäherungen an diese). Protokolliert wurden insofern insbesondere die sichtbaren Artikulationen im Bereich der Lippen und Vorderzunge; phonetische Transkriptionssymbole erlaubten eine schnelle Niederschrift, wie sie bei der raschen Abfolge der einzelnen Testwörter geboten war.

Die Probanden zeigten vor dem Versuchsaufbau und der für sie ungewohnten Testsituation keinerlei Scheu und waren eifrig bemüht, ihre Aufgabe gut zu erfüllen. Zeigte eine Versuchsperson zu irgendeinem Zeitpunkt des Tests Zeichen von An-

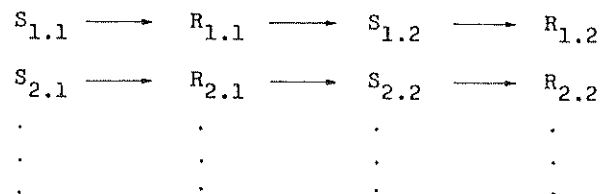
³⁴Herrn Mag.(grad.) J. Petersen vom Institut für Phonetik der Universität Kiel, der diese Funktion wahrnahm, spreche ich an dieser Stelle meinen Dank für die Anteilnahme aus.

spannung oder Ermüdung, hielt der Versuchsleiter das Testband an und sorgte für Auflockerung. Die relativ kurze Dauer der Einzeltests läßt die Wirkung von Ermüdungsfaktoren als unwahrscheinlich erscheinen. - Im folgenden werden diese Einzeltests ausführlich beschrieben.

2.6.4. Test III: Perzeption-Produktion: Nonsensewörter

Der erste der Produktionstests bestand aus 75 Paaren von Nonsensewörtern bzw. solchen englischen Wörtern, die den Vpp zum Untersuchungszeitpunkt unbekannt waren; die Konstruktion des Testmaterials erfolgte nach den gleichen Prinzipien wie bei den Hörtestmaterialien. Die Stimuli wurden den Probanden über Kopfhörer vorgespielt; die Vpp hatten die Aufgabe, die gehörten Wörter spontan nachzusprechen. Um Zufallseffekte bei der Dekodierung der Stimuli auszuschalten, wurde jedes Stimuluswort nach der ersten Schülerimitation ein weiteres Mal vorgespielt und von den Probanden wiederholt. Damit ergab sich die folgende Abfolge:

Perzeption - Produktion - Perzeption - Produktion



Die Pausendauer zwischen dem ersten Stimuluswort einer Folge und seiner Wiederholung betrug - 2.8 sec., die zwischen dem zweiten Wort einer Folge und dem ersten Stimuluswort der nächsten Folge ebenfalls - 2.8 sec. Als Gesamtdauer ergibt sich damit eine Zeit von - 8.6 min; da jedoch mindestens zwei Erholungspausen eingelegt wurden (von Fall zu Fall auch mehr), betrug die durchschnittliche Testdauer etwa 10-12 Minuten. Die relativ schnelle Abfolge der einzelnen Stimuluswörter verursachte den Probanden im übrigen keinerlei Schwierigkeiten, und keine der insgesamt zwölf Versuchspersonen zeigte

sich unfähig, sich dem Rhythmus des Testes anzupassen.

Anweisungen wurden vor dem Test mündlich erteilt. Wie vor den Hörtests wurde den Probanden mitgeteilt, daß es sich bei den Testwörtern um englische Wörter, von dem gleichen Engländer gesprochen, handele, daß die Wörter ihnen jedoch unbekannt seien. Als Aufgabe wurde ihnen erteilt, die über Kopfhörer gehörten Wörter so gut sie es vermochten nachzusprechen. Auf die schnelle Abfolge der Testwörter wurden sie ausdrücklich hingewiesen. Auch der Sinn der Wortwiederholung wurde den Probanden mitgeteilt, der darin bestünde, die eigene Aussprache gegebenenfalls zu korrigieren, falls sie ihre ersten Realisationen für verbesserungsbedürftig hielten.

Bereits bei der Konstruktion der Testwörter wurde das Prinzip verfolgt, mit einer möglichst geringen Anzahl von Testwörtern einen größtmöglichen Effekt zu erzielen. 51 der 75 Testwörter (= 68%) fungierten als Testwörter für jeweils nur eines der vier den Gegenstand der Untersuchung bildenden Aussprachprobleme (TH-, R-, V- oder FVC-Test), 22 Stimuli (= 29%) hingegen als Testwörter für gleichzeitig zwei der vier Einzelprobleme, während 2 Stimuli (= 3% der Gesamtheit) gar bezüglich dreier verschiedener Eigenschaften zu beurteilen waren. So war beispielsweise das Stimuluswort /ðens/ nur bezüglich einer Eigenschaft - des initialen /ð/ - relevant, während ein Stimulus wie /skraid/ sowohl als Testwort für den R-Test wie auch für den FVC-Test fungierte. Der 75 Stimuluspaare enthaltende Test konnte so das Material für nicht weniger als 202 Einzelbeurteilungen liefern. Die Anteile der einzelnen Aussprachprobleme verteilten sich dabei wie folgt:

TH-Test:	55	Testwörter	(= 28%)
R-Test:	58	"	(= 29%)
V-Test:	36	"	(= 18%)
FVC-Test:	52	"	(= 26%)

Auf die Art der Auswertung der auf Tonband aufgezeichneten Schülerproduktionen und die sich dabei ergebenden Probleme wird weiter unten eingegangen.

2.6.5. Test IV: Perzeption-Produktion: englische Einzelwörter

Auch bei Test IV handelt es sich um einen kurzen Nachsprechttest, der allerdings mit "echtem" englischen Material durchgeführt wurde, das den Schüler(inne)n beider Gruppen vertraut war. Die Auswahl der Testwörter mußte den Leistungsstand der Vpp berücksichtigen und erfolgte anhand des im Unterricht der Hauptschüler(innen) (Gruppe "LR") verwendeten Lehrwerkes.³⁵ Da die Gymnasialgruppe ein anderes Lehrbuch benutzte, mußte durch eine Gegenprobe sichergestellt werden, daß das Wortmaterial des Tests auch den Angehörigen dieser Gruppe vertraut war; in einzelnen Fällen war eine Revision des Testmaterials erforderlich.³⁶

Die Vpp hörten über Kopfhörer die ein- oder zweisilbigen Stimuluswörter und waren angehalten, diese spontan nachzusprechen; die Testwörter wurden ihnen (nur einmal) in einem konstanten Rhythmus (alle drei Sekunden ein Wort) vorgespielt. Einem ersten Testteil, der aus 66 Einzelwörtern bestand und 4.1 min. dauerte, folgte ohne Pause ein kurzer zweiter Teil mit zehn in rascher Abfolge aufeinanderfolgenden Sätzen (durchschnittliche Wortanzahl: 4.3); der Sinn dieses kurzen Zusatzteils lag darin, das Dekodierungsvermögen der

³⁵H. Friedrichs, Peter Pim and Billy Ball 1; Englischs Lehrwerk für Hauptschulen und sechsjährige Grundschulen, 26. Auflage Berlin 1970. Zum Zeitpunkt der Testdurchführung waren die Schüler(innen) auf S. 32 dieses Unterrichtswerkes angelangt.

³⁶Die Testwörter dieses in der Chronologie der Einzeltests letzten Produktionstests waren mit den Wörtern von Test V (s.u.) ungefähr zur Hälfte identisch; dort waren sie jedoch in kurze Sätze eingebettet und lagen zudem in verschrifteter Form vor.

Vpp bei längeren Äußerungen zu testen. Die Gesamtdauer des Tests betrug exakt 5.0 min. - Während die Probanden mit den Einzelwortstimuli dieses Tests keinerlei Schwierigkeiten hatten und die gehörten Wörter spontan imitierten, erwiesen sich die Sätze für fast die Hälfte der Vpp als schwierig und führten nur zu mangelhaften Imitationen; sie wurden deshalb nicht mit in die Wertung einbezogen.

Nach Beendigung des Tests wurden die Vpp befragt, ob sie in der Lage gewesen seien, sämtliche der ihnen vorgespielten Stimuluswörter zu identifizieren, die ihnen aus ihren Unterrichtserfahrungen vertraut sein mußten; die Frage wurde in nahezu allen Fällen bejaht.³⁷

Die 66 Stimuluswörter des Einzelworttests, die teilweise wiederum unter dem Gesichtspunkt der mehrfachen Verwendbarkeit³⁸ ausgewählt worden waren, waren Gegenstand von insgesamt 100 Einzelurteilen. Dabei waren die Anteile an den vier Ausspracheproblemen wie folgt verteilt:

TH-Test:	16	Testwörter	(= 16%)
R-Test:	23	"	(= 23%)
V-Test:	38	"	(= 38%)
FVC-Test:	23	"	(= 23%)

Probleme der Auswertung dieser Produktionsdaten werden weiter unten behandelt.

³⁷Angesichts der Ergebnisse dieses Tests, die auf einzelne Identifikationsfehler bzw. -mängel verweisen, muß die Verlässlichkeit dieser Schüleraussagen bezweifelt werden. Siehe im einzelnen weiter unten, bes. S. 115-117.

³⁸Siehe oben, S. 47. So konnte beispielsweise das Testwort bad als Testwort sowohl im V-Test als auch im FVC-Test fungieren und ein Testwort wie friend gar in drei der vier Tests (R-, V- und FVC-Test).

2.6.6. Produktion-Perzeption-Produktion: englische Sätze

Der auf Test III nach einer Pause von ca. 5 min. unmittelbar folgende Test V hatte wie Test IV tatsächliches englisches Wortmaterial zum Inhalt, das aus 35 kurzen englischen Sätzen mit durchschnittlich 5.4 Wörtern (Mindestzahl: drei Wörter; Höchstzahl: 11 Wörter) bestand. Die Auswahl der Testwörter und -sätze mußte den Leistungsstand der Vpp berücksichtigen und erfolgte nach den gleichen Prinzipien wie bei Test IV (s.o., S. 48). Der insgesamt gesehen spärliche Wortschatz der Probanden - vor allem im Hinblick auf die den Gegenstand dieser Untersuchung bildenden Ausspracheprobleme - machte es zudem notwendig, einige Wörter zu wiederholten Malen in die Testsätze einzubauen. Darüber hinaus fand hier das bereits im ersten Produktionstest verwendete Prinzip Anwendung, vor allem solche Testwörter einzusetzen, die mehrere lautliche Eigenschaften der fraglichen Ausspracheprobleme zugleich aufwiesen.

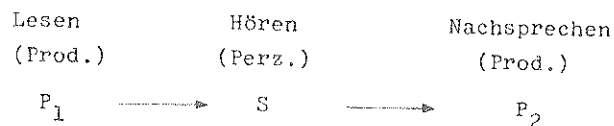
Von den insgesamt 188 Wörtern der 35 Stimulussätze fungierten insgesamt 49 als Testwörter, die bezüglich der untersuchten Problemlaute und -lautverbindungen 111 Beurteilungen ermöglichen; diese Zahl ist noch mit 2 zu multiplizieren, da die Versuchspersonen jeden Satz zweimal produzierten.³⁹ Die Anteile der Testwörter an den vier den Gegenstand des Testes bildenden Ausspracheproblemen waren dabei wie folgt verteilt:

TH-Test:	44	Testwörter	(=20%)
R-Test:	32	"	(=14%)
V-Test:	82	"	(=37%)
FVC-Test:	64	"	(=29%)

Die Vpp hatten die Aufgabe, die Testsätze von DIN A 6-Karten abzulesen; danach wurde ihnen der gleiche Stimulussatz, von

³⁹Auf die Unterschiede zwischen ersten und zweiten Schülerproduktionen und die sich daraus ergebenden Beurteilungsschwierigkeiten wird weiter unten, S. 181-182, eingegangen.

unserem englischen Informanten gesprochen, über die Außenlautsprecher des zweiten Tonbandgerätes zugespielt. Die Probanden wurden aufgefordert, sich diesen Beispielsatz anzuhören und ihn dann ein weiteres Mal zu produzieren. Alle Schülerproduktionen wurden auf Band aufgezeichnet. Damit ergab sich die folgende Abfolge:



Die Probanden konnten das Versuchstempo weitgehend selber bestimmen, da ihnen freigestellt wurde, sich die auf den Karten in orthographischer Form vorliegenden Sätze erst sorgfältig durchzulesen; individuelle Unterschiede traten hier denn auch reichhaltig zutage. Die durchschnittliche Versuchsdauer betrug etwa 22 min.; falls notwendig, legte der Versuchsleiter zwischendurch Erholungspausen ein.

Die Versuchsanordnung bereitete den Vpp keinerlei Schwierigkeiten; keiner der Probanden erwies sich als unfähig, die ihm gestellte Aufgabe zu erfüllen. Probleme der Auswertung der in den drei Produktionstests gesammelten Daten wie auch die Ergebnisse des Tests werden weiter unten in Kapitel 5 behandelt. Zum Abschluß dieses die Art der Datenerhebung beschreibenden Kapitels wird in der folgenden Tabelle 6 (siehe S. 52) ein Überblick über alle Tests gegeben. Die allen fünf Tests gemeinsamen phonetischen Kontexte, die die Grundlage für die Korrelation der Hör- und Produktionsfähigkeiten abgeben, sind mit * gekennzeichnet.

Test	Kontext	Perzeption		Produktion		
		Diskr.	Ident.	Nons.W.	engl.W.	engl.S.
TH: /θ/	*1 #_v	18	15	5x2	3	4(x2)
	2 #_r	4	2	3x2	1	1(x2)
	3 C ¹ _s#	2	-	-	-	-
	4 V_s#	4	-	-	-	-
	5 V_t#	4	2	1x2	-	-
	*6 V_#	9	9	3x2	2	2(x2)
	7 C ² _1#	12	10	3x2	1	-
TH: /ð/	*1 #_v	18	18	6x2	4	11(x2)
	2 V_d#	4	4	2x2	-	-
	3 V_z#	4	4	1x2	1	-
	4 V_#	18	14	4x2	1	-
	5 V_v	-	-	-	3	4(x2)
V: /e/	*1 #_C ³ ₁	8	4	2x2	1	1(x2)
	*2 C ³ ₁ _C ³ ₁	20	12	4x2	10	10(x2)
	1 #_C ³ ₁	8	4	2x2	-	-
	*2 C ³ ₁ _C ³ ₁	20	12	4x2	13	9(x2)
	*1 #_C ³ ₁	8	4	2x2	1	2(x2)
	*2 C ³ ₁ _C ³ ₁	20	12	4x2	13	19(x2)
R	*1 #_v	21	18	10x2	10	6(x2)
	*2 #C ¹ _v	15	12	11x2	10	9(x2)
	3 #C ² _v	6	2	6x2	-	-
	4 V_v	16	4	2x2	3	1(x2)
FVC	*1 V̂C ³ ₁ #	32	18	5x2(+8)	14	21(x2)
	*2 V:C ³ ₁ #	15	8	4x2(+8)	3	3(x2)
	*3 V̂VC ³ ₁ #	13	12	4x2(+8)	6	8(x2)

Tab. 6: Übersicht über alle Tests

3.0.0. "Testing the Test":

Kritische Prüfung der Perzeptionstests

Ein wichtiges Nebenanliegen dieser Arbeit sehen wir darin, die verwendeten Testverfahren auf ihre Leistung und Aussagekraft hin zu prüfen und die Ergebnisse solch kritischer Prüfung ausführlicher, als es sonst der Fall ist, darzustellen; wir hoffen damit, Entscheidungshilfen für die Testkonstruktion und -durchführung bei ähnlich gelagerten Untersuchungen liefern zu können.

Testformat, Testkonstruktion, -anordnung und -durchführung sind nur vereinzelt Gegenstand kritischer Untersuchungen gewesen,¹ obwohl die Art, in der Testmaterial konstruiert, angeordnet und dargeboten wird, beträchtlichen Einfluß auf die gewonnenen Ergebnisse ausüben kann. Solche Einflüsse müssen in ihren Wirkungen erkannt und möglichst genau eingeschätzt werden; insbesondere gilt es sicherzustellen, daß nicht zufällige Einflüsse für einen beobachteten experimentellen Effekt verantwortlich sind.²

In diesem Kapitel sollen Antworten auf die folgenden Fragen gegeben werden:

- Eignen sich die von uns durchgeführten beiden Arten von Hörtests (Test I und II: Hördiskrimination und -identifikation) grundsätzlich zur Erhebung von verlässlichen Befunden bei Vpp der hier untersuchten Altersgruppe?
- Treffen die Vpp ihre Entscheidungen frei von zufälligen Einflüssen?

¹Siehe aber E.J. Brière, "Phonological Testing Reconsidered", Language Learning, 17 (1967), 163-171. Die Literatur über Hörtests wird kritisch kommentiert bei Nemer, An Experimental Study of Phonological Interference, S. 11-35.

²Siehe L. Sachs, Statistische Auswertungsmethoden, 3. Aufl. Berlin 1972, S. 431.

- Bestehen Unterschiede zwischen beiden Arten von Perceptionstests (Hördiskrimination und -identifikation)?
- Gibt es Anzeichen für Leistungsunterschiede, die ihren Grund in der Geschlechtszugehörigkeit der Vpp haben?
- Lassen sich Leistungsunterschiede in der Abfolge von insgesamt sieben Einzeltests beobachten?
- Ist für beide Arten von Hörtests ein "Positionseffekt" zu verzeichnen?³
- Gibt es eine optimale Testdauer für Hörtests der von uns durchgeführten Art?

Zusätzlich zu jeder der bisher genannten Einzelfragen stellen wir die Frage:

- Welche Aussagen lassen sich über mögliche Unterschiede im Testverhalten der beiden Gruppen (Hauptschüler(innen): "LR", Gymnasiast(inn)en: "M") machen?

3.1.0. Zufallsfreiheit der Entscheidungen

Tabelle I (siehe Anhang, S. 233) enthält die absoluten Fehlerzahlen der Vpp beider Gruppen ("LR" und "M") sowie der Kontrollgruppe ("K") für Test I (Hördiskrimination: 4 Blöcke) und Test II (Höridentifikation: 3 Blöcke). Anhand der beobachteten Fehlerhäufigkeiten klären wir zunächst die Frage, ob die Vpp beider Gruppen bei einer statistischen Sicherheit von $S = 1 - \alpha$ ihre Entscheidungen frei von Zufällen getroffen haben.

3.1.1. Teststatistik

Wir stellen die beobachteten Fehlerhäufigkeiten in Form von Kontingenztafeln dar⁵ und prüfen auf Gemeinsamkeit der beiden

³Zur Bedeutung siehe im einzelnen weiter unten, 3.5.

⁴Alle statistischen Tests werden für die Irrtumswahrscheinlichkeit $\alpha = 5\%$ durchgeführt, d.h. unsere Aussagen gelten mit der statistischen Sicherheit $S = 95\%$.

⁵Siehe Sachs, Statistische Auswertungsmethoden, S. 357.

Grundgesamtheiten. - Kontingenztafeln (hier: Vierfeldertafeln) sind von der allgemeinen Form

	B ₁	B ₂
A ₁	n ₁₁	n ₁₂
A ₂	n ₂₁	n ₂₂

Die Anzahl der beobachteten Häufigkeiten der Merkmalkombinationen (A_i, B_j) sei n_{ij}; i, j = 1, 2. Die Frage lautet, ob der Anteil von A₁ bei B₁ und B₂ an den Grundgesamtheiten als gleich angenommen werden kann. Es gilt:

relative Häufigkeit schätzt Wahrscheinlichkeit

$$\hat{p}_1 = \frac{n_{11}}{n_{11} + n_{21}} \quad - \quad p_1$$

$$\hat{p}_2 = \frac{n_{12}}{n_{12} + n_{22}} \quad - \quad p_2$$

Die zu testende Hypothese lautet dann:

$$H_0 : p_1 = p_2 \quad (\text{Nullhypothese})$$

$$H_1 : p_1 \neq p_2 \quad (\text{Alternativhypothese})$$

Es wird ein Näherungsverfahren mit der Normalverteilung angewendet (gut für n > 40).⁶ Man bildet

$$n_1 = n_{11} + n_{21}$$

$$n_2 = n_{12} + n_{22}$$

$$n = n_1 + n_2$$

$$\bar{p} = \frac{n_{11} + n_{12}}{n}$$

⁶Vgl. Graf/Henning/Stange, Formeln und Tabellen der mathematischen Statistik, 2. Aufl., von K. Stange und H.-J. Henning, Berlin 1966, S. 89.

$$\left. \begin{aligned} x_1 &= n_{11} \pm \frac{1}{2} \\ x_2 &= n_{12} \pm \frac{1}{2} \end{aligned} \right\} \text{ je nachdem, ob + oder - einen Wert} \\ \text{naher an } \begin{Bmatrix} n_{1p} \\ n_{2p} \end{Bmatrix} \text{ ergibt.}$$

Die Prufgroe ist
$$\hat{u} = \frac{|x_1 n_2 - x_2 n_1|}{\sqrt{p(1-p) n_1 n_2 n}} \quad (1)$$

Den Zahlenwert fur den Schwellenwert $u_{1-\frac{\alpha}{2}}$ entnimmt man statistischen Tabellen; er betragt fur $\alpha = 5\%$: $u_{0,975} = 1.96$ (fur den zweiseitigen Test). Ist die Prufgroe groer als der Schwellenwert, so wird H_0 (bei entsprechender Irrtumswahrscheinlichkeit α) zugunsten von H_1 verworfen.

Als Zufallsbereich (gute Naherung fur $np(1-p) > 9$) fur die Anzahl der Fehlentscheidungen x bei insgesamt n Entscheidungen, wenn die Wahrscheinlichkeit p betragt, erhalt man

$$np - \left(\frac{1}{2} + u_{1-\frac{\alpha}{2}} \sqrt{np(1-p)}\right) \leq x \leq np + \left(\frac{1}{2} + u_{1-\frac{\alpha}{2}} \sqrt{np(1-p)}\right) \quad (2)^7$$

3.1.2. Zufallsschranken fur Test I und II

Die theoretische Fehlerwahrscheinlichkeit p betragt bei Test I (vier Entscheidungsmoglichkeiten pro Testwortreihe) $p = 0.75$ und bei Test II (drei Entscheidungsmoglichkeiten pro Reihe) $p = 0.667$. Fur die ersten drei Blocke von Test I ist die Anzahl der Testwortreihen $n = 75$, fur Block 4 $n = 74$, durch Einsetzen dieser Werte in Formel (2) erhalten wir als Zufallsbereich fur Test I

$$p = 0.75; \quad n_{1,2,3} = 75 : \quad 48 \leq x \leq 64 \\ n_4 = 74 : \quad 48 \leq x \leq 63,$$

fur Test II

$$p = 0.667; \quad n_1 = 65 : \quad 35 \leq x \leq 51 \\ n_2 = 66 : \quad 36 \leq x \leq 52.$$

⁷Vgl. Graf/Henning/Stange (1966), S. 47.

$$n_3 = 70 : \quad 38 \leq x \leq 54$$

Da die beobachteten Fehlerhaufigkeiten (vgl. Tab. I) ausnahmslos auerhalb dieser Zufallsintervalle liegen,⁸ kann mit 95% Sicherheit angenommen werden, da nicht das angenommene p gilt. Unsere beiden Tests erweisen sich somit als geeignete Instrumente, tatsachliche Fahigkeiten der Hordiskrimination und -identifikation bei unseren Probanden zu ermitteln.

3.1.3. Gruppenunterschiede (absolute Fehlerhaufigkeit)

Anhand der folgenden Vierfeldertafeln ermitteln wir mogliche Unterschiede zwischen beiden Gruppen von Vpp; die Zeilen enthalten die Anzahl der beobachteten absoluten Fehlerhaufigkeiten (F) sowie der richtigen Entscheidungen (R) jeweils fur alle Angehorigen der Gruppen "LR" und "M".

	"LR"	"M"
F	462	693
R	1930	5287
Σ	2392	5980

Test I

Tab. 7: Absolute Fehlerhaufigkeiten
(Test I: Hordiskrimination)

Wir setzen diese Werte in Formel (1) ein und erhalten $\hat{p} = 0.138$ sowie als Wert fur die Prufgroe $\hat{u} = 9.23 > u_{0,975} = 1.96$; damit ist nachgewiesen, da die Gruppe "LR" signifikant mehr Fehler als die Gruppe "M" macht. Die relativen Fehlerhaufigkeiten errechnen sich als $\hat{p}_{LR} = 19\%$ und $\hat{p}_M = 12\%$.

Tabelle 8 (siehe S. 58) enthalt die entsprechenden Fehlerhaufigkeiten fur Test II (Horidentifikation); durch Einsetzen der Werte in Formel (1) erhalten wir $p = 0.181$ und als Wert

⁸Nur eine der insgesamt 28 Vpp (SA₂/"LR") erzielt im dritten Block von Test II ein Ergebnis,² das sich der unteren Zufallsschranke $\hat{p}_{u} = 38$ nahert.

	"LR"	"M"
F	482	539
R	1126	3481
Σ	1608	4020

Test II

Tab. 8: Absolute Fehlerhäufigkeiten
(Test II: Höridentifikation)

für die Prüfgröße $\hat{u} = 14.53 > u_{0.975} = 1.96$; auch bei diesem Test besteht zwischen beiden Gruppen ein signifikanter Unterschied (bei relativen Fehlerhäufigkeiten von $\hat{p}_{LR} = 30\%$ und $\hat{p}_M = 13\%$).⁹

3.2.0. Unterschiede zwischen Test I und Test II

Eine andere Anordnung der beobachteten Fehlerhäufigkeiten, wie sie in den folgenden Tabellen 9 - 11 vorgenommen wurde, ermöglicht eine Beantwortung der Frage, ob zwischen beiden Arten von Hörtests ein Unterschied besteht und einer der beiden Tests sich möglicherweise als schwieriger erweist. In Tabelle 9 sind die Falsch/Richtig-Entscheidungen der Hauptschüler(innen) (Gruppe "LR") für beide Tests getrennt

	Test I	Test II
F	462	482
R	1930	1126
Σ	2392	1608

"LR"

Tab. 9: Absolute Fehlerhäufigkeiten
(Test I und II); Gruppe "LR"

aufgeführt. Bei Einsetzen in Formel (1) erhalten wir für

⁹ Daß unsere Kontrollgruppe signifikant weniger Fehler als die Gruppe "M" macht, ist an den absoluten Fehlerzahlen der Tab. I unmittelbar abzulesen und erfordert keinen weiteren Nachweis.

$\bar{p} = 0.236$ und als Wert für die Prüfgröße $\hat{u} = 7.75 > u_{0.975} = 1.96$; Test II erweist sich damit als signifikant schwerer als Test I. Die Prüfung für die Gruppe "M" (siehe Tabelle 10)

	Test I	Test II
F	693	539
R	5287	3481
Σ	5980	4020

"M"

Tab. 10: Absolute Fehlerhäufigkeiten
(Test I und II); Gruppe "M"

ergibt den gleichen Befund ($\bar{p} = 0.123$; $\hat{u} = 2.68 > 1.96$), obgleich der Effekt hier nicht so deutlich ist.

Während sich somit Test II für beide Gruppen von Vpp gegenüber Test I - der Diskriminationsaufgabe - als signifikant schwerer erweist, tritt bei den Angehörigen unserer Kontrollgruppe der entgegengesetzte Effekt ein (siehe Tabelle 11).

	Test I	Test II
F	56	22
R	1738	1184
Σ	1794	1206

"K"

Tab. 11: Absolute Fehlerhäufigkeiten
(Test I und II); Gruppe "K"

Wir erhalten $\bar{p} = 0.026$ und $\hat{u} = 2.27 > 1.96$. Für die Angehörigen dieser Gruppe erweist sich damit Test II gegenüber Test I als signifikant leichter.

Da die Testwortreihen in Test II nur geringfügig länger waren als in Test I (- 8.2 sec. gegenüber - 7.6 sec.), kann kaum davon ausgegangen werden, daß dieser geringe zeitliche Unterschied für den bei den Gruppen "LR" und "M" beobachteten

Effekt verantwortlich zu machen ist. Wenn zudem ausgeschlossen werden kann, daß allein zunehmende "Testmüdigkeit" gegen Ende der Serie aus sieben Einzeltests (Blöcken) den Effekt bewirkt hat (siehe unten, 3.4.0.), muß angenommen werden, daß der signifikante Unterschied in der Schwierigkeit beider Tests seine Ursache in der Art der gestellten Aufgabe selbst hat: Die Diskriminationsaufgabe bei Anwesenheit nur einer einzigen Distinktion in jeder Testwortreihe erweist sich als leichter als die Identifikationsaufgabe ("recognition and matching") bei drei Distinktionen in der Vergleichswortreihe.

3.3.0. Einfluß der Geschlechtszugehörigkeit

Kontrollierte Tests sprachlicher Fähigkeiten bei Kindern und Heranwachsenden haben häufig den Nachweis für Leistungsunterschiede in Abhängigkeit von der Geschlechtszugehörigkeit erbracht: Mädchen schneiden bei solchen Untersuchungen gewöhnlich besser ab als Jungen.¹⁰ Die mögliche Wirksamkeit dieser Variablen bei den hier durchgeführten Hörtests soll im folgenden untersucht werden.

Die entsprechenden Fehlerdaten, die aus Tabelle I (siehe Anhang, S. 233) gewonnen wurden, werden in den Tabellen 12 und 13 wiederum in Form von Kontingenztafeln dargestellt, und zwar für beide Tests getrennt. Für Test I (Diskrimination)

	♀	♂
F	482	673
R	3405	3812
Σ	3887	4485

Test I

Tab. 12: Absolute Fehlerhäufigkeiten (alle ♀ und alle ♂ beider Gruppen); Test I

¹⁰Vgl. Politzer (1972) sowie die dort angegebene Literatur. Siehe auch H. Winitz u. M. Lawrence, "Children's Articulation and Sound Learning Ability", JSHR, 4 (1961), 259-268.

ergibt sich bei Einsetzen in Formel (1) $\bar{p} = 0.138$, für die Prüfgröße $\hat{u} = 3.41 > 1.96$ und damit ein signifikanter Unterschied: Jungen machen (signifikant) mehr Fehler als Mädchen ($\hat{p}_\eta = 12\%$, $\hat{p}_\sigma = 15\%$). - Die gleiche Prüfung für Test II

	♀	♂
F	498	523
R	2115	2492
Σ	2613	3015

Tab. 13: Absolute Fehlerhäufigkeiten (alle ♀ und alle ♂ beider Gruppen); Test II

ergibt für $\bar{p} = 0.181$ und für die Prüfgröße den Wert $\hat{u} = 1.63 < 1.96$; eine Geschlechtsabhängigkeit ist somit für Test II nicht nachzuweisen. - Faßt man die Ergebnisse beider Tests zusammen, erweist sich der Effekt erneut als nicht wirksam.¹¹ In den von uns durchgeführten Hörtests unterscheiden sich die Leistungen von Mädchen und Jungen nicht signifikant.

3.4.0. Testverhalten in Abhängigkeit von der Gesamttestzeit

Abbildung 3 stellt das Testverhalten der beiden Gruppen von Vpp über die Gesamttestzeit von ca. sieben Wochen dar, wobei die Unterschiede zwischen beiden Gruppen, die wir bereits als signifikant ermittelt haben (siehe 3.1.3.), unmittelbar ins Auge fallen (siehe S. 62). Mit den Werten von Tabelle II (siehe Anhang, S. 234), die die Fehlerzahlen der Vpp beider Gruppen in Prozent ausdrückt, wäre eine Auswertung als "Split-plot-plot-Versuchsplan"¹² durchzuführen, die alle möglicherweise wirksamen Effekte (Gruppenzugehörigkeit, Geschlechtsabhängigkeit, Einfluß der Zeitvariablen) testet. Wegen der ungleichen Zellenbesetzung ist eine solche SPP-Auswertung

¹¹ $\bar{p} = 0.155$, $\hat{u} = 1.42 < 1.96$.

¹²Siehe etwa B.J. Winer, Statistical Principles in Experimental Design, 2. Aufl. New York 1971, S. 366-371.

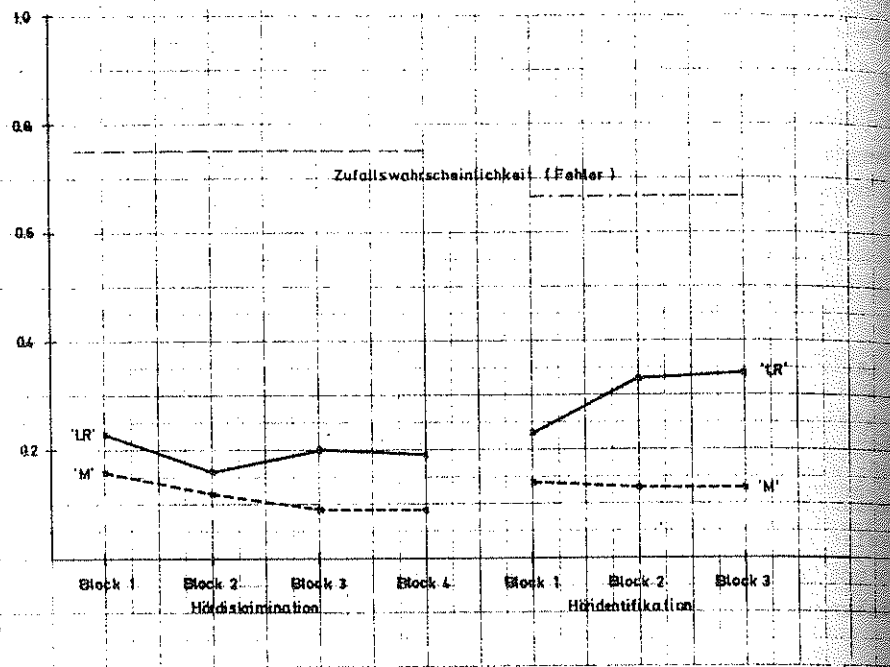


Abb. 3: Testverhalten der beiden Gruppen über die Gesamttestzeit (Fehlerhäufigkeit/Block)

jedoch nicht möglich.- Faßt man Jungen und Mädchen zusammen (s.o., 3.3.0.) und führt die Auswertung für beide Gruppen getrennt durch, so erhält man eine einfache Blockvarianzanalyse, die nichtparametrisch nach FRIEDMAN getestet wird.¹³ Man erhält bei beiden Gruppen signifikante Unterschiede der Fehlerraten in den einzelnen Blöcken.¹⁴ Mithilfe der multiplen

¹³Siehe Sachs, Statistische Auswertungsmethoden, S. 422-426. Bei diesem Test handelt es sich um eine Rangvarianzanalyse zum Zweck des verteilungsunabhängigen Vergleichs mehrerer abhängiger Stichproben von Meßwerten.

¹⁴Eine solche Analyse ist selbstverständlich nur sinnvoll und erlaubt Schlüsse auf potentiell wirksame Einflüsse

Vergleiche nach WILCOXON und WILCOX¹⁵ ergeben sich für die Größenordnung der Blöcke die Anordnungen

"LR": $D_2 < D_4 < D_3 < D_1 < Id_1 < Id_2 < Id_3$ ¹⁶

- Fehlerhäufigkeit +

"M": $D_4 < D_3 < D_2 < Id_2 < Id_3 < Id_1 < D_1$

In beiden Fällen unterscheiden sich die beiden "kleinsten" (mit der geringsten Fehlerhäufigkeit) und größten Blöcke voneinander signifikant (zwischen den Blöcken in der Überlappungszone und allen anderen treten keine signifikanten Unterschiede auf); die Größenordnung ist jedoch für beide Gruppen eine andere.

Bei der Gruppe der Hauptschüler(innen) ("LR") kann der Effekt der Trennung von Diskriminations- und Identifikationstests in der bereits nachgewiesenen (s.o., 3.2.0.) größeren Schwierigkeit der zweiten Testart gegenüber der ersten begründet sein. Die beobachtete Abfolge jedoch und insbesondere das Fehlen eines deutlichen "Schwelleneffekts" für Id_1 - den ersten Block einer neuen Testart, der sich als fehleranfälliger als die späteren Id_2 und Id_3 hätte erweisen sollen¹⁷ - legen den Schluß nahe, daß eine gewisse Testmüdigkeit gegen Ende der rund siebenwöchigen Testperiode den Effekt in erheblichem Maße mitbewirkt hat (siehe auch Abb. 3, oben S. 62). Die Beobachtungen des Versuchsleiters während der Durchführung der Einzeltests mit dieser Gruppe ließen einen solchen Effekt in den Ergebnissen erwarten.

der Gesamttestdauer nur dann, wenn von einer Homogenität der einzelnen Blöcke in beiden Tests ausgegangen werden kann. Dies wurde bei der Testkonstruktion berücksichtigt: siehe weiter oben, S. 34-36; 39-41.

¹⁵Siehe Sachs, Statistische Auswertungsmethoden, S. 426-429.

¹⁶D = Diskrimination; Id = Identifikation.

¹⁷Vgl. die Abfolge bei "M"; für D_1 läßt sich dieser Schwelleneffekt auch bei "LR" beobachten.

Bei der Gruppe "M" zeigt sich ein deutlicher Schwelleneffekt für die jeweils ersten Blöcke jeder Testart (D_1 und Id_1); der Effekt ist für den Eingangstest der gesamten Serie von sieben Testblöcken (D_1) noch deutlicher als für den ersten Block des (schwierigeren) Identifikationstests. Eine gegen Ende der Testserie zunehmende Testmüdigkeit hatte der Versuchsleiter bei dieser Gruppe nicht feststellen können; ein solcher Effekt läßt sich aus der beobachteten "Größenordnung" der einzelnen Blöcke auch nicht ablesen. Die Abfolge erlaubt hingegen, die Wirksamkeit eines "Lerneffektes" bei dieser Gruppe zu konstatieren, der für die zeitliche Abfolge der vier Blöcke des Diskriminations-tests am deutlichsten ist: $D_1 > D_2 > D_3 > D_4$, wobei die Ergebnisse der Blöcke 3 und 4 signifikant besser ausfallen als die des ersten Blocks.

Das Bild von einer insgesamt ausgeglicheneren und konstanteren Leistungskurve der Gruppe "M" gegenüber der Gruppe "LR", wie es Abb. 3 zeigt, wird in der statistischen Analyse bestätigt. Bei der Gruppe der Hauptschüler sind Konzentrationsmängel und eine gegen Ende der Testserie zunehmende Testmüdigkeit für die schlechteren Ergebnisse mitverantwortlich zu machen. Die Wirkung solcher Faktoren bei der Gruppe "M" kann weitgehend ausgeschlossen werden.

3.5.0. "Positionseffekt"

Die durchschnittliche Länge der einzelnen Testwortreihen, die wir oben mit 7.6 s (Hördiskrimination) angegeben haben, gibt zu bedenken, ob die Position des zu diskriminierenden Testwortes jeder Reihe, das den "kritischen" Laut enthält (Test I), bzw. des mit dem Mottowort zu identifizierenden Stimulus der Vergleichswortreihe (Test II) einen Effekt bewirkt hat, den wir als "Positionseffekt" bezeichnen. Bei der beträchtlichen Länge der Testwortreihen kann mit einem solchen Effekt gerechnet werden; die gestellte Diskriminations- bzw. Identifikationsaufgabe forderte den Probanden nicht nur Aufmerksamkeit für die Testdauer von ca. 20 Minuten ab (siehe hierzu unten, 3.6.0.), sondern hat möglicherweise im Falle der einzelnen Testwortreihen

ihre Gedächtniskapazität überfordert.

3.5.1. Ausgangshypothese

Wenn mit einem Positionseffekt der oben beschriebenen Art gerechnet werden kann, ist anzunehmen, daß er bei den beiden von uns durchgeführten Testarten in unterschiedlicher Weise in Erscheinung tritt. Bei der Diskriminationsaufgabe (vier Stimulusworte pro Reihe; eine einzige Distinktion) dürfte er sich vor allem dann zeigen, wenn das unterschiedliche Wort, das den "kritischen" Laut enthält, die erste Position einnimmt: Die früheren Gehörseindrücke laufen Gefahr, von den späteren "verwischt" bzw. überdeckt zu werden. - Für die Identifikationsaufgabe ("Mottowort", nach einer kurzen Pause gefolgt von drei Vergleichswörtern, die alle voneinander verschieden sind¹⁸ und von denen eines mit dem Mottowort identisch ist) läßt sich keine unmittelbar einleuchtende Voraussage bezüglich eines möglichen Positionseffektes machen. Man kann davon ausgehen, daß die Aufgabe um so leichter ist, je geringer die Entfernung zwischen Mottowort und identischem Vergleichswort ist; in diesem Falle wäre damit zu rechnen, daß sich ein Positionseffekt dort einstellt, wo das Vergleichswort die letzte Position (3) einnimmt. Mit gleicher Berechtigung kann argumentiert werden, daß das Mottowort, in Position 0 stehend, selbst am meisten gefährdet ist, von den späteren Gehörseindrücken verdeckt zu werden.

3.5.2. Statistischer Nachweis

Dem statistischen Nachweis liegt die im Anhang, S. 235 beigelegte Tabelle III zugrunde bzw. eine daraus angefertigte Hilfstabelle (relative Fehlerhäufigkeiten / Position in %). Für beide Tests wurde für beide Gruppen von Vpp jeweils getrennt die Frage gestellt, ob ein signifikanter Positionseffekt vorhanden ist. Die Rangvarianzanalyse nach FRIEDMAN (s.o., S. 61 u. Anm. 11) ergibt bei Test I (Hördiskrimination) signifikante Un-

¹⁸Siehe aber die oben, S. 39-41 gemachte Einschränkung, die Wörter des FVC-Tests betreffend.

terschiede ("LR": $\hat{\chi}_n^2 = 9.71 > \chi_{3,0.95}^2 = 7.81$; "M": $\hat{\chi}_n^2 = 19.25 > \chi_{3,0.95}^2 = 7.81$).

Diese Unterschiede wurden mithilfe von WILCOXON-WILCOX näher untersucht; für die Abfolge der Fehleranfälligkeit der einzelnen Positionen ergibt sich:

"LR"

4	<	3	<	2	<	1
---	---	---	---	---	---	---

-fehleranfälligt

"M"

3	<	4	<	2	<	1
---	---	---	---	---	---	---

Position 1 erweist sich damit für beide Gruppen als die fehleranfälligste: "Unterschiedliche" Stimuluswörter, die die erste Position in einer Testwortreihe einnahmen, wurden signifikant häufiger fehlerhaft beurteilt als solche, die die letzte Position innehatten. Bei der Gruppe "M" erweist sich Position 3 als die leichteste, so daß auch hier ein signifikanter Unterschied zu Position 1 nachgewiesen werden konnte.

Hingegen ergab die Rangvarianzanalyse bei Test II (Höridentifikation) bei beiden Gruppen keine unterschiedliche Fehlerhäufigkeit in Abhängigkeit von der Position des identischen Stimuluswortes in der Vergleichswortreihe: Ein Positionseffekt ist hier nicht nachzuweisen.¹⁹

3.5.3. Wertung

Der für Test I statistisch nachgewiesene Positionseffekt muß als ein unerwünschter experimenteller Effekt bezeichnet werden, der seine Ursache unmittelbar in der gewählten Darbietungsart hat. Er ließe sich auch kaum dadurch verringern, daß man die Pausen zwischen den einzelnen Stimuli der Testwortreihen verkürzt, die in unserem Test - 1.6 s betragen. - Daß in Hörtests dieser Art mit solchen Effekten gerechnet werden muß, unterstreicht die Notwendigkeit, für ausreichende Homogenität des Testmaterials Sorge zu tragen, vor allem für systema-

¹⁹Für die Kontrollgruppe läßt sich für keinen der beiden Tests ein Positionseffekt nachweisen.

tische Variation der Positionen, damit alle getesteten Probleme in gleichem Maße betroffen sind. Diesen Anforderungen genügt unser Test; eine weitere statistische Korrektur scheint nicht angebracht.

Der nichtsignifikante Befund für Test II bezüglich der Wirksamkeit des Positionseffektes scheint dieser Testart einen Vorzug gegenüber Test I einzuräumen. Das Ergebnis der statistischen Prüfung sollte jedoch mit Vorsicht interpretiert werden, da nach wie vor nicht ausgeschlossen werden kann, daß das Mottowort in der Position 0 von einem solchen Effekt tatsächlich betroffen war. Die vorliegenden Fehlerdaten erlauben keine Prüfung dieser Frage.

3.6.0. Optimale Testdauer

Eine Möglichkeit, wenigstens annäherungsweise die optimale Testdauer für Hörtests der hier vorgestellten Art zu ermitteln, besteht darin, die Fehlerhäufigkeit in verschiedenen Testteilen miteinander zu vergleichen. Wir haben zu diesem Zweck unsere beiden Tests willkürlich in jeweils drei Teile unterteilt und die ermittelten Fehlerhäufigkeiten pro Testdrittel in der Tabelle IV (siehe Anhang, S.236) zusammengetragen.

3.6.1. Statistischer Nachweis

Die statistische Prüfung wurde für beide Gruppen von Vpp getrennt durchgeführt. Für Test I (alle vier Blöcke zusammengekommen) ergibt die Rangvarianzanalyse nach FRIEDMAN bei der Gruppe "LR" keinen signifikanten Befund ($\hat{\chi}_n^2 = 1.19 < \chi_{2,0.95}^2 = 5.99$), d.h. die Fehlerzahlen in den drei Testabschnitten bleiben relativ konstant und unterscheiden sich nicht signifikant.

Hingegen erhält man für die Gruppe "M" signifikante Unterschiede für die Fehlerraten in den drei Abschnitten ($\hat{\chi}_n^2 = 8.20 > \chi_{2,0.95}^2 = 5.99$); nach WILCOXON-WILCOX ergibt sich für die Abhängigkeit der Fehlerzahlen von der Testdauer (~ 20 Minuten)

$\boxed{2.} < \boxed{3.} < 1.$ 20
 - fehleranfällig +

d.h. die Angehörigen dieser Gruppe machen in der mittleren Testphase signifikant weniger Fehler als in der Anfangs- und Schlußphase, wobei sich die Anfangsphase als die fehleranfälligste erweist.

Die statistische Prüfung für Test II, der sich für beide Gruppen als der schwierigere erwiesen hat (s.o., 3.2.0.), ergibt signifikante Unterschiede sowohl für die Gruppe "LR" ($\hat{\chi}_a^2 = 12.25 > \chi_{2,0.05}^2 = 5.99$) als auch für "M" ($\hat{\chi}_r^2 = 23.43 > 5.99$). Mithilfe der Multiplen Vergleiche (WILCOXON-WILCOX) ergeben sich für die Abhängigkeit der Fehlerzahlen von der Testdauer (~17.5 Minuten) die Abfolgen

"LR": $\boxed{1.} < \boxed{2.} < 3.$
 "M": $\boxed{1.} < \boxed{2.} < 3.$

Beide Gruppen machen somit im letzten Testdrittel signifikant mehr Fehler als im ersten; bei "M" ist darüber hinaus auch der Unterschied zwischen dem 2. und dem 1. Drittel signifikant.

In diesem Zusammenhang ist auch die Frage zu stellen, inwieweit das von uns praktizierte Verfahren, den Probanden bei jedem Einzeltest (Block) fünf Testwortreihen als "practice period" vorzugeben, die dann an das Bandende kopiert wurden und nur dort in die Fehlerwertung eingingen, gerechtfertigt werden kann.²¹ In Tab. V (Anhang, S. 237) sind die entsprechenden Fehlerzahlen zusammengestellt. Ein Vergleich der Fehlerzahlen im "Vorspann" mit denen der identischen Testwortreihen am Bandende ergibt bei Anwendung des Vorzeichentests von

²⁰Die Ziffern stehen für "1., 2. und 3. Testdrittel".

²¹Siehe oben, S. 34 und 37.

DIXON u. MOOD²² (bei getrennter Durchführung für beide Gruppen und beide Tests) keine signifikanten Leistungsunterschiede. Die Zunahme der absoluten Fehlerzahlen in der letzten Phase der Tests, wie sie vor allem bei Test II beobachtet wurde, sollte allerdings zu Bedenken Anlaß geben.

3.6.2. Wertung

Eine Abhängigkeit der Fehlerhäufigkeit von der Testzeit ist für Test I mit dem statistischen Verfahren, das wir angewendet haben, nicht nachweisbar; insbesondere ist für diese Testart keine signifikante Fehlerzunahme gegen Ende des etwa 20 Minuten währenden Tests festzustellen, woraus geschlossen werden darf, daß die Testdauer hier nahezu optimal gewählt ist.

Die für Test II ermittelten signifikanten Befunde (für beide Gruppen von Vpp) lassen hingegen den Schluß zu, daß die Dauer für diesen Test, der sich bereits als der schwierigere von beiden Testarten erwiesen hat, nicht optimal gewählt war. Dieser Test scheint die Probanden häufig bis an die Grenze ihres Konzentrations- und Aufnahmevermögens geführt zu haben. Künftige Untersuchungen ähnlicher Art mit Vpp unserer Altersgruppe (10 - 11 Jahre) sollten die Testdauer geringer bemessen und zwölf Minuten nicht überschreiten lassen.

Beobachtungen des Versuchsleiters während der Durchführung der einzelnen Tests haben wie auch die statistische Analyse der vorliegenden Daten gezeigt, daß umfangreiche "practice periods" allenfalls bei Eingangstests (am Beginn einer Serie gleichartiger Einzeltests bzw. bei Einführung eines neuen Testformats) erforderlich sind, wo sie den auch bei unserer Untersuchung beobachteten "Schwelleneffekt" (siehe oben, 3.4.0.) vermindern helfen; ansonsten kann auf sie verzichtet werden.

²²Siehe Sachs, Statistische Auswertungsmethoden, S.247-249.

3.7.0. Trennung der Gruppen

Wir beschließen dieses Kapitel, in dem wir uns ausführlich mit den beiden verwendeten Hörtests und den durch sie verursachten Einflüssen sowie mit dem unterschiedlichen Testverhalten beider Gruppen befaßt haben, mit der Frage, inwieweit die rohen Fehlerdaten eine Trennung beider Gruppen von Vpp nicht nur als Gesamtheiten, sondern auch auf individueller Ebene nahelegen.

Die Analyse, die von den Fehlerdaten der Tabelle II (siehe Anhang, S. 234) ausging, wurde mithilfe eines Programms zur Cluster-Analyse der "International Mathematical and Statistical Library" (IMSL) auf der Rechenanlage der Abteilung Medizinische Statistik und Dokumentation der Rheinisch-Westfälischen Technischen Hochschule Aachen durchgeführt. Die Ergebnisse von Test I und II wurden zusammen betrachtet.

Die insgesamt 28 Vpp beider Gruppen ("LR" und "M") werden nach dem Ergebnis der Cluster-Analyse den folgenden zwei Gruppen zugewiesen:

- Gruppe 1: alle Angehörigen der Gruppe "LR";
zusätzlich BT_2 , DM, FT, WF_1 von "M";²³
- Gruppe 2: die restlichen 16 Angehörigen von "M".

Das Ergebnis dieser Analyse liefert somit auch auf individueller Ebene eine nahezu saubere Trennung der beiden Gruppen von Vpp. Die Mittelwerte (Fehlerraten in Prozent) der beiden Cluster-Gruppen können der folgenden Tabelle entnommen werden (siehe Seite 71).

²³Vgl. auch oben, Tab. 5, S. 43.

		Gruppe 1	Gruppe 2
Test I	Block 1	23.44	12.57
	" 2	17.79	9.84
	" 3	19.05	8.00
	" 4	17.41	7.45
Test II	Block 1	21.63	12.96
	" 2	28.52	10.13
	" 3	29.85	10.64

Tab.14: Fehlerraten (in %) der beiden Cluster-Gruppen (Test I und II)

4.0.0. Beurteilung der Schülerproduktionen

Die Fehlerdaten für die Perzeptionsaufgabe lassen sich unmittelbar den von den Probanden ausgefüllten Antwortbögen entnehmen und in quantitative Daten überführen; die Art der gemachten Perzeptionsfehler ist dabei leicht zu bestimmen, insofern alle Fehler als kategorische interpretiert werden und die Kategorienklassen durch den Versuchsaufbau vorgegeben werden.

Produktionsfehler hingegen, wie sie in den hier durchgeführten drei Produktionstests ermittelt worden sind, sind selten von der kategorischen Art, und ihre Zuweisung zu Klassen von Kategorien gestaltet sich häufig problematisch. Dies trifft in besonderem Maße für "Lernerfehler" ("learner's errors") zu, wie sie beim Erwerb einer Fremdsprache auftreten: Die Literatur zur Fehleranalyse¹ enthält zahlreiche Hinweise auf den approximativen Charakter sprachlicher Äußerungen des Fremdsprachenlerner und unterstreicht die Notwendigkeit der Erforschung systematischer Tendenzen in diesem Bereich, die Aufschluß über spezifisches Lernverhalten geben und Detailkenntnisse der zugrundeliegenden Prozesse vermitteln soll.

Um die Aufdeckung solcher systematischer Fehler im Bereich der vier von uns untersuchten Ausspracheprobleme des Englischen für Deutsche geht es auch in dieser Untersuchung. Dies setzt eine qualitative Beurteilung der Schülerproduktionen voraus.

4.1.0. Auditive Beurteilung durch zwei geschulte Beobachter

4.1.1. Auditive Beurteilung und Transkription: Beobachter 1

Das Material aus den drei Produktionstests (rund 6000 Schüleräußerungen, die auf Tonträger gespeichert vorlagen) wurde zunächst vom Verf. über Kopfhörer mehrfach abgehört und in eine enge phonetische Transkription überführt.² Da das

¹Siehe S. 10, Anm. 30.

²Vgl. D. Abercrombie, "Phonetic Transcriptions", Le Maître Phonétique, 3^{me} série, No. 100 (1953), 32-34.

vom Verf. benutzte IPA-System³ wie jedes der bekannten alphabetischen Transkriptionssysteme⁴ bei "defektiven" und/oder etablierte lautsprachliche Normen verletzenden Äußerungen mitunter auf seine Grenzen stößt, obwohl ein beträchtlicher Vorrat an diakritischen Zeichen zur Verfügung steht,⁵ wurden diese engen Transkriptionen von Fall zu Fall durch zusätzliche Kommentare ergänzt. Dabei wurde Bezug sowohl auf artikulatorische Parameter (z.B. "stärker gerundet"; "zusätzliche velare Artikulation") als auch auf auditive Eindrücke (z.B. "stark aspiriert"; "klingt wie") genommen.

Im Falle der häufig schwer zu beurteilenden Realisationen von /ø,ø/, deren auditive Beurteilung bei Fehlen der visuellen Information über den Artikulationsort auch phonetisch geschulten Beobachtern schwerfällt,⁶ standen außerdem die während der Schüleraufnahmen angefertigten Protokolle zur Verfügung (s.o., S. 45), die den genannten Informationsmangel ausglich. Lückenlose Protokolle über diese Problemlaute waren allerdings nur für Test III (Nonsensewörter) geführt worden; bei Test IV und V (engl. Einzelwörter und Sätze) war die Protokollführung lückenhaft, da der Versuchsleiter durch Bedienung eines der beiden Tonbandgeräte in Anspruch genommen war. Für die Beurteilung der /ø,ø/- Realisationen dieses Tests muß demzufolge mit einer Fehlermarge gerechnet werden; sie kann für den zweiten Beobachter (s.u.), der sich zuvor als Angehöriger der Kontrollgruppe der Serie von Perzeptionstests unterzogen hatte,

³Siehe The Principles of the International Phonetic Association, repr. London 1971.

⁴Analphabetische Systeme phonetischer Transkription haben u.a. Jespersen und Pike entwickelt: Siehe O. Jespersen, Articulation of Speech Sounds, Represented by Means of Alphabetic Symbols, Marburg 1889 (vgl. auch dsb., Phonetische Grundfragen, Leipzig 1904, Kap. 2, S. 14-32 ("Lautschrift")); K.L. Pike, Phonetics, repr. Ann Arbor 1969, S. 153-156. Solche alphabetischen Transkriptionssysteme sind für die phonetische Praxis von geringem Wert, da sie zu mühsam zu handhaben sind.

⁵Siehe aber J.L.M. Trim, "Some Suggestions for the Phonetic Notation of Sounds in Defective Speech", Speech, 17 (1953), 21ff.

⁶Siehe unten, 5.1.3., bes. S. 100 und Anm. 11.

ziemlich genau angegeben werden und beträgt rund 6%.⁷

Für die Transkription der Vokalrealisationen wurde ein verkürztes Verfahren angewendet, das dennoch eine hinreichend genaue Spezifikation der Abweichungen gestattete. Normgerechte Realisationen im Bereich der palatalen Kurzvokale wurden der Einfachheit halber mit den phonologischen Symbolen /i, e, æ/ gekennzeichnet, deren Stellung in bezug auf das System der Kardinalvokale bekannt ist;⁸ die Notation von Abweichungen erfolgte sodann mit Zuhilfenahme der entsprechenden diakritischen Zeichen, die die bei der Vokalartikulation isolierten Parameter (Öffnungsgrad; artikulierender Zungenteil; Rundung) genau zu spezifizieren vermögen. Die notierten auditiven Qualitäten sind somit letztlich auf das Kardinalvokalsystem bezogen, welches als das einzige praktisch bewährte System der auditiven Beurteilung von Vokalen angesehen werden muß.⁹ Dauer wurde in jedem Falle phonetisch notiert, wobei vier Abstufungen der Dauer unterschieden wurden (kurz; halblang; lang; überlang).

Die angefertigten engen phonetischen Transkriptionen liefern ein hinlänglich genaues Bild von der Qualität der Schülerproduktionen und gestatten insbesondere eine exakte Interpretation der Abweichungen bezüglich aller involvierten phonetischen Parameter. Auf die Durchführung signalphonetischer Messungen, die in Ergänzung zu einer auditiven Beurteilung quantitative Daten über akustische Eigenschaften der produzierten Signale liefern, wurde verzichtet: Das hier vorliegende Material ist für solche Untersuchungen zu umfangreich und z.T. auch ungeeignet.¹⁰

⁷Siehe im einzelnen weiter unten, 5.1.8., S. 118-119. In den Hörtests wurden /ø,ɔ/ von diesem Beobachter in 98% der Fälle korrekt beurteilt.

⁸Siehe oben, S. 3, Anm. 9.

⁹Vgl. P. Ladefoged, "The Value of Phonetic Statements", Language, 36 (1960), 387-396.

¹⁰So sind beispielsweise Formantmessungen an von Kindern produzierten Vokalen mithilfe der üblichen Breitbandsonagramme

Gute Kenntnisse englischer lautsprachlicher Normen können beim Verf. vorausgesetzt werden (Anglistik-Studium; einjähriger USA-Aufenthalt; mehrere drei- bis sechswöchige USA- und Englandaufenthalte). Dennoch war die Hinzuziehung eines zweiten geschulten Beobachters, der darüber hinaus Englisch als Muttersprache spricht, angezeigt.

4.1.2. Auditive Beurteilung: Beobachter 2

Sämtliche Schüleräußerungen, die das Korpus zu den hier untersuchten Ausspracheproblemen des Englischen für Deutsche abgeben, wurden zusätzlich zu der auditiven Beurteilung durch den Verf. von einem phonetisch geschulten "native speaker" (Promotion im Hauptfach Phonetik) über Kopfhörer abgehört und in einem gegenüber der oben geschilderten Weise abgekürzten Verfahren beurteilt. Diesem Beobachter wurden für jede Vp Datenblätter zur Verfügung gestellt, auf denen die zu beurteilenden Stimuluswörter und -sätze in verschrifteter Form vorlagen. Im Fall von Test III (Nonsensewörter) handelte es sich dabei um breite (phonologische) Transkriptionen der Testwörter, bei Test IV und V um orthographische Versionen der (tatsächlichen) englischen Wörter und Sätze. Die zu beurteilenden Laute bzw. Lautverbindungen unserer vier Problembereiche waren in jedem Falle zweifelsfrei gekennzeichnet.

Die Anfertigung vollständiger (enger) phonetischer Transkriptionen ist ein zeitraubendes Verfahren, das den Fortgang der Untersuchung beträchtlich verzögert hätte. Diesem Beurteiler wurde deshalb nahegelegt, auf vollständige Transkription

problematisch, da die Filterbandbreite des Geräts Männerstimmen "favorisiert"; die Breitbandanalyse von Stimmlagen, deren Grundfrequenz von (annähernd) derselben Größenordnung ist wie die Bandbreite des Analysefilters, zeigt eine verwirrende Mischung von Formant- und Obertonstrukturen, die nur schwer zu interpretieren ist: Siehe G. Fant, "Analysis and Synthesis of Speech Processes", in B. Malmberg (Hrsg.), Manual of Phonetics, repr. Amsterdam 1974, S. 173-277, bes. 177-181.

der Schüleräußerungen zu verzichten, stattdessen nur die gekennzeichneten Problemlaute bzw. -lautverbindungen qualitativ zu beurteilen und diese Urteile ggf. zu kommentieren. Das Urteil konnte auch in diesem Falle in Form einer phonetischen Transkription des fraglichen Problemlautes abgegeben werden, sofern dafür gesorgt war, daß bei der Transkription einheitliche Kriterien zugrundegelegt wurden und beide Beurteiler sich bezüglich der Verwendung spezifischer Konventionen verständigten.¹¹

4.2.0. Auditive Beurteilung und phonetische Parameter

Auch diese Beurteilung durch einen zweiten geschulten Beobachter, der zudem "native speaker" ist, muß zuerst und vor allem anderen als eine phonetische Beurteilung aufgefaßt werden, da uns daran gelegen ist, auf dem Wege der auditiven (ohrenphonetischen) Analyse die Abweichungen der Schülerrealisationen von einer nicht allzu eng gefaßten Norm "des Englischen" zu bestimmen. Es war daher wichtig, diese Abweichungen im Hinblick auf die möglichenfalls beteiligten phonetischen Parameter selber zu bestimmen und für jedes der vier untersuchten Ausspracheprobleme diese Parameter explizit zu machen.¹²

Bei dieser Klärung kann die phonetische Fachliteratur zu Einzelfragen nützliche Aufschlüsse geben. Sie gibt jedoch in der Regel nur Aufschlüsse über "normales" Verhalten, welches frei ist von Störungen, wie sie etwa bei der Aneignung neuer Artikulationsgewohnheiten durch die Lerner einer Fremdsprache ge-

¹¹ Etwa bezüglich der Anwendung vereinfachter Verfahren wie des o.a. zur Notation der Vokalrealisationen.

¹² Unter phonetischen Parametern verstehen wir in erster Linie artikulatorische Parameter wie Artikulationsart, Artikulationsstelle, Art und Richtung des erzeugenden Luftstrommechanismus u. dgl. Eine vollständige Aufzählung findet sich bei G.E. Peterson u. J.E. Shoup, "A Physiological Theory of Phonetics", JSHR, 9 (1966), 5-67. Anders als diese Autoren, die die prosodischen Parameter von den phonetischen Parametern trennen und beide unter dem Oberbegriff "physiological speech parameters" vereinen, machen wir keine Unterscheidung zwischen phonetischen und prosodischen Parametern.

geben sind. Es ist deshalb damit zu rechnen, daß nicht alle an abweichenden und/oder normverletzenden lautlichen Realisationen unserer Probanden beteiligten Parameter von vornherein bekannt sind und in ihrer Bedeutung richtig eingeschätzt werden. Die auditive Analyse eines Datenkorpus wie des unsrigen kann hier erste Aufschlüsse erbringen; sodann ist zu wünschen, daß die ermittelten Befunde zu einem späteren Zeitpunkt durch gezielte experimentalphonetische Untersuchungen sowohl im artikulatorischen wie auch im akustischen Bereich ergänzt werden. Bevor wir uns jedoch der Analyse der von uns erhobenen Fehlerdaten zuwenden, werfen wir einen kurzen Blick auf die bei der Artikulation unserer Problemlaute beteiligten phonetischen Parameter.

4.2.1. Phonetische Parameter: /θ, ð/

Die klassischen Parameter der Konsonantenklassifikation sind Artikulationsort, Artikulationsart sowie die Art der Stimmbeteiligung (Zustand der Glottis). Die englischen interdentalen Frikative /θ/ und /ð/ werden von der Vorderzunge (aktiver Artikulator) an den oberen Schneidezähnen (passiver Artikulator) gebildet, wobei die Zungenspitze sich in (losem) Kontakt mit der Kante sowie der Innenseite der Schneidezähne befindet und die Seiten der Vorderzunge einen festeren Kontakt mit den Seitenzähnen bilden, so daß der erzeugende Luftstrom zwischen dem relativ flachen Vorderzungenteil und den Schneidezähnen zentral entweicht. Echte interdentaler Artikulationen sind bei englischen Sprechern selten; bezüglich des genauen Artikulationsortes ist jedoch ein hohes Maß an individueller Variation zu verzeichnen.¹³

Der artikulierende Luftstrom erzeugt ein breitbandiges Friktionsgeräusch von geringer Intensität, das den Frequenzbereich zwischen -1.4 und -8 kHz ausfüllt.¹⁴ Die spektralen Eigenschaften sowie die Intensität dieses Friktionsrauschens stellen diejenigen akusti-

¹³ Vgl. Gimson (1970), S. 183-184; siehe auch P. Stevens, "Spectra of Fricative Noise in Human Speech", Lang. & Speech, 3 (1960), 32-49. (34). - Mit Peterson u. Shoup (1966) wäre vielleicht besser von "linguadentalen" Frikativen zu sprechen.

¹⁴ Vgl. Stevens (1960), 41. Stevens macht auch darauf aufmerksam, daß die Konfiguration der Zähne die Qualität des produzierten Lautes in beträchtlichem Maße beeinflusst (34).

schen Kennzeichen ("cues") dar, die für die Perzeption der verschiedenen Frikative verantwortlich sind.¹⁵

Fehlerhafte ("defektive") Realisationen der englischen /θ, ð/ können sowohl den artikulatorischen Parameter "Artikulationsstelle" wie auch den Parameter "Artikulationsart" betreffen: Möglich sind sowohl (post-)dentale "gerillte" Frikative wie auch solche Laute, die an der richtigen Artikulationsstelle, aber ebenfalls mit "gerillter" anstatt flacher Zungenstellung gebildet sind. Diese Laute sind aufgrund ihrer Geräuscheigenschaften (höherfrequentes Rauschen; größere Intensität) den s-artigen Lauten zuzurechnen.- Zu kräftiger Druck der artikulierenden Vorderzunge an der dentalen Artikulationsstelle kann andererseits einen Mangel an Friktionsgeräusch hervorrufen und zu einer plosivartigen Lösung des derart gebildeten (inter- oder postdentalen) "Verschlußlautes" führen: Der resultierende Laut wird häufig als "echter" Verschlußlaut klassifiziert.

/θ/ und /ð/ werden gemeinhin nach der unterschiedlichen Art der Stimmbeteiligung unterschieden, was jedoch die tatsächlichen Verhältnisse unzureichend beschreibt. Manche Autoren ziehen es aus diesem Grunde vor, den Gegensatz als einen zwischen "starken" und "schwachen" Konsonanten zu beschreiben.¹⁶ Die genaue Natur des Gegensatzes bedarf weiterhin der Klärung; unsere Problematik wird jedoch nicht davon berührt.

4.2.2. Phonetische Parameter: /r/

Das wichtigste Allophon des englischen /r/ ist ein stimmhafter postalveolarer Dauerlaut, bei dessen Bildung die gehobene

¹⁵Vgl. G.W. Hughes u. M. Halle, "Spectral Properties of Fricative Consonants", *JASA*, 28 (1956), 303-310; J. Heinz u. K. Stevens, "On the Properties of Voiceless Fricative Consonants", *JASA*, 33 (1961), 589-596.- [θ] und [ð] werden allerdings in der Perzeption weniger durch ihre Geräuschteile als durch die unterschiedlichen Vokalübergänge unterschieden: siehe K.S. Harris, "Cues for the Discrimination of American English Fricatives in Spoken Syllables", *Lang. & Speech*, 1 (1958), 1-7.

¹⁶Z.B. O'Connor (1967), S. 37. Vgl. auch A. Malécot, "The Lenis-Fortis Opposition: Its Physiological Parameters", *JASA*, 47 (1970), 1588-1592.

Zungenspitze dem oberen Zahndamm angenähert ist, ohne diesen zu berühren. Die Seiten der Zunge bilden einen Kontakt mit den oberen Molaren, während die Zungenmasse insgesamt zurückgezogen und in ihrer Mitte gesenkt ist. Der artikulierende Luftstrom kann durch die nur approximierten aktiven und passiven Artikulatoren ungehindert und ohne Entstehen von Friktionsgeräusch entweichen.

Die Lippenposition wird durch den vokalischen Kontext bestimmt und kann sämtliche Parameterwerte (von gespreizt bis gerundet) annehmen.

Fehlbildungen können sowohl den Ortsparameter wie auch den Parameter "Rundung" involvieren: Bei zu starker Annäherung der Zungenspitze an die postalveolare Artikulationsstelle kann es zu Kontaktbildung kommen, die Friktionsgeräusch verursacht; es entstehen Laute, die in ihrem Gehörseindruck dem palato-alveolaren Frikativ [ʒ] ähnlich sind.¹⁷ Übertriebene Lippenrundung und Mangel an Vorderzungenartikulation resultiert in (friktionslosen Dauer-)Lauten des Typs [w] oder [v], wie sie als r-Ersatz häufig in der Kindersprache beobachtet worden sind.¹⁸ Da die deutschen r-Laute (wenigstens im norddeutschen Bereich) in ihrer überwiegenden Mehrzahl eine dorso-velare Artikulationsstelle haben,¹⁹ kann bei deutschen Englischlernenden mit Doppelartikulationen vom Typ [ɹw] gerechnet werden.

¹⁷Siehe auch Jones, *An Outline of English Phonetics*, S. 199; Gimson (1970), S. 211.

¹⁸Vgl. Gimson (1970), S. 208; siehe auch N.V. Smith, *The Acquisition of Phonology*, Cambridge (Engl.) 1973, Appendix C, S. 241-243.

¹⁹Zu ⁰r/ siehe H. Ulbrich, *Instrumentalphonetisch-auditive R-Untersuchungen im Deutschen* (=Schriften zur Phonetik, Sprachwissenschaft und Kommunikationsforschung, 13), Berlin 1972; W. Meyer-Eppler, "Zur Spektralstruktur der /r/-Allophone des Deutschen", *Acustica*, 9 (1959), 246-250. Siehe auch B.P.O. u. L.M. Hildebrandt, "Das deutsche R; Regelmäßigkeiten in der gegenwärtigen Reduktions-Entwicklung und Anwendung im Fremdsprachenunterricht", *Linguistics*, 11 (1965), 5-20.

4.2.3. Phonetische Parameter: /i, e, æ/

Vokale werden nach den Parametern "Öffnungsgrad", "artikulierender Zungenteil" sowie "Lippenrundung" klassifiziert; für spezifische einzelsprachliche Vokale sind darüber hinaus Dauereigenschaften relevant.

Abweichende Realisationen von /i, e, æ/ können anhand dieser vier Parameter eindeutig bestimmt werden. Da es sich bei allen drei Vokalen um palatale Kurzvokale handelt, die nur durch ihre Öffnungsgradeigenschaften unterschieden sind, ist vor allem für diesen zuletzt genannten Parameter mit Abweichungen zu rechnen. Beträchtliche Variation der allophonischen Dauer²⁰ läßt allerdings auch für den Dauerparameter abweichende Realisationen erwarten.

4.2.4. Phonetische Parameter: "Final voiced consonants"

Daß die im Englischen gegenüber dem Deutschen bewahrte sog. "Stimmhaftigkeitsopposition" im Auslaut²¹ nicht nur einfache paradigmatische Gegensätze in auslautenden Konsonanten betrifft, sondern durch die Länge des vorausgehenden Vokals (oder Diphthongs) mechanisch geregelt wird, ist seit Sweet bekannt und bereits vor nahezu 70 Jahren in einer beispielhaften experimentalphonetischen Untersuchung von E. A. Meyer nachgewiesen worden.²² Die Verhältnisse lassen sich in die folgenden

²⁰Siehe Gimson (1970), S. 94-95.

²¹Die Verhältnisse im Deutschen beschreibt G.M. Bonnin, "Some Acoustic Aspects of Final Stop Allophones in Contemporary German", *Phonetica*, 11 (1964), 65-100; siehe auch K. Ezawa, Die Opposition stimmhafter und stimmloser Verschlusslaute im Deutschen (=Tübinger Beiträge zur Linguistik 29), Tübingen 1972. Siehe aber auch H. Pilch, "Das Lautsystem der hochdeutschen Umgangssprache", *Zeitschr. f. Mundartforschung*, 33 (1966), 247-266.

²²E.A. Meyer, Englische Lautdauer; Eine experimentalphonetische Untersuchung (= Skrifter utgifna af K. Humanistiska Vetenskaps-Samfundet i Uppsala VIII.3), Uppsala u. Leipzig 1903; vgl. auch W. Ruth, Experimentalphonetische Untersuchung über die Dehnung kurzer Vokale im Standard English (=Sitzungsberichte der Akademie der Wissenschaften in Wien,

Regeln fassen:

- (1) Phonologisch lange Vokale sind von geringerer (phonetischer) Dauer, wenn sie von (einem oder mehreren) "fortis" Konsonanten gefolgt werden, als wenn sie in finaler Position vorkommen oder vor "lenis" Konsonant(en).
- (2) Phonologisch kurze Vokale sind nur vor "fortis" Konsonant(en) phonetisch kurz; werden sie von "lenis" Konsonant(en) gefolgt, sind sie von größerer phonetischer Dauer.

Der allophonisch bedingte Wechsel, der durch die Art der Folgekonsonanten automatisch geregelt ist, kann eine Umkehrung der phonetischen Dauerverhältnisse in den beiden beteiligten phonologischen Klassen bewirken: Messungen haben z.B. ergeben, daß phonologisch langes /i:/ in einem Wort wie seat typisch von kürzerer Dauer ist (~12.3 cs) als phonologisch kurzes /i/ in einem Wort wie hid (~14.7 cs).²³

Untersuchungen von Raphael mit synthetischem Material haben gezeigt, daß für die Wahrnehmung der "Stimmbeteiligungseigenschaften" wortfinaler Konsonant(verbindingen) die Länge des vorausgehenden Vokals nicht nur ein hinreichendes, sondern gar notwendiges Hörindiz²⁴ ist, während die tatsächliche Anwesenheit von Stimmhaftigkeit während der Verschlussphase nur geringen Indizwert besitzt.²⁵ Eine weitere Untersuchung²⁶

phil.-hist. Klasse, 222. Bd., 1. Abhdlg.), Wien u. Leipzig 1943.- Den Einfluß des finalen Konsonanten auf die Dauer des vorausgehenden Vokals hat als erster H. Sweet erkannt (Handbook of Phonetics, 1877, S. 59-60).

²³Vgl. die Beispiele bei Gimson (1970), S. 95.

²⁴Lüdtkes Vorschlag einer Übersetzung von "acoustic cue" (wohl in Anlehnung an frz. "index acoustique"): vgl. H. Lüdtke, "Zur Theoriebildung in der Phonetik", *Polia Linguistica*, 5 (1972), 333-354 (337).

²⁵Siehe L.J. Raphael, "Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English", *JASA*, 51 (1972), 1296-1303; vgl. auch P. Denes, "Effect of Duration on the Perception of Voicing", *JASA*, 27 (1955), 761-764.

²⁶Siehe L.J. Raphael et al., "Vowel and Nasal Duration as Cues to the Perceptual Categorization of Word-Final Consonants in English", Paper given at the Eighth International Congress on Acoustics, London 1974, Contributed Papers Vol. I, S. 18.

hat den Nachweis geführt, daß in englischen Wörtern, die in der Folge V+Nasal+C enden, die Dauer des Nasals ein wichtigeres Hörindiz für die Wahrnehmung der Stimmbeteiligungseigenschaften des finalen Konsonanten ist als die Vokaldauer. Es wird der Schluß gezogen, daß die Dauer der V + Nasalverbindung insgesamt die Wahrnehmung des Folgekonsonanten als "stimmhaft" (lenis) oder "stimmlos" (fortis) beeinflusst.²⁷

Durchgehende Stimmhaftigkeit und stimmhafte Verschlüßlösung sind im konsonantischen Auslaut englischer Wörter in der Regel nicht anzutreffen; auch die (phonologisch) "stimmhaften" Konsonanten des Englischen sind in finaler Position entweder "entstimmt" ("devoiced") oder völlig stimmlos, und finale Verschlüßlaute können darüber hinaus auch nicht nach dem Grad der Aspiration unterschieden werden, da sowohl stimmhafte wie stimmlose Plosive im Auslaut leicht behaucht sind.²⁸

Die Tatsache, daß die für die Wahrnehmung der Stimmbeteiligung auslautender Konsonanten primären akustischen und prosodischen Eigenschaften nicht im finalen Konsonanten selber, sondern im vorausgehenden Vokal (bzw. in -Nasal- od. -/l/-Verbindungen) enkodiert sind,²⁹ läßt für deutsche Englischlernende, die mit einem solchen Fall von "paralleler Enkodierung" der (akustischen und artikulatorischen) Information nicht vertraut sind, Lernschwierigkeiten besonderer Art erwarten.³⁰ Abweichende Realisationen können sowohl die phone-

²⁷Vgl. auch Jones, *An Outline of English Phonetics*, S. 233.- Wir vermuten, daß die Folge V + /l/ einen ähnlichen Einfluß auf die Wahrnehmung des finalen Konsonanten ausübt.

²⁸Siehe Gimson (1970), S. 153.

²⁹Zum Begriff der Enkodierung siehe A.M. Liberman, F.S. Cooper, D.P. Shankweiler u. M. Studdert-Kennedy, "Perception of the Speech Code", *Psychological Review*, 74 (1967), 431-461; A.M. Liberman, "Some Characteristics of Perception in the Speech Mode", *Perception and Its Disorders*, 48 (1970), 238-254.

³⁰Zu diesem Bereich liegt jetzt eine Untersuchung von W.J. Barry vor, die das Problem sowohl auf der Wahrnehmungs- wie auch auf der Produktionsseite untersucht hat: siehe W.J. Barry, "Parallel Encoding as a Source of Interference", Vortrag gehalten auf dem 4th Annual Phonetics Symposium,

tische Dauer des dem Endkonsonanten vorausgehenden Vokals (bzw. der V + Nasal- oder V + /l/-Verbindung) wie auch die Realisationsart des Endkonsonanten selber betreffen; mit Produktionsfehlern ist im Übrigen nicht nur für mit lenis-Konsonant(en) endende Lautfolgen zu rechnen, sondern auch für Folgen mit auslautendem fortis-Konsonant.³¹

4.3.0. Beurteilung nach Akzeptabilität

Die phonetisch-parametrische Beurteilung der unser Korpus ausmachenden Schülerproduktionen durch zwei geschulte Beobachter, die über genügende Erfahrungen in der auditiven Analyse phonetischer Daten sowie in der Handhabung eines Transkriptionssystems verfügen, ermöglicht zwar eine genaue Einschätzung der Normgerechtigkeit bzw. Fehlerhaftigkeit der Schüleräußerungen; sie gestattet darüber hinaus auch, die konstatierten Abweichungen dieser Realisationen von der zu erfüllenden Norm zu klassifizieren und eine Typologie der qualitativen Fehler aufzustellen, welche Aufschluß über individuelles Lernverhalten und die beim Erwerb einer fremden Aussprache verwendeten Lernerstrategien gibt. Hingegen ermöglicht sie von vornherein keine Überführung des so analysierten Materials in quantitative Daten, wie sie für eine statistische Fehleranalyse zu fordern sind. Aus diesem Grunde war es notwendig, beiden Beurteilern ein weiteres Kriterium an die Hand zu geben, welches eine quantitative Aufschlüsselung der Produktionsdaten erlaubte.

4.3.1. Beurteilungskriterien

Ergänzend zur phonetisch-parametrischen Beurteilung wurden sämtliche Schülerrealisationen von beiden Beobachtern nach Akzeptabilität beurteilt, wobei der erste Beobachter (Verf.) von einem einfachen Kriterium ausging und die Schüler-

The Language Centre, University of Essex, January 1974 (erscheint in den *Occasional Papers: Phonetics*).

³¹Die Hypothesen formuliert genau Barry (1974), 3-4.

produktionen entweder als "abweichend" oder "nichtabweichend" in bezug auf die zu fordernde Realisationsnorm einstuft. Dem zweiten Beobachter (der zugleich "native speaker" ist) stand ein abgestuftes Kriterium zur Verfügung, welches die Unterscheidung zwischen "guten", "(noch) akzeptablen" und "nicht akzeptablen" Äußerungen gestattete.³² Die Toleranzbreite dieses Kriteriums, das unserer Ansicht nach weder zu eng noch zu weit gefaßt ist, erlaubt eine gute Handhabung.

4.3.2. Normbezogenheit

Zu betonen ist, daß die Einstufung einer Schüleräußerung als "nichtabweichend" (Beurteiler 1) oder "gut" (Beurteiler 2) nicht gleichbedeutend ist mit der Feststellung ihrer vollständigen Übereinstimmung mit dem Modellwort oder -satz, das die Schüler nachahmten. Eine derart kritische Anwendung der Kriterien hätte im Falle der Beurteilung durch den Verfasser zur Zurückweisung von weit mehr als der Hälfte aller Äußerungen führen müssen und im Falle des zweiten Beobachters zu einer geringen Anzahl von Urteilen in der Kategorie "gut". Sie hätte darüber hinaus in jedem einzelnen Falle einen Vergleich der Schülerimitationen mit den Vorbildrealisationen erforderlich gemacht; letztere waren aber auf den zu beurteilenden Schülerbändern nur im Fall von Test V enthalten, wo die Sätze den Probanden über Außenlautsprecher zugespielt worden waren und insofern mit aufgenommen wurden.

Eine unkritische Anwendung des Akzeptabilitätskriteriums kann andererseits darin bestehen, dem geringen Ausbildungsstand der Probanden, die erst vier bis fünf Monate Englischunterricht hatten, von vornherein Rechnung zu tragen und bei der Einstufung ihrer Äußerungen mit unangebrachter Großzügigkeit

³²Das gleiche Kriterium hat Brière in seiner Untersuchung zur Aufstellung einer Hierarchie von Lernschwierigkeiten für die von ihm untersuchten Lautklassen verwendet; vgl. Brière, A Psycholinguistic Study of Phonological Interference, S. 54.

zu verfahren. Das hieße, anstelle einer hinreichend genau spezifizierbaren Aneignungsnorm des Englischen, die von Fremdheitsmerkmalen frei ist, eine ungenügend spezifizierte Schulnorm des Englischen für Deutsche zugrundelegen und dabei gewisse Abweichungen in der lautsprachlichen Realisation als gegeben hinzunehmen. Bei solcher Anwendung des Kriteriums sind Aussagen über das Auftreten von Fehlern und ihre Größenordnung nur beschränkt möglich. Beide Beurteiler sind überzeugt davon, nicht in der hier geschilderten Weise verfahren zu sein.

Im Übrigen sind beide Beobachter mit der Realisationsnorm des Informanten, der die nachzunehmenden Stimuluswörter und -sätze, die in den Tests Verwendung fanden, auf Band sprach, gut vertraut. Die Gebrauchsnorm dieses einen Sprechers jedoch als die allein verbindliche anzusehen, ist aus unmittelbar einsichtigen Gründen nicht angebracht.

4.4.0. Nichtübereinstimmung der Akzeptabilitätsurteile

Die Frage nach der Reliabilität der Akzeptabilitätsurteile sowie nach dem Grad der Übereinstimmung zwischen mehreren Beobachtern, denen die gleichen Äußerungen zur Beurteilung vorlagen, ist nur selten problematisiert worden und wird in den meisten Untersuchungen gar nicht erst gestellt.³³ Sie

³³So auch bei Nemser (1971), der die Beurteilung der Produktionsfehler seiner Probanden (Ungarn mit geringen Englischkenntnissen) offensichtlich allein vorgenommen hat, ohne diese Urteile von anderen geschulten Beobachtern bestätigen zu lassen. - Brière (1968), S. 54-55, berichtet von etwas mehr als 4% Nichtübereinstimmung zwischen seinen Beurteilern, die aber nur entgegengesetzte Urteile in den Kategorien "good" und "not acceptable" betrafen. Er schildert eine aufwendige Methode, diese nichtübereinstimmenden Urteile zu bereinigen. Auch hier verblieben schließlich noch 6 (von insgesamt 6720!) Äußerungen, über die keine Übereinkunft erzielt werden konnte; bei diesen sechs handelte es sich um Vokalrealisationen von [a] (2) und [y] (4).

wird hier in erster Linie für die voneinander abweichenden Akzeptabilitätsurteile gestellt, die jedoch, wie im folgenden klar wird, eine Abweichung in den Parameterurteilen beider Beobachter implizieren.

4.4.1. Kriterienvergleich und Bestimmung der Urteilsabweichungen

Da beiden Beobachtern verschiedene Akzeptabilitätskriterien zur Verfügung standen, ist zunächst ein Kriterienvergleich notwendig sowie eine Festlegung dessen, was als nicht-übereinstimmendes Urteil anzusehen ist.

Die möglichen Kombinationen der verschiedenen Kriterien ergeben sich aus Abbildung 4:

	Beob. 2			
Beob. 1		++	+	-
+				X
-	X			

Abb. 4: Kriterienvergleich: Beob. 1 u. 2³⁴

Von den 2x3 = 6 Kombinationsmöglichkeiten, die sich aus dem Kriteriumvergleich ergeben, wurden nur zwei, die in Abb. 4 mit "x" gekennzeichnet sind, als nichtübereinstimmende Urteile aufgefaßt - nämlich alle jene Fälle, in denen (a) der erste Beobachter eine Lautrealisation als "nichtabweichend" eingestuft hatte, die vom zweiten Beobachter als "nicht akzeptabel" bezeichnet wurde, und (b) der erste Beobachter "abweichend" beurteilt hatte, während der zweite Beobachter die Realisation als "gut" bezeichnete.³⁵ Es verbleibt eine "Grauzone" möglicher Nichtübereinstimmung in allen Fällen, in denen der mit dem feiner abgestuften Kriterium ausgestattete zweite Beobachter

³⁴Beobachter 1: "+" = "nichtabweichend", "-" = "abweichend"; Beobachter 2: "++" = "gut", "+" = "(noch) akzeptabel", "-" = "nicht akzeptabel".

³⁵Ähnlich Brière (1968), S. 54-55.

eine Äußerung als "(noch) akzeptabel" eingestuft hatte, die von dem ersten Beobachter entweder als "abweichend" oder "nicht abweichend" beurteilt worden war; sie wird hier im Vertrauen auf die bessere Kenntnis englischer lautsprachlicher Realisationsnormen seitens des "native speaker" vernachlässigt, dessen Urteile in dieser Kategorie unbesehen hingenommen wurden. - Als Nichtübereinstimmung besonderer Art wurden schließlich jene nur (vereinzelt auftretenden) Fälle von Abweichungen angesehen, bei denen beide Beobachter einen anderen phonetischen Parameter beteiligt sahen.

4.4.2. Nichtübereinstimmung in den unterschiedlichen Aussprachebereichen

Tab. 15 gibt einen Überblick über die in allen drei Produktionstests nach Anwendung des oben genannten Kriterienvergleichs (bzw. nach dem Vergleich der phonetisch-parametrischen Fehlerspezifikation) nichtübereinstimmend beurteilten Schülerproduktionen; die Zahlen (relative Häufigkeiten in Prozent) sind für die vier Ausspracheprobleme und die drei Tests sowie für beide Gruppen von Vpp getrennt aufgeführt. Die Summierung ergibt (nach Wichtung der entsprechenden Durchschnittszahlen),

Gruppe	Test	TH	R	V	FVC	\bar{x} ³⁶
"LR"	III	0.11	0.03	0.15	0.13	0.10
	IV	0.07	0.04	0.21	0.19	0.14
	V	0.06	0.04	0.10	0.13	0.09
	Σ_{LR} ³⁶	0.08	0.03	0.14	0.14	0.10
"M"	III	0.04	0.03	0.16	0.12	0.08
	IV	0.03	0.04	0.18	0.11	0.10
	V	0.02	0.02	0.09	0.11	0.07
	Σ_M ³⁶	0.03	0.03	0.13	0.11	0.08
	Σ_{LR+M}	0.05	0.03	0.13	0.12	0.09

Tab. 15: Nichtübereinstimmung der Akzeptabilitätsurteile (in %)

³⁶Bei den hier angegebenen Durchschnittszahlen handelt es sich um gewichtete Durchschnitte.

daß von den insgesamt 6310 Schüleräußerungen 9 % (das sind 564) als nichtübereinstimmend im oben festgelegten Sinne beurteilt wurden.³⁷ Weniger vermerkwürdig als diese Zahl ist jedoch, daß sich in der Beurteilung der von uns getesteten Ausspracheprobleme des Englischen für Deutsche deutliche Unterschiede abzeichnen: So sind, wie aus Tabelle 15 unmittelbar hervorgeht, die Schülerrealisationen der englischen palatalen Kurzvokale /i, e, ə/ sowie die Realisationen der sog. "final voiced consonants" häufiger nichtübereinstimmend beurteilt worden als die r-Realisationen und die Produktionen der interdentalen Frikative. Der (nichtparametrische) statistische Test unter Einbeziehung aller zwölf Vpp beider Gruppen ergibt einen signifikanten Befund ($\hat{\chi}_a^2 = 22.10 > \chi_{3,0.05}^2 = 7.81$): Vokalrealisationen sowie "final voiced consonants" wurden signifikant häufiger nichtübereinstimmend beurteilt als r- und TH-Realisationen. Der Effekt läßt sich im übrigen auch für jeden einzelnen der drei Produktionstests nachweisen; die Prüfung bei Test V (englische Sätze) ergibt jedoch signifikante Unterschiede in der Nichtübereinstimmung der Beobachter nur für die "final voiced consonants" gegenüber den r-Realisationen, während sich für die Vokalrealisationen eine merkliche Verbesserung der Urteilsübereinstimmungen gegenüber Test III und IV beobachten läßt, in denen isolierte Nonsense- bzw. englische Einzelwörter Verwendung fanden. Die relative Natur der Vokalqualität und ihre Abhängigkeit von kontextuellen Einflüssen ist bekannt;³⁸ im Falle unseres Tests läßt sich nachweisen, daß die Übereinstimmung in der Beurteilung der Vokalrealisationen unserer Probanden dort merklich verbessert ist, wo die Einbettung der Testwörter in größere zusammenhängende Kontexte erfolgt und einen verlässlicheren Bezugsrahmen für die Beurteilung abgibt.

Daß der Grad der Übereinstimmung in den Akzeptabilitätsurteilen der beiden Beobachter mit dem Ausmaß der in den ver-

³⁷Vgl. auch Brière (1968), S. 54-55.

³⁸P. Ladefoged, "The Nature of Vowel Quality", in *dsb.*, *Three Areas of Experimental Phonetics*, S. 50-142.

schiedenen Ausspracheprobembereichen anzutreffenden Produktionschwierigkeiten der Vpp hoch korreliert,³⁹ erscheint vermerkwürdig; eine Erklärung für diese Beobachtung kann im Rahmen dieser Untersuchung nicht gegeben werden.

4.4.3. Nichtübereinstimmung bei der Beurteilung individueller Leistungen

Der statistisch signifikante Befund bezüglich der Nichtübereinstimmung der Beurteiler in den verschiedenen Ausspracheprobembereichen wurde zum Anlaß genommen, der Frage nachzugehen, ob einzelne Individuen unserer beiden Gruppen (signifikant) häufiger nichtübereinstimmend beurteilt worden sind als andere. Die statistische Prüfung erweist die Nichtwirksamkeit eines solchen Effektes ($\hat{\chi}_r^2 = 10.39 < \chi_{11,0.05}^2 = 19.7$): Alle zwölf Probanden, die sich den Produktionstests unterzogen, geben somit in gleichem Ausmaß Anlaß zu nichtübereinstimmenden Urteilen.⁴⁰

4.4.4. Entscheidung bezüglich der Nichtübereinstimmungen: Beobachter 3

Anstatt die nichtübereinstimmend beurteilten Äußerungen aus den drei Produktionstests bei gemeinsamem Abhören durch beide Beobachter erneut beurteilen zu lassen und so zu einer Entscheidung über die rund 9 % des gesamten Materials ausmachenden Zweifelsfälle zu kommen, entschloß sich der Versuchsleiter, diese strittigen Fälle durch Hinzuziehung eines dritten Beobachters, dessen phonetische Ausbildung der der beiden anderen vergleichbar und der darüber hinaus ebenfalls "native speaker" ist, entscheiden zu lassen. Bei diesem Beobachter

³⁹Wir greifen damit auf die Ergebnisse der Datenanalyse vor: siehe im einzelnen weiter unten, Kap. 5.

⁴⁰Um unerwünschte Beurteilereffekte von vornherein auszuschalten, wurde die Reihenfolge der vom zweiten Beobachter zu beurteilenden Vpp randomisiert, so daß er keine Kenntnis von der Gruppenzugehörigkeit der einzelnen Probanden hatte.

handelte es sich um den Sprecher aller in den Perzeptions- und Produktionstests verwendeten Aufnahmen.⁴¹

Die strittigen Schüleräußerungen wurden von den Original-Schülerbändern kopiert und zu drei Testbändern (je eines für Test III, IV und V) zusammengestellt. Die Äußerungen wurden in der durch den jeweiligen Test gegebenen Reihenfolge kopiert; eine weitere Randomisierung schien nicht notwendig. Dem Beurteiler lagen vorbereitete Datenblätter der gleichen Art vor, wie sie von den anderen Beurteilern benutzt worden waren und auf denen der zu beurteilende Laut zweifelsfrei gekennzeichnet war. Es wurde ihm keine Mitteilung darüber gemacht, welcher Art die von den beiden anderen Beobachtern abgegebenen Urteile waren.

Auch von diesem Beobachter wurde gefordert, bei Feststellung von Abweichungen bzw. Normverletzungen die beteiligten phonetischen Parameter möglichst genau zu spezifizieren. Darüber hinaus wurde er aufgefordert, die Akzeptabilität der auditiv analysierten Laute bzw. Lautfolgen nach dem abgestuften Kriterium, das auch dem zweiten Beobachter zur Verfügung gestanden hatte, zu beurteilen.

Eine Analyse der von diesem dritten Beobachter abgegebenen Urteile, die eine Entscheidung über die strittigen Schüleräußerungen herbeiführen sollte, zeigt, daß sich die Akzeptabilitätsurteile dieses Beobachters zu fast gleichen Teilen auf die drei zur Verfügung stehenden Kategorien verteilen und nicht etwa deutlich erkennbar einen der beiden ersten Beobachter in seinen Urteilen bestätigen.

⁴¹Die Verwendung dieses Beobachters bei der Beurteilung der Zweifelsfälle war aus bestimmten Gründen, auf die erst später eingegangen werden kann, angezeigt: siehe weiter unten, 5.3.5., bes. S. 175-176

4.5.0. Phonetische Fehleranalyse

Die Feststellung der Art der in den Schüleräußerungen gemachten Produktionsfehler und ihre phonetisch-parametrische Spezifikation ist das erste Ziel einer Fehleranalyse, die darauf aus ist, eine Typologie der in den vier untersuchten Aussprachebereichen beobachteten Fehler aufzustellen. So aufschlußreich individuelle Fehlrealisationen auch sein mögen, die einen Einblick in die Dynamik individueller Aneignung neuer Aussprachegewohnheiten gestatten, wird es uns dennoch in erster Linie darum zu tun sein, die systematischen Lernerfehler aufzudecken, die ein Symptom für Lernschwierigkeiten sind, wie sie auch bei größeren Populationen als der von uns untersuchten zwölfköpfigen Gruppe auftreten.

4.6.0. Quantifizierung der Produktionsdaten

Auf den nichtkategorischen Charakter vieler Produktionsfehler, die beim Erwerb neuer Aussprachegewohnheiten auftreten, ist weiter oben bereits hingewiesen worden; kategorische "Falsch/Richtig"-Entscheidungen werden dem spezifischen Charakter vieler Lernerfehler im Bereich der Aussprache, welche erkennen lassen, daß die geforderte Zielartikulation wenigstens approximiert wird, nicht gerecht. Es liegt insoweit nahe, sich bei der Einschätzung des Grades der Fehlerhaftigkeit eines abgestuften Urteilkriteriums zu bedienen, wie es in den Akzeptabilitätsurteilen des zweiten (und den ergänzenden Urteilen des dritten) Beobachters vorliegt.

Um eine Quantifizierung der beobachteten Aussprachefehler und die Aufstellung von Fehlerhierarchien zu ermöglichen, wurden den drei möglichen Akzeptabilitätsurteilen des zweiten Beobachters willkürlich Indexwerte zugeordnet, wobei Urteile der Kategorie "gut" den Indexwert "1", solche der Kategorie "(noch)akzeptabel" den Wert "0.5" und die der Kategorie

"nicht akzeptabel" den Wert "0" erhielten.⁴² Dieses Verfahren gestattet, auch die als "noch akzeptabel" eingestuften Realisationen bei der Ermittlung der Produktionsleistungen zu berücksichtigen und damit die Dynamik des Lernprozesses besser abzubilden.

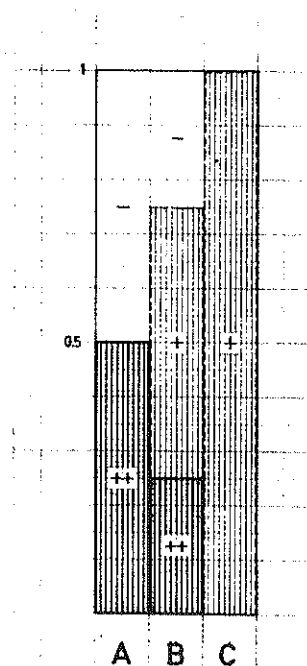


Abb. 5

Dem numerischen Indexwert zwischen 0 und 1, mit dem die Güte der von den Probanden erbrachten Produktionsleistung erfaßt wird, haftet insofern eine gewisse Schwäche an, als er keine eindeutigen Rückschlüsse auf die ihm zugrundeliegenden Urteile erlaubt: Abb. 5 macht deutlich, daß z.B. einem Indexwert 0.5 50 Urteile der Kategorie "gut" und 50 der Kategorie "nicht akzeptabel" zugrundeliegen können (Fall A), aber auch 25 Urteile der Kategorie "gut", 50 Urteile "(noch) akzeptabel" und 25 "nicht akzeptabel" (Fall B),⁴³ gegebenenfalls aber auch alle 100 Urteile in der Kategorie "(noch) akzeptabel" gefällt wurden (Fall C).

Es ist deshalb unumgänglich, die Urteilkategorien, die der Errechnung des Produktionsindex zugrundeliegen, in jedem Falle explizit zu machen und

⁴² Ähnlich Brière (1968), S. 55

⁴³ Der "Produktionsindex" errechnet sich in diesem Falle als $\frac{25 \times 1 + 50 \times 0.5}{100} = \frac{25 + 25}{100} = 0.5$; der Divisor "100" wird eingeführt, um diese Zahlen den in den Perzeptionstests ermittelten Fehlerhäufigkeiten (in %) vergleichbar zu machen bzw. den daraus (durch Subtraktion der Fehlerraten von 100) ermittelten Perzeptionsindexwerten.

die Vergleichbarkeit verschiedener Indexwerte unter Beweis zu stellen. Dies sollen Histogramme der in Abb. 5 für einen theoretischen Fall wiedergegebenen Art leisten, die wir weiter unten für die beurteilten Ausspracheleistungen in jedem der vier Problembereiche erstellen.

Im folgenden wenden wir uns der Analyse der in fünf Einzeltests erhobenen Fehlerdaten zu und stellen die ermittelten Befunde zu den Hör- und Produktionsschwierigkeiten zweier Gruppen deutscher Englischlernender für die von uns untersuchten Ausspracheprobleme vor.

5.0.0. Lautwahrnehmung und Lautproduktion:
Testergebnisse

5.1.0. TH-Test

Eine spezifische Eigenart des Testmaterials macht es notwendig, die Versuchsergebnisse für die stimmlosen und stimmhaften interdentalen Frikative getrennt zu referieren und die beiden Laute /θ/ und /ð/ überhaupt getrennt voneinander zu betrachten. Anlaß hierfür ist die Tatsache, daß /ð/ in initialer (prä vokalischer) Position nur in einer eng begrenzten Anzahl von Wörtern vorkommt, die größtenteils zur geschlossenen Klasse der sog. "Formwörter" gehören¹ und im Englischen besonders häufig vorkommen. Dies wie auch die Tatsache, daß diese Wörter nur im Falle von emphatischer Betonung akzenttragend sein können, gewöhnlich jedoch unakzentuiert sind und spezifischen Abschwächungserscheinungen unterliegen, läßt für die Produktion dieser Laute das Wirken besonderer "Performanzfaktoren" erwarten, die die korrekte Produktion in besonderem Maße erschweren.² Die weitere Erörterung dieser Faktoren wird bei der Interpretation der Produktionsfehler vorzunehmen sein.

5.1.1. Perzeption: /θ/

Das englische stimmlose /θ/ kommt in einer Anzahl verschiedener Positionen im Wort vor,³ von denen die wichtigsten getestet wurden. Eine Übersicht über die einzelnen phonetischen Kontexte, in denen dieser Problemlaut in den Testwörtern

¹Aufgezählt bei Scherer u. Wollmann, Englische Phonetik und Phonologie, S. 126-127; siehe auch Jones, An Outline of English Phonetics, S. 184.

²Eine besondere Erschwernis ist vor allem dadurch gegeben, daß initiales /ð/ häufig nach dem auslautenden -s der Flexionsendung der 3. sg. present tense steht und besondere Gefahr läuft, in dieser Position in zusammenhängender Rede assimiliert zu werden. Vor einer solchen progressiven Assimilation des Artikulationsortes warnen die einschlägigen Handbücher ausdrücklich: siehe etwa Gimson (1970), S. 184-185.

³Eine ausführliche Übersicht über die Phonetik des (britischen) Englisch gibt Gimson (1970), S. 239-255.

vorkam, ist in Tabelle 6 (s.o., S. 52) enthalten. Aus der Aufstellung geht hervor, daß "marginale" phonetische Kontexte für diesen Laut entsprechend selten zur Beurteilung angeboten wurden. Die Häufigkeit der Testkontexte genau nach den für das Englische ermittelten Phonemfrequenzen für alle möglichen Kontexte zu errechnen,⁴ schien uns eine zu weitgehende Forderung angesichts der Vielzahl der untersuchten Problembereiche; das Testmaterial hätte um ein Vielfaches vermehrt werden müssen, um solchen Forderungen zu genügen.

5.1.2. Fehlervorgabe

Die Konstruktion von Hörtests der hier durchgeführten Art setzt theoretische Überlegungen bezüglich der Art der erwarteten Perzeptionsfehler voraus, da diese Fehler in den einzelnen Testwortreihen vorgegeben werden; man beobachtet sodann, wie gut die perzeptorische Trennung zwischen dem kritischen Testlaut und den vorgegebenen Konfusionslauten ist. Es scheint plausibel, hierbei von tatsächlich beobachteten Fehlern bei der Produktion dieser kritischen Laute auszugehen, wobei das Bestehen eines Zusammenhangs zwischen Produktion und Perzeption vorausgesetzt wird. Zu fragen ist, ob nur von solchen (tatsächlich beobachteten) Produktionsfehlern auszugehen ist, wie sie beim Erwerb neuer Aussprachegewohnheiten etwa im Fremdsprachenunterricht auftreten, oder ob man darüber hinaus auch solche Fehler mit berücksichtigen soll, wie sie beim Erstspracherwerb zu verzeichnen sind. Schließt man die letztgenannten Fehler mit ein, setzt man voraus, daß der Prozeß des Erlernens einer zweiten Sprache nicht grundsätzlich verschieden ist von dem des Erlernens der ersten und der Erwerb neuer Aussprachegewohnheiten sich auf ähnliche Weise vollzieht - eine Auffassung, die bis heute ebenso wenig bewiesen ist wie ihr Gegenteil.⁵

⁴So ist Nemser in seiner Untersuchung vorgegangen: vgl. Nemser, An Experimental Study of Phonological Interference, S. 55.

⁵Vgl. Corder, "The Significance of Learner's Errors", passim. Siehe auch H. Wode, "Natürliche Zweitsprachigkeit:

Bei der Vorgabe der Verwechslungsmöglichkeiten gehen wir von diesem doppelten Ansatz aus, der uns sinnvoller zu sein scheint als das Ausgehen von abstrakt-typologischen Vergleichen auf der Phonemebene und den im Anschluß daran formulierten Fehlerhypothesen. Wir glauben, daß die im Erst- wie auch im Zweitsprachenerwerb tatsächlich beobachteten Produktionsfehler eine sinnvolle Eingabe in Perzeptionsaufgaben der hier gestellten Art darstellen.

Als möglicher Fehlerreflex gegenüber dem "kritischen" Testlaut /θ/ kommt selbstverständlich in erster Linie das von deutschen Englischlernenden häufig substituierte [s] in Frage, wobei im Augenblick davon abgesehen werden kann, daß die Substitution unmodifizierter s-Artikulationen in der lautsprachlichen Realisation von /θ/ durch Deutsche vermutlich selten ist und die korrekte Artikulationsart (Zungenkonfiguration) bzw. der richtige Artikulationsort häufig wenigstens approximiert werden. Die didaktischen Handbücher vermerken darüber hinaus zum Teil auch Substitution von [t] für /θ/, was häufig auf Einflüsse der Schrift zurückgeführt wird.

Nirgendwo finden sich Hinweise darauf, daß deutsche Englischlernende für /θ/ den auditiv sehr ähnlichen labiodentalen Frikativ [f] substituieren, der im Erstsprachenerwerb (englischsprachiger Kinder) als häufiger produktiver Reflex für /θ/ auch dann noch zu vermerken ist, wenn /θ/ bereits als erworben gelten kann.⁶ Auch [s] und s-artige Artikulationen sind in der Literatur zum Erstsprachenerwerb als Fehlerreflexe belegt, Substitution von [t] für /θ/ hingegen nur

Probleme, Aufgaben, Perspektiven", *Ling. Berichte*, 32 (1974) 15-36, wo die einschlägige Literatur kritisch gesichtet wird und eine Reihe von Arbeitshypothesen aufgestellt werden.

⁶Siehe Ch. A. Ferguson, "Fricatives in Child Language Acquisition", *Papers and Reports on Child Language Development*, Stanford University, 1973, 63-85 (dort auch weitere Literatur). Siehe auch D.L. Olmsted, *Out of the Mouth of Babes* (= *Janua Linguarum*, Ser. minor, 117), Den Haag 1971, passim.

vereinzelt.⁷ - [f], [s] und [t] sind demnach diejenigen Konfusionslaute, die in beiden Perzeptionstests dem "kritischen" Laut gegenübergestellt wurden.

5.1.3. Hördiskrimination

Die Ergebnisse von Test I zur Hördiskrimination von /θ/ sind in den Tabellen 16a (Gruppe "LR") und 16b (Gruppe "M") zusammengestellt; Abb. 6 (s. S. 99) stellt sie in

Kontext	TW ⁸	FZ	FR	Fehlerreflexe		
				/f/	/s/	/t/
*1 # _ v	144	33	0.23	0.79	0.18	0.03
2 # _ r	32	12	0.38	0.67	-	0.33
3 C ¹ _ s#	16	13	0.81	-	-	1.00
4 V _ s#	32	6	0.19	0.67	-	0.33
5 V _ t#	32	9	0.28	1.00	0.00	-
*6 V _ #	72	19	0.26	0.63	0.21	0.16
*7 C ² _ #	96	24	0.25	0.79	0.08	0.13
Σ	424	116	0.27	0.67	0.10	0.23
gemeins. Kontexte ⁹	312	76	0.24	0.74	0.16	0.10
gewichtete Durchschn. ¹⁰	-	-	-	0.75	0.16	0.09

Tab. 16a: Fehlerraten und Fehlerreflexe: Hördiskrimination (/θ/); Gruppe "LR" (8 Vpp)

⁷Ferguson (1973) verzeichnet keine einzige solche Substitution; siehe aber N.V. Smith, *The Acquisition of Phonology*, Cambridge (Engl.) 1973, Appendix C, s.v. thank-you, thick, thigh, thin etc. (S. 254-255).

⁸Verwendete Abkürzungen: TW = Anzahl der Testwörter (= abgegebene Hörerurteile); FZ = (absolute) Fehlerzahlen; FR = Fehlerraten (= relative Fehlerhäufigkeiten).

⁹Als "gemeinsame Kontexte" bezeichnen wir diejenigen phonetischen Umgebungen für den kritischen Laut, in denen alle drei Fehlerreflexe - hier [f],[s],[t] - phonotaktisch möglich sind und in Testwortreihen tatsächlich als Verwechslungsmöglichkeiten vorgegeben wurden (diese Kontexte sind mit einem * gekennzeichnet).

¹⁰Die Wichtung der relativen Häufigkeiten wurde nach Nenser vorgenommen; die gewichteten Durchschnittszahlen

Kontext	TW	FZ	FR	Fehlerreflexe		
				/f/	/s/	/t/
*1	360	53	0.15	0.85	0.13	0.02
2	80	29	0.36	0.55	-	0.45
3	40	22	0.55	-	-	1.00
4	80	8	0.10	0.62	-	0.38
5	80	24	0.30	0.96	0.04	-
*6	180	33	0.18	0.76	0.06	0.18
*7	240	37	0.15	0.86	0.03	0.11
Σ	1060	206	0.19	0.71	0.05	0.24
gemeins. Kontexte	780	123	0.16	0.82	0.07	0.11
gewichtete Durchschn.	-	-	-	0.83	0.08	0.09

Tab. 16b: Fehlerraten und Fehlerreflexe:
Hördiskrimination (/θ/); Gruppe "M" (20 Vpp)

anschaulicher Form dar und erlaubt einen unmittelbaren Vergleich der beiden Gruppen von Vpp.

Bemerkenswert erscheint zunächst das unerwartet gute Diskriminationsvermögen beider Gruppen bezüglich der hier getesteten schwierigen Distinktionen: /θ/ wird in 73 (Gruppe "LR") bzw. 81% ("M") der Fälle erfolgreich diskriminiert, was insbesondere angesichts der beträchtlichen auditiven Ähnlichkeit von [θ] und [f] erstaunt. Die sechs Angehörigen der aus kompetenten Beobachtern und "native speakers" bestehenden Kontrollgruppe erzielen in diesem Test zwar deutlich bessere Leistungen; eine durchschnittliche Fehlerhäufigkeit von 5% erweist jedoch die relative Schwierigkeit der Diskriminationsaufgabe auch für diesen Kreis von Hörern, die mit den getesteten Distinktionen wohlvertraut sind und zudem weitaus bessere Abhörbedingungen

errechnen sich aus dem Produkt der in einem Kontext beobachteten relativen Fehlerhäufigkeit und dem Quotienten aus der absoluten Fehlerzahl (in diesem Kontext) und der durchschnittlichen Fehlerzahl in gemeinsamen Kontexten. Vgl. Nemser (1971), S. 167, Anm. 131.

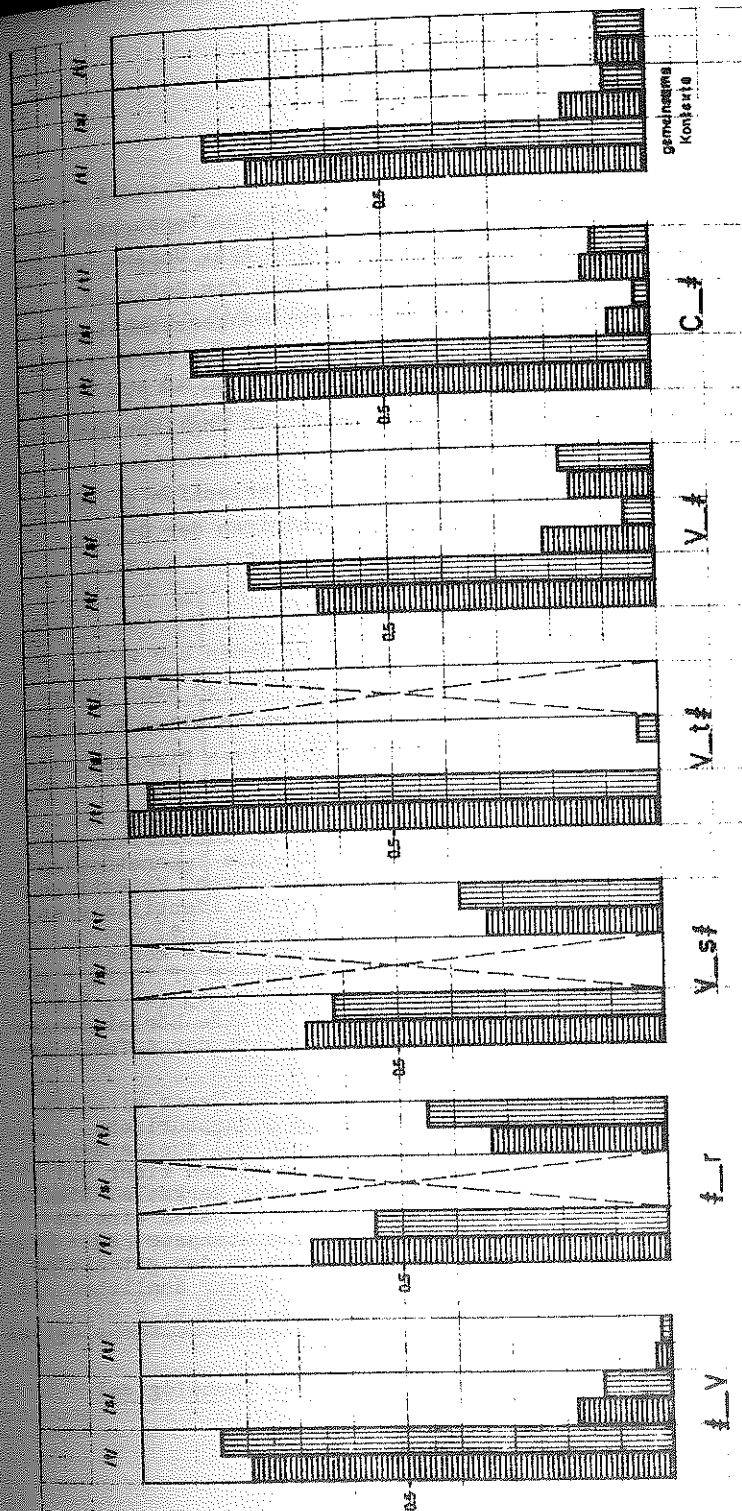


Abb. 6: Hördiskrimination: /θ/;
Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten
("LR": ; "M":)

vorfanden.¹¹ - Ein Blick auf Art und Häufigkeit der auditiven Verwechslungen zeigt, daß bei beiden Gruppen von Versuchspersonen [f] der überwältigende Fehlerreflex ist. Unsere Ergebnisse stimmen in dieser Hinsicht gut mit den Befunden früherer Untersuchungen allgemeiner Perzeptionsgesetzmäßigkeiten¹² wie auch mit den Ergebnissen Nemser's zur Perzeption von /θ/ bei ungarischen Englischlernenden überein.¹³ [s] und [t], den häufigen Lautsubstitutionen deutscher Englischlernender bei der Realisation des stimmlosen interdentalen Frikativs, kommt als Fehlerreflexen in der Perzeption gegenüber [f] kaum Bedeutung zu, wobei bei den Angehörigen der Gruppe "LR" (Hauptschüler(innen)) [s] allerdings häufiger an Verwechslungen beteiligt war als [t].

Für die einzelnen phonetischen Kontexte, in denen der kritische Laut zur Beurteilung angeboten wurde, ergeben sich zwar unterschiedliche Verhältnisse; jedoch erweist sich in allen Positionen [f] als der primäre Fehlerreflex in der auditiven Diskrimination.

¹¹Siehe oben, 2.5.1., S. 23-24.- In einer Identifikationsaufgabe, in der englischsprachige Hörer, die als "phonetically trained" bezeichnet werden, neben sog. "defektiven" Frikativen auch normale englische Frikativrealisationen zu identifizieren (= transkribieren!) hatten, wurden [f], [v], [z], [θ] und [ð] nur in jeweils 57, 50, 33, 30 und 13% der Fälle korrekt identifiziert: siehe D.J. Sharf, "Distinctiveness of 'Defective' Fricative Sounds", *Lang. & Speech*, 11 (1968), 38-45 (44). Daß unsere untrainierten Probanden signifikant bessere Hörleistungen erbracht haben, mag in der unterschiedlichen Art der Aufgabe begründet sein.

¹²Mohr und Wang haben in psychoakustischen Tests mithilfe der Methode des "semantic differential" die perzeptorische Distanz von Sprachlauten (Vokale und Konsonanten) zueinander ermittelt und diskutieren ihre Ergebnisse im Rahmen einer universalen Markiertheitstheorie für phonologische Merkmale. Die Ähnlichkeitsurteile der (untrainierten) Hörer ergaben eine geringere perzeptorische Distanz für das Konsonantenpaar [θ] - [f] als für das Paar [θ] - [t] ([s] gehörte nicht zu den getesteten Lauten). Nach diesen Ergebnissen ist [θ] [f] sogar ähnlicher als seiner stimmhaften Entsprechung [ð]. Siehe B. Mohr u. W.S.-I. Wang, "Perceptual Distance and the Specification of Phonological Features", *Phonetica*, 18 (1968), 31-45 (37).

¹³Siehe Nemser, *An Experimental Study of Phonological Interference*, S. 60-62.

auch wenn in den Kontexten #__r und V__s# die perzeptorische Trennung der möglichen Fehlerreflexe [f] und [t] für [θ] schwieriger zu sein scheint als in anderen Testkontexten.

Abb. 7 zeigt die unterschiedliche Anfälligkeit der einzelnen phonetischen Kontexte für perzeptorische Verwechslungen, wobei deutlich zutage tritt, daß für beide Gruppen von Vpp das Wirken der gleichen Gesetzmäßigkeiten angenommen werden

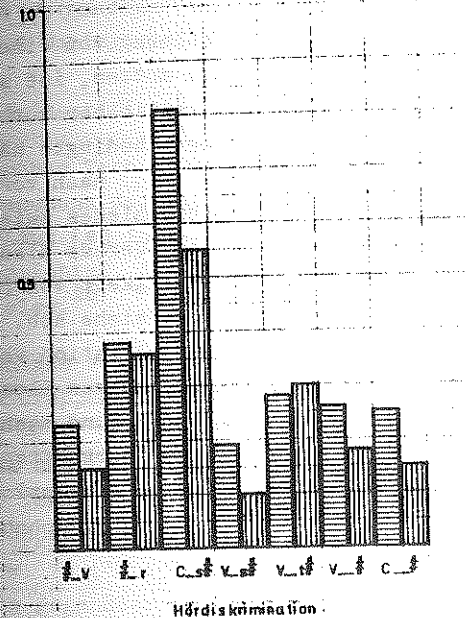


Abb. 7: Hördiskrimination: /θ/; Fehlerraten in den einzelnen Testkontexten ("LR": [solid bar]; "M": [hatched bar])

muß.¹⁴ Als besonders fehleranfällig erweisen sich für beide Gruppen von Vpp die Kontexte #__r und C__s#, wobei für den letzten Fall eine Erklärung darin gesucht werden muß, daß die Probanden nach vier Monaten Englischunterricht mit dem kritischen Laut in dieser höchst marginalen Umgebung noch nie konfrontiert worden waren, der im Englischen nur in flektierten Formen wie tenths, months, fifths etc. vorkommt. Die Tatsache, daß die Entscheidungen der Vpp hier in vielen Fällen offensichtlich zufällig gefällt wurden und der marginale Status dieses Kontextes überhaupt waren Anlaß genug, diese Umgebung bei allen weiteren Tests auszuschließen.

Die relativ hohe Fehleranfälligkeit der Umgebung #__r, mit der die Probanden zum Zeitpunkt der Untersuchung von einem Wort wie three her hinreichend vertraut sein mußten, ist schwieriger zu erklären, zumal im Kontext selber unmittelbar vorhandene phonetische Information - die kontextuell bedingte häufige Realisation von /r/ als "alveolar tap" ([r]) nach vorausgehendem /θ/¹⁵ - die Diskrimination des Unterschiedes zwischen /θr-/ /fr-/ und /tr-/ hätte erleichtern sollen. Gerade diese Tatsache kann jedoch die vergleichsweise hohe Häufigkeit von [t] als Fehlerreflex in dieser Umgebung erklären (siehe Abb. 6, S. 99), da in der norddeutschen Aussprache die Realisation von /r/ als [r] in der Umgebung #t__ (wie auch nach anderen Flosiven) häufig anzutreffen ist. Der hier erhobene Befund macht erneut deutlich, wie wichtig die Berücksichtigung herkunftsgebundener Unterschiede für die Erklärung spezifischer Perzeptions- und Produktionsmuster beim Erwerb neuer Aussprachegewohnheiten sein kann.

¹⁴ Der einzige Fall, bei dem sich eine phonetische Umgebung bei der Gruppe "M" als fehleranfälliger erweist als bei der Gruppe "LR", ist für den Kontext V__t# zu verzeichnen. Diese Unregelmäßigkeit ist leicht mit dem Hinweis auf die geringe Anzahl von Testwörtern mit [θ] in dieser Umgebung zu erklären, wo geringe Schwankungen in den absoluten Fehlerzahlen u.U. hohe Differenzen in den relativen Fehlerhäufigkeiten nach sich ziehen.

¹⁵ Vgl. Gimson (1970), S. 208. In der Realisationsnorm unseres Informanten wird tatsächlich /r/ im Kontext #θ__ als [r] realisiert.

Ein Vergleich der Leistungen beider Gruppen von Vpp in diesem Test ergibt, daß in nahezu allen getesteten phonetischen Kontexten die Hörerlebnisse der Gruppe "M" besser definiert sind als die der Gruppe "LR". Hierfür sind nicht nur die (signifikant) besseren Hörleistungen insgesamt ein Anzeichen, sondern vor allem auch die schärfere Trennung der (fehlerhaften) [f]-, [s]- und [t]- Reflexe untereinander, wie sie aus Abb. 6 unmittelbar herauszulesen ist. Inwieweit die bessere Fähigkeit der Gruppe "M", die wahrnehmungsrelevanten Eigenschaften aus den einzelnen Stimuluswörtern zu extrahieren und miteinander zu vergleichen, allein durch Aufmerksamkeitsfaktoren bedingt ist oder aber durch tatsächlich bessere Diskriminationsfähigkeiten bei dieser Gruppe, muß dahingestellt bleiben.

5.1.4. Höridentifikation

Gegenüber Test I wurde in Test II (Höridentifikation) auf Untersuchung des Identifikationsvermögens für den "kritischen" Laut [θ] in zwei spezifischen phonetischen Kontexten verzichtet, die im Englischen nur eine untergeordnete Rolle spielen.¹⁶

Die Ergebnisse von Test II sind in den Tabellen 17a und 17b für beide Gruppen von Vpp getrennt zusammengestellt, während Abbildung 8a (s. S. 105) die Art und Häufigkeit der Fehlerreflexe in den einzelnen Testkontexten darstellt. Für die gegenüber Test I signifikant schlechteren Hörleistungen beider Gruppen in diesem Test (Erfolgsquoten von "nur" 62 ("LR") bzw. 76 ("M") Prozent) kann, wie weiter oben gezeigt wurde, die gegenüber der einfachen Diskriminationsaufgabe erschwerte Aufgabenstellung verantwortlich gemacht werden.

Die Ergebnisse dieses Tests stimmen im übrigen gut mit denen

¹⁶ Außer der Umgebung C__s#, für deren Ausscheiden Gründe bereits genannt wurden, wurde auch auf den Kontext V__s# verzichtet, dessen Vorkommen im Englischen auf flektierte Formen beschränkt ist.

Kontext	TW	FZ	FR	Fehlerreflexe		
				/f/	/s/	/t/
*1 #_v	120	42	0.35	0.74	0.14	0.12
2 #_r	16	8	0.50	0.63	-	0.37
3 v_t#	16	7	0.44	1.00	0.00	-
*4 v_#	72	29	0.40	0.65	0.21	0.14
*5 c_#	80	33	0.41	0.24	0.45	0.31
Σ	304	119	0.39	0.65	0.16	0.19
gemeins. Kontexte	272	104	0.38	0.55	0.26	0.19
gewichtete Durchschn.	-	-	-	0.55	0.26	0.19

Tab. 17a: Fehlerraten und Fehlerreflexe: Höridentifikation (/θ/); Gruppe "LR" (8 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe		
				/f/	/s/	/t/
*1	300	58	0.19	0.90	0.07	0.03
2	40	16	0.40	0.81	-	0.19
3	40	15	0.37	1.00	0.00	-
*4	180	52	0.29	0.83	0.09	0.08
*5	200	51	0.26	0.57	0.24	0.19
Σ	760	192	0.25	0.79	0.11	0.10
gemeins. Kontexte	680	161	0.24	0.77	0.13	0.10
gewichtete Durchschn.	-	-	-	0.77	0.13	0.10

Tab. 17b: Fehlerraten und Fehlerreflexe: Höridentifikation (/θ/); Gruppe "M" (20 Vpp)

von Test I überein. Wiederum erweist sich [f] als der über-
ragende Fehlerreflex für [θ], wohingegen [s] und [t] nur selten
an Verwechslungen beteiligt sind. Die einzige Ausnahme bildet
der Kontext C_#, wo auslautendes postkonsonantisches [θ] bei
den Angehörigen der Gruppe "LR" häufiger mit [s] als mit [f]
verwechselt worden ist. Als Erklärung für diese Tatsache genügt
es nicht, auf das Testverhalten dieser Gruppe zu verweisen,

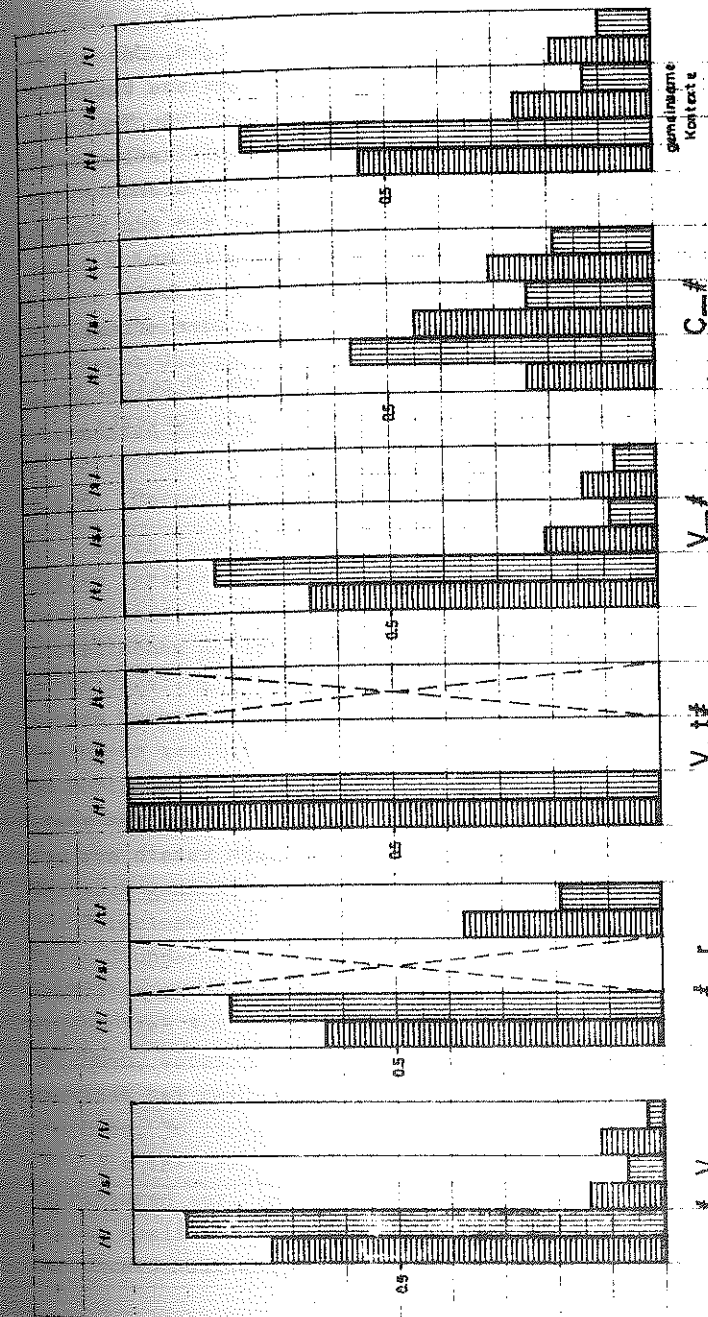


Abb. 8a: Höridentifikation: /θ/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten ("LR": ; "M":

die, wie weiter oben gezeigt wurde, im Verlauf dieses Tests eine zunehmende Testmüdigkeit an den Tag legte. Der beobachtete Effekt ist möglichenfalls auch auf gewisse Unterschiede der Testwörter von Test II gegenüber denen von Test I zurückzuführen. Vor allem wäre eine genauere Spezifikation des Konsonanten in diesem Kontext wünschenswert, da wir Grund zu der Annahme haben, daß in Perzeptionsaufgaben wie den hier gestellten die spezifischen phonetischen (segmentellen) Kontexte die Perzeption stärker beeinflussen als globale Positionseigenschaften wie wortinitial, -medial oder -final.

Eben diese Positionseigenschaften üben, wie Abb. 8b zeigt, wo die Fehleranfälligkeit der einzelnen phonetischen Kontexte

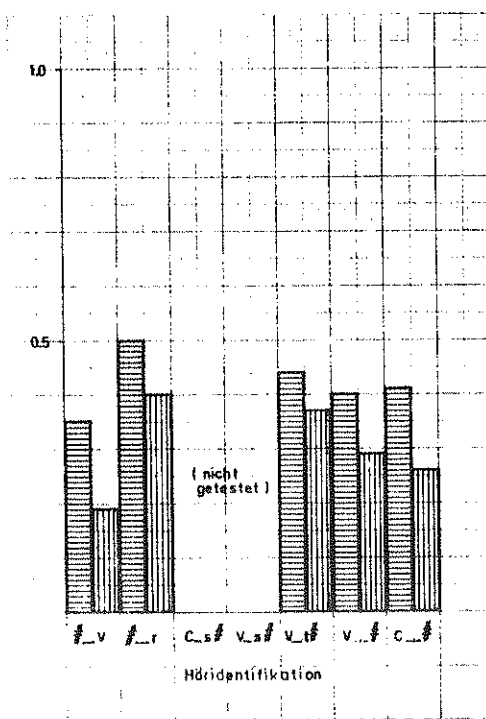




Abb. 8b: Höridentifikation: /θ/
Fehlerraten in den einzelnen Testkontexten
("LR":  ; "M": )

in Test II dargestellt ist, offensichtlich keinen tiefgreifenden Einfluß auf die Perzeption aus. Allein für die initiale Position ist (ähnlich wie auch in Test I) ein geringer Vorteil gegenüber den "medialen" Positionen, vielleicht auch der finalen Position, zu verzeichnen, wofür allgemeine informationstheoretische Aspekte der Sprache verantwortlich sein mögen, die den Empfänger einer Nachricht zwingen, dem Beginn einer Äußerung größere Aufmerksamkeit zu schenken. Inwieweit solche Gesetzmäßigkeiten, die für den Ablauf normaler Kommunikationsprozesse wichtig sind, in Testsituationen wie der unseren überhaupt Gültigkeit haben, wo die Aufmerksamkeit der Probanden zudem systematisch auf die im Material enthaltenen phonetischen Feinheiten gelenkt wird, muß jedoch in Frage gestellt werden.

5.1.5. Perzeption-Produktion: Nonsensewörter

Durch Verwendung von Nonsensewörtern ähnlich denen, die in beiden Perzeptionstests verwendet wurden, in diesem ersten Produktionstest konnte erreicht werden, daß /θ/ in sämtlichen phonetischen Kontexten mit Ausnahme der nach Test I ausgeschiedenen Umgebung C_s# in diesem Test zur Nachahmung angeboten werden konnte. Jedes Stimuluswort wurde den Probanden über Kopfhörer zweimal vorgespielt und von ihnen zweimal realisiert, um Hörfehler auszuschalten und den Vpp Gelegenheit zu geben, ihre erste Realisation ggf. zu korrigieren. Es zeigte sich, daß die Vpp von dieser Möglichkeit nur in einer geringen Zahl von Fällen Gebrauch gemacht haben und stattdessen ihre erste Realisation des Stimuluswortes nach erneutem Hören exakt reproduzierten. Beide Gruppen ("LR": 4 Vpp; "M": 8 Vpp) legten hier das gleiche Verhalten an den Tag. Bei insgesamt 408 Äußerungspaaren der Angehörigen der Gruppe "LR" waren nur 75 "Korrekturen" zu verzeichnen, wobei 39 nach Ansicht der Beurteiler ein besseres Resultat bei der zweiten Realisation des Stimuluswortes erbrachten, 36 hingegen ein schlechteres. Die 8 Vpp der Gruppe "M", die insgesamt 816 Äußerungspaare produzierten, weisen 136 Korrekturen auf, von denen 78 Verbesserungen gegenüber der ersten Realisation erbrachten, 58 hingegen Verschlechterungen. Der Unterschied zwischen erster und zweiter Realisationen ist

statistisch gesehen nicht signifikant (DIXON-MOOD-Vorzeichen-test), und der geringen Unterschiede wegen ist nichts dagegen einzuwenden, beide Realisationen mit in die Beurteilung einzubeziehen.

Tabelle 18a und b weist die Fehlerraten sowie die Art der Fehlerreflexe für die beiden Gruppen von Vpp nach, wobei, wie bereits aus oben angestellten Überlegungen zu erwarten war, die Art der beobachteten Produktionsfehler nicht in allen Fällen eine eindeutige Einordnung in vorgegebene Fehlerkategorien ermöglicht. Zwar fallen die meisten Produktionsfehler in die Phonembereiche der in den Perzeptionstests vorgegebenen englischen Konsonantenphoneme /f/, /s/ und /t/; es verbleibt jedoch eine "Unsicherheitszone" (24% aller Äußerungen der Mitglieder der Gruppe "LR", 11% der der Mitglieder von "M"), in der eine Kategorisierung der Produktionsfehler auf Schwierigkeiten stößt. Tabelle 19 (s. S. 109) gibt eine Übersicht über die hier verzeichneten Fehler für die einzelnen phonetischen Kontexte.

Kontext	TW	FZ	FR	Fehlerreflexe			
				/f/	/s/	/t/	and.
*1 #_v	40	23	0.58	0.83	0.13	0.00	(0.04)
2 #_r	24	13	0.54	0.23	0.15	0.39	0.23
3 v_s#	8	8	1.00	0.25	-	0.00	0.75
4 v_t#	8	7	0.88	1.00	0.00	-	0.00
*5 v_#	24	21	0.88	0.67	0.00	0.00	0.33
*6 c_#	24	23	0.96	0.65	0.00	0.00	0.35
Σ	128	95	0.74	0.63	0.05	0.00	0.27
gemeins. Kontexte	88	67	0.76	0.72	0.04	0.00	0.24
gewichtete Durchschn.	-	-	-	0.72	0.04	0.00	0.24

Tab. 18a: Fehlerraten und Fehlerreflexe: Perz - Prod (Nonsensewörter): /θ/; Gruppe "LR" (4 Vpp)¹⁷

¹⁷ Eingeklammerte Häufigkeitszahlen deuten an, daß es sich hierbei um Fehlrealisationen einer einzigen Vp handelt.

Kontext	TW	FZ	FR	Fehlerreflexe			
				/f/	/s/	/t/	and.
*1	80	11	0.14	0.82	(0.18)	0.00	0.00
2	48	7	0.15	0.00	1.00	0.00	0.00
3	16	13	0.81	0.85	-	0.00	(0.15)
4	16	14	0.88	0.93	0.00	-	(0.07)
*5	48	22	0.41	0.73	0.09	0.00	0.18
*6	48	41	0.85	0.90	0.00	0.00	0.10
Σ	256	108	0.42	0.80	0.10	0.00	0.10
gemeins. Kontexte	176	74	0.42	0.82	0.09	0.00	0.09
gewichtete Durchschn.	-	-	-	0.84	0.05	0.00	0.11

Tab. 18b: Fehlerraten und Fehlerreflexe: Perz - Prod (Nonsensewörter): /θ/; Gruppe "M" (8 Vpp)

Kontext	Fehler	Häufigkeit
1 #_v	[pf]	1
2 #_r	[g'] [pr] [ϕ]	je 1
3 v_s#	[fts] [fst] [fθ] [st] [gt]	3 2 je 1
4 v_t#	[fθt]	1
5 v_#	[ft] [fs] [θt] [fθ] [ϕ]	4 3 2 je 1
6 c_#	[mfθ] ¹⁸ [mf] [mpf] [ϕ]	6 je 2

Tab. 19: Idiosynkratische Produktionsfehler: Perz - Prod (Nonsensewörter); Fehlerdaten beider Gruppen

¹⁸ Testwort: /nanθ/.

Die Aufstellung zeigt, daß diese Fehler seltener in initialer als in medialer und finaler Position anzutreffen sind und gehäuft dort auftreten, wo /θ/ Element einer Konsonantenverbindung (z.B. /θs/, /θt/, /pθ/, /nθ/) ist: Die Identifikation und korrekte Nachahmung solcher Clusters scheint größere Schwierigkeiten zu bereiten als die Produktion von /θ/ an der Wortgrenze (oder) in der Nachbarschaft von Vokalen. Eine Ausnahme stellt allerdings auslautendes postvokalisches /θ/ dar, das insbesondere von den Angehörigen der Gruppe "LR" häufig als komplexe Konsonantenverbindung realisiert wird.

Abb. 9 stellt die Fehleranfälligkeit der einzelnen phonetischen Kontexte für /θ/ in diesem Test dar. Während für die Angehörigen der Gruppe "LR" für alle Kontexte schlechtere Produktions- als Perzeptionsleistungen festzustellen sind und überhaupt nur 26% der Nachahmungen als korrekt eingestuft wurden, produzierten

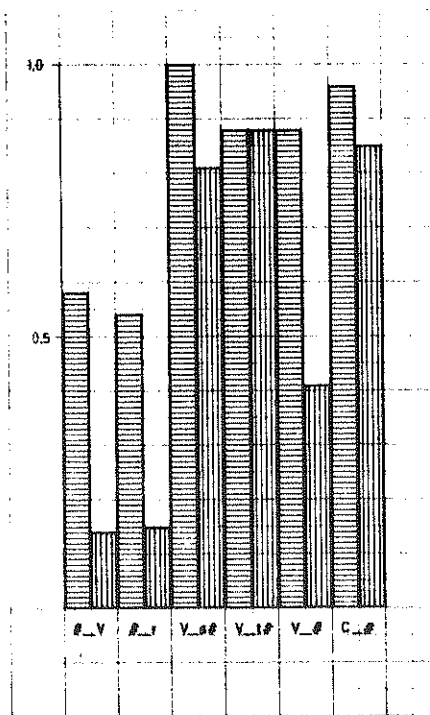


Abb. 9: Perz - Prod (Nonsensewörter): /θ/; Fehlerraten in den einzelnen Testkontexten ("LR": ■; "M": ▨)

die Angehörigen der Gruppe "M" den kritischen Laut in immerhin 58% der Fälle korrekt.¹⁹ Besonders bemerkenswert ist dabei die geringe Anzahl von Produktionsfehlern in den Kontexten #_V und #_r, also für den kritischen Laut in initialer Position, der hier auch von den Angehörigen der Gruppe "LR" häufiger richtig wiedergegeben wurde als in anderen Umgebungen. Für die Gruppe "M" ist die relative Fehlerhäufigkeit für die initial-prävokalische Position (wenn auch nur geringfügig) niedriger als die in beiden Perzeptionstests ermittelte, und noch bemerkenswerter erscheint die gegenüber Test I und II deutlich verbesserte Produktionsleistung für /θ/ im Kontext #_r. Da korrekte Imitation richtige Identifikation voraussetzt, kann die Verbesserung der Leistungen in diesen beiden Kontexten bei den Angehörigen von "M" allenfalls auf die andere Zusammensetzung der (verringerten) Gruppe zurückgeführt werden, die, wie weiter oben (S. 42) gezeigt wurde, vielleicht ein leichtes Übergewicht zugunsten der mit besseren Hörleistungen aus den Perzeptionstests hervorgegangenen Probanden zeigt.

Die Fehlerdaten der Tabellen 18a und b (s.o., S.108f.) sind in Abb. 10 (s.S. 112 - 113) nochmals für alle Testkontexte dargestellt. Von einem radikal veränderten Bild der Produktionsfehler gegenüber den in den Hörtests gemachten Perzeptionsfehlern kann keine Rede sein;²⁰ [f] ist in der Mehrzahl der Kontexte nach wie vor der am häufigsten für [θ] substituierte Laut, während Substitution von [s] relativ selten vorkommt und sich solche Substitutionen nur bei einzelnen Angehörigen der je-

¹⁹Vgl. Nemser's Zahlen für einen ähnlichen "Repetition Test": durchschnittliche Fehlerhäufigkeit 79%, mit nur geringer Variation in den einzelnen Testkontexten (Nemser (1971), S. 81).

²⁰Nemser hat in seiner Untersuchung zur Kontaktsituation Ungarisch-Englisch für eben diese Art von Imitationsaufgabe ein tatsächlich verändertes Fehlerbild gegenüber den Hörtestbefunden ermittelt. Überwogen bei den letzteren wie auch in unseren Tests zunächst die [f]-Reflexe, waren bei der Imitation mehr [s]- als [f]-Substitutionen zu verzeichnen. Vgl. Nemser (1971), S. 74-81.

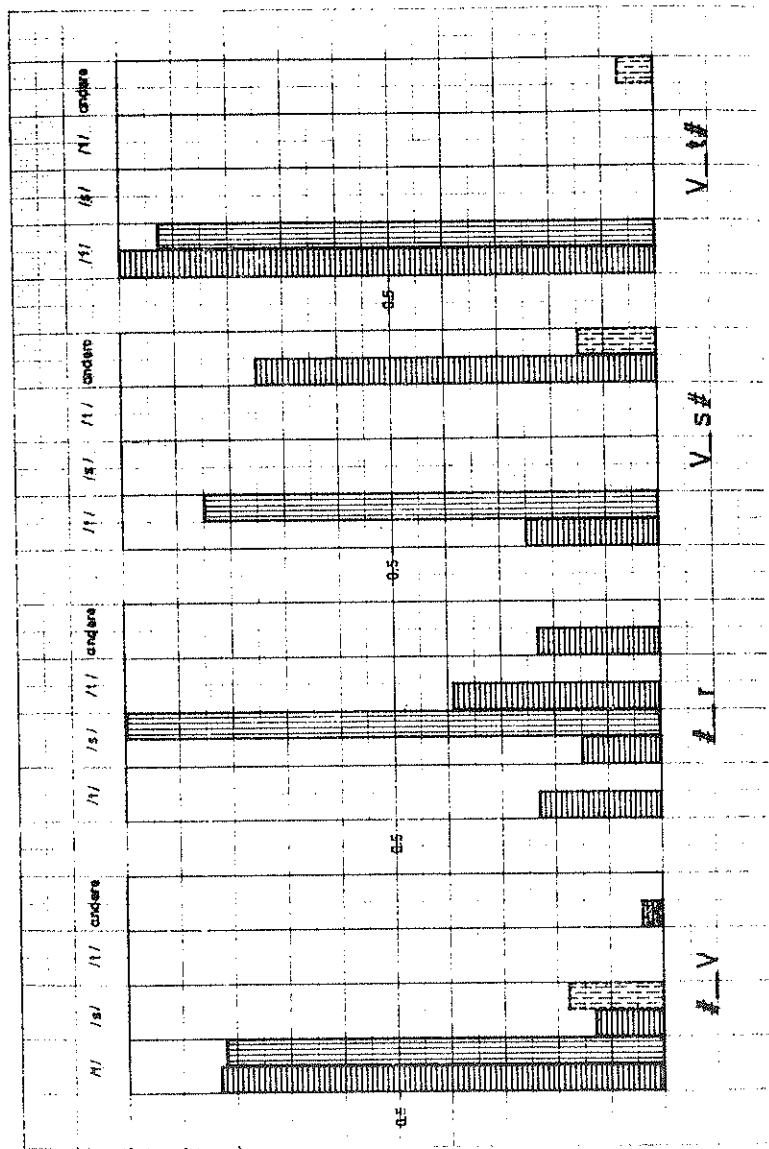


Abb. 10: Perzeption - Produktion (Nonsensewörter): /θ/;
 Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten
 ("L": ; "M": ; "N":)²¹

²¹Die mit unterbrochenen Linien gezeichneten Blöcke deuten an, daß es sich hierbei um Fehlrealisationen einer einzigen Vp handelt.

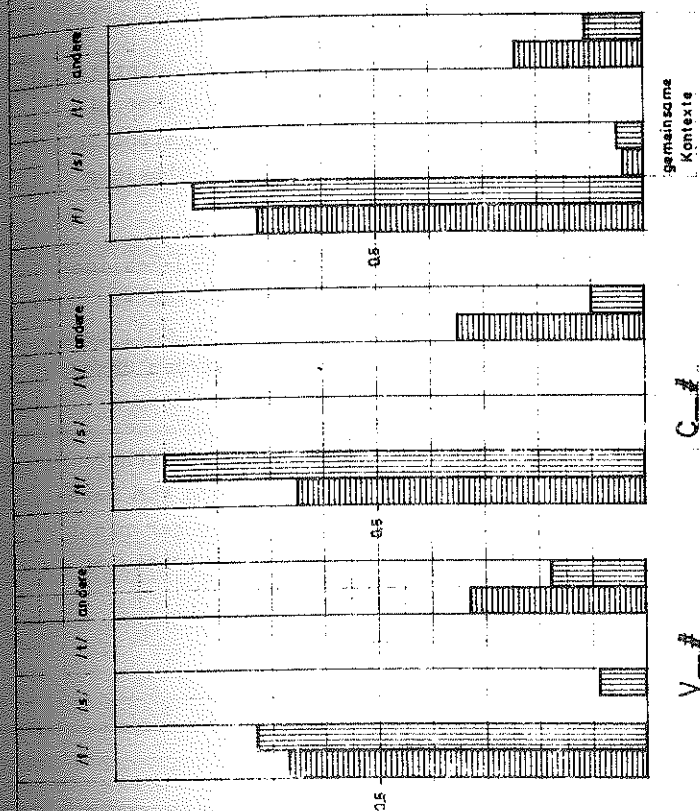


Abb. 10 (Fortsetzung)

weiligen Gruppe finden. [t] scheidet als Fehlerreflex für die Angehörigen von "M" vollständig aus und ist nur im Kontext #_r bei zwei Angehörigen der Gruppe "LR" anzutreffen.

Nicht erklärt werden kann die geradezu kategorische Substitution von [s] für [θ] in initialer Position vor /r/ durch immerhin drei Angehörige von "M", die auch mit den Perzeptionsfehlern der entsprechenden Vpp für diesen Kontext nicht in Einklang gebracht werden kann.

Verstöße gegen die "Stimmbeteiligungsart" in den [θ]-Realisationen der Probanden wurden nur in drei Fällen beobachtet (zweimal [ð], einmal [z] jeweils in initialer Position); beide Fälle, in denen [ð] für [θ] substituiert wurde, wurden dennoch als korrekte Realisationen angesehen.

Hinzuweisen ist auch auf die Tatsache, daß der "kritische" Laut [θ] häufig Anlaß zu Fehlrealisationen (in Form von Metathesen und Substitutionen) dort gibt, wo er selbst gar nicht gefordert wird und an die Stelle von [s] oder [f] tritt. So finden sich Realisationen der Testwörter

- /θrɪf/ als [θrɪθ]
- /θa:ft/ als [fa:θt]
- /nu:fs/ als [nu:fθ], [nu:θfs], [nu:θ]
- /sɔ:θ/ als [θɔ:f], [θɔ:θt], [θɔ:θ]
- /bru:s/ als [bru:θ]
- /s:θ/ als [θɜ:θ]
- /saɪðʃ/ als [θaɪft]
- /skraɪt/ als [θkwaɪk].²²

Bei Verwechslungen von [θ] und [f] können Identifikationsfehler zur Erklärung solcher Fehlrealisationen führen, kaum aber für Substitutionen von [θ] für [s], wie sie in unseren Testmaterialien häufig zu verzeichnen sind. Bei diesen "Versprechern" kann

²²Die letzten drei Äußerungen sind für eine Angehörige der Gruppe "LR" (SA₂) belegt, die zusätzlich zu den vier Mitgliedern dieser Gruppe zu Produktionstests herangezogen wurde. An der Fehlerwertung nimmt diese Vp nicht teil.

man von Überkompensation seitens des Lernenden sprechen, die durch die Anwesenheit von [θ] an anderer Stelle im Wort oder überhaupt im größeren Kontext des gesamten Tests bewirkt wird. Fehler dieser Art sind im übrigen nicht auf sprachliches Material wie das hier verwendete Nonsensematerial beschränkt und finden sich auch in dem englischen Material (Einzelwörter und Sätze) der Tests IV und V. Jeder Englischlehrer ist mit Realisationen wie [θɪθ rɔ̃ mæ pənθɪl] für This is my pencil oder [θɪθtɪ] für sister vertraut, wie sie sich in unserem Testmaterial häufig finden. Sie zeigen, wie bewußt das Bemühen des Lerners ist, sich die ungewohnten Artikulationen anzueignen; bemerkenswert dabei ist, daß das kritische [θ] dabei ausschließlich für [s] eintritt - die häufigste Fehlrealisation von [θ].

5.1.6. Perzeption-Produktion: englische Einzelwörter

Die Verwendung von "echten" englischen Wörtern in den beiden letzten Produktionstests machte des geringen Ausbildungsstandes der Probanden wegen Kürzungen des Testmaterials und das Ausscheiden mancher phonetischer Kontexte notwendig, die unsere Aussagen über Art und Häufigkeit der hier produzierten Fehlerreflexe in ihrer Verlässlichkeit stark einschränken: Die hier erhobenen Befunde sind in vieler Hinsicht lückenhaft und bedürfen dringend der Ergänzung durch weitere Materialerhebungen. Wir glauben dennoch, daß die Ergebnisse dieser kurzen Tests einige Einsicht in die Dynamik des Lernprozesses gestatten und zumindest die Formulierung von Hypothesen erlauben, die in weiterführenden Untersuchungen auf ihre Gültigkeit hin geprüft werden sollten.

Die auditive Darbietung fremder Stimuluswörter in einer Testsituation, in der die Vpp aufgefordert werden, ihr Diskriminationsvermögen für Lautpaare wie [θ]-[f] unter Lewis zu stellen oder solche Laute zu identifizieren, stellt eine außerordentliche Verschärfung der Situation solchen Bedingungen gegenüber dar, wie sie in natürlichen Kommunikationssituationen und auch in der spezifischen Situation des Fremdsprachenunterrichts anzutreffen sind. Wo die visuelle Information über den Artikulationsort

vorenthalten wird und die Probanden ihre Entscheidungen allein anhand der auditiven Eindrücke vornehmen müssen, ist auch für Sprecher/Hörer, die mit den genannten Distinktionen wohlvertraut sind, mit hohen Fehlerraten zu rechnen. Verbesserte Hör- und Identifikationsleistungen sind allerdings dort zu gegenwärtigen, wo den Vpp ihre Kenntnis der Stimuluswörter, die den kritischen Laut enthalten, zu Hilfe kommt und die Semantisierbarkeit der ihnen vorgespielten Testwörter eine eindeutige Identifikation des Testlautes erlaubt.

Diesen Effekt zeigen deutlich die Ergebnisse von Test IV, der mit englischem Wortmaterial (Einzelwörtern) durchgeführt wurde, das allen Vpp vertraut war. Die Fehlerraten für die vier phonetischen Kontexte, in denen /θ/ enthalten war, zeigt Tabelle 20. Die Ergebnisse sind um so eindrucksvoller, als die schnelle Abfolge der den Vpp über Kopfhörer vorgespielten isolierten

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_v	12	1	0.08	24	-	0.00
2 #_r	4	2	0.50	8	1	0.13
3 v_#	8	2	0.25	16	-	0.00
4 c_#	4	-	0.00	8	-	0.00
Σ	28	5	0.18	56	1	0.02

Tab. 20: Perzeption-Produktion (engl. Einzelwörter): /θ/; Fehlerraten in den einzelnen Testkontexten

Einzelwörter eine rasche Identifikation und spontane Imitation verlangte. Einige der hier beobachteten Aussprachefehler sind möglichenfalls denn auch auf Identifikationsfehler zurückzuführen (so [frl:] für [θrl:], das zweimal vorkam); ansonsten wurde zweimal [s] für [θ] substituiert, einmal (mit gleichzeitigem Verstoß gegen die Stimmbeteiligungseigenschaften) [z] und ein weiteres Mal [f].

Welch bedeutende Rolle die Semantisierbarkeit der Testwörter für die korrekte Identifikation und Nachahmung und auch für die Art der Produktionsfehler spielt, verdeutlichen die Fehlerzahlen für zwei zusätzlich eingefügte Testwörter mit /θ/ in finaler postvokalischer Stellung, bei denen es sich zwar um englische Wörter handelt, die den Vpp jedoch noch nicht bekannt waren (es handelt sich um die Wörter north und south, die nicht in die Wertung einbezogen wurden). Sie wurden nur in 50% ("LR") bzw. 75% ("M") der Fälle korrekt wiedergegeben, was das auslautende /θ/ anbetrifft, wobei für letzteres fünfmal [f] substituiert wurde und je einmal [θ̥], [θ̄] und [θ̥̄]. Die Fehler-typik ähnelt hier stark der des Nonsensewörter-Tests und weist erneut darauf, daß für auditiv dargebotene Stimuli, die den Vpp unbekannt sind und für sie keine semantische und/oder orthographische Repräsentation haben, besondere Gesetzmäßigkeiten gelten.

5.1.7. Produktion-Perzeption-Produktion: englische Sätze

Erst die in Test V (englische Sätze) erhobenen Produktionsdaten bringen das aus der Unterrichtssituation vertraute Fehlerbild an den Tag: Die zuerst von Karten, auf denen die Testsätze in orthographischer Form vorlagen, abgelesenen und nach einer Modellrealisation vom Tonband erneut produzierten Sätze weisen in der verhältnismäßig geringen Zahl von Fällen, wo sie fehlerhaft realisiert wurden, nahezu ausschließlich Substitution von [s] und s-artigen Lauten für das geforderte [θ] auf (82% aller Fehlrealisationen);²³ die restlichen Produktionsfehler umfassen hingegen f-artige Laute bzw. Kontaminationen wie [θ̄] und [θ̥].²⁴

²³Wir sehen im Augenblick davon ab, daß manche dieser Fehler bei den ersten Realisationen (beim Lesen der Testsätze) auftreten und sich in den zweiten Realisationen nicht wieder zeigen.

²⁴Ein aufschlußreicher Fall von Metathese findet sich in der Wiedergabe des Testsatzes "Are you fat or thin?" durch einen Angehörigen von "M" als [ə ju fət o θin].

Die Fehlerraten in den drei Testkontexten für die Zweitrealisationen der 12 Vpp verzeichnet Tab. 21.²⁵ Die insgesamt geringe Zahl von Fehlrealisationen, die sich zudem konzentriert bei einzelnen Angehörigen der beiden Gruppen finden und nicht gleichmäßig gestreut sind, ist zweifellos darauf zurückzuführen, daß Fehlidentifikationen bei diesem Test, dessen Material den Vpp in verschrifteter Form vorlag, ausgeschlossen waren.

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_v	16	4	0.25	32	1	0.03
2 #_r	4	3	0.75	8	1	0.13
3 v_#	8	2	0.25	16	3	0.19
Σ	28	9	0.32	56	5	0.09

Tab. 21: Prod. - Perz. - Prod. (engl. Sätze): /θ/; Fehlerraten in den einzelnen Testkontexten

5.1.8. Beurteilung der /θ/-Realisationen nach Akzeptabilität

Die lückenlose Protokollführung über die /θ/- und /ð/-Realisationen der zwölf Probanden bei der Durchführung von Test III (Nonsensewörter) bewahrte die visuelle Information über den Artikulationsort unzweideutig und konnte bei der späteren auditiven Auswertung der Schülerproduktionen durch den Verfasser zur Beurteilung mit herangezogen werden. Dem zweiten Beobachter lag diese Information nicht vor; der Vergleich mit den Protokollunterlagen zeigte, daß er dennoch in - 94% aller zu beurteilenden Fälle ein richtiges Urteil über den hier beteiligten phonetischen Parameter Artikulationsort abgegeben hatte. In der Fehlerstatistik werden die 6% abweichend beurteilten Fälle nicht berücksichtigt, die der

²⁵Gründe für die Nichtberücksichtigung der Erstrealisationen (beim Ablesen der Sätze) werden ausführlicher weiter unten, S. 182 gegeben.

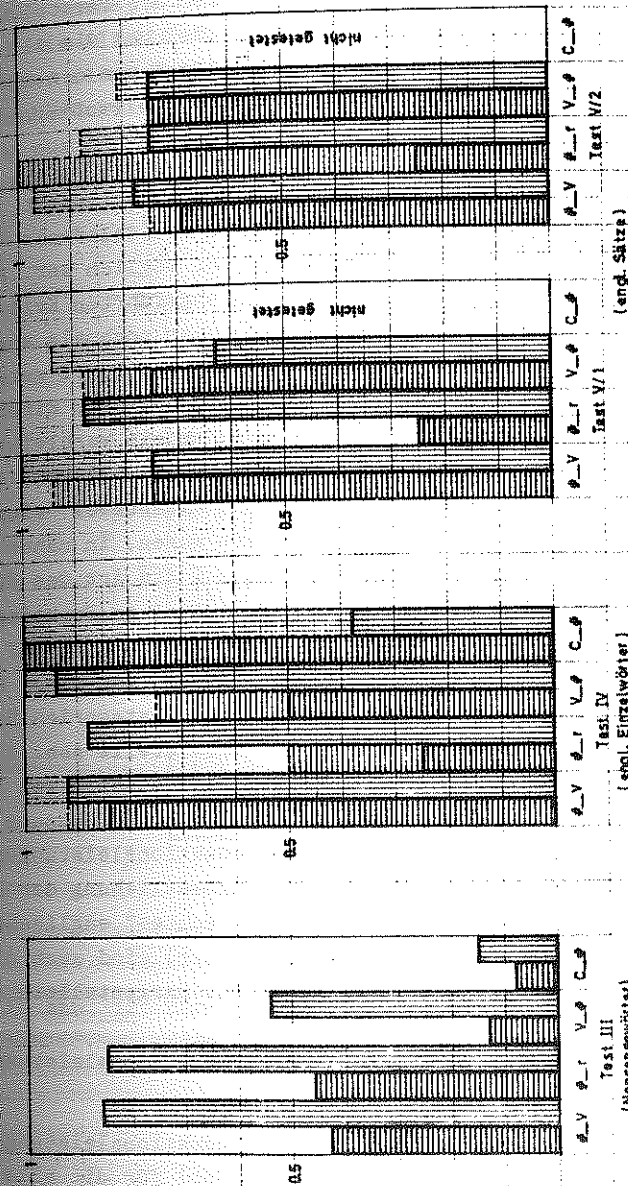


Abb. 11: Akzeptabilitätsbeurteilung der Produktionsleistungen: /θ/ ("LR": ; "M":)

Einfachheit halber anhand der Protokollunterlagen entschieden wurden; es wurde auch darauf verzichtet, bei der Darstellung der Produktionsleistungen (s.o., Abb. 11, S. 119) die als "(noch) akzeptabel" eingestuften Realisationen von den als "gut" bezeichneten getrennt darzustellen.

Für Test IV und V lagen vollständige Protokollunterlagen, in denen die wichtige Information über den Artikulationsort festgehalten wurde, nicht vor, so daß eine zweifelsfreie Entscheidung über die Schülerproduktionen oft nicht möglich war. Als "nichtübereinstimmend" im Sinne des oben angeführten Kriterienvergleichs beurteilte Äußerungen konnten zwar nochmals abgehört und dem dritten Beobachter zur Entscheidung zugeführt werden; alle Urteile wurden hier jedoch anhand der auditiven Eindrücke gefällt, so daß ein gewisses Maß an Unsicherheit verbleibt. So ist insbesondere nicht von vornherein auszuschließen, daß manche vom zweiten Beobachter als "(noch) akzeptabel" eingestuften Äußerungen tatsächlich fehlerhaft waren. In nahezu allen Fällen wurde bei solchen Urteilen auf die Art des Friktionsgeräusches Bezug genommen (häufig "Mangel an Friktion", aber auch "zusätzliche alveolare Friktion"), die auch auf Beteiligung des Ortsparameters schließen lassen. Von einer kategorischen Zuweisung dieser Äußerungen in die Kategorien "falsch" oder "richtig" wurde abgesehen; stattdessen stellt Abb. 11 die abgestuften Urteile für die Schülerrealisationen von /θ/ in den getesteten Umgebungen dar, die auch zur Berechnung der Indexwerte, mit der die Güte der Produktionsleistungen erfaßt wird, herangezogen wurden. Zur Verdeutlichung der Leistungsverbesserung in Test IV und V (mit "echtem" englischen Material) gegenüber Test III (Nonsensewörter) sind auch die Leistungsdaten für den letztgenannten Test dargestellt.

5.1.9. Vergleich der Perzeptions- und Produktionsleistungen (/θ/)

Abbildung 12 stellt die Hör- und Produktionsleistungen der beiden Gruppen ("LR": 4 Vpp; "M": 8 Vpp) für die allen fünf Tests gemeinsamen Kontexte von /θ/ dar. Von den maximal sieben

Kontexten, in denen /θ/ zur Beurteilung oder Nachahmung angeboten wurde, konnten nur zwei berücksichtigt werden, nämlich die allen Tests gemeinsamen Umgebungen #_V und V_r (der allen Tests ebenfalls gemeinsame Kontext #_r blieb unberücksichtigt, weil die Vpp zum Zeitpunkt der Untersuchung nur ein entsprechendes Wort - three - in ihrem Vokabular hatten).

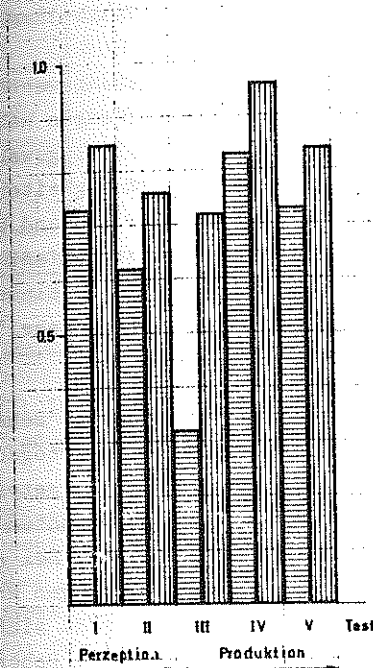


Abb. 12: Hör- und Produktionsleistungen: /θ/ ("LR": [horizontal lines]; "M": [vertical lines])

Im einzelnen liegen den individuell errechneten und über die Gruppe gemittelten Leistungswerten zugrunde:

- Test I (Hördiskrimination): 324 Hörerurteile;²⁶
 Test II (Höridentifikation): 288 Hörerurteile;
 Test III (Produktion: Nonsensewörter): 86 Realisationspaare;
 Test IV (Produktion: engl. Einzelwörter): 60 Realisationen;
 Test V (Produktion: engl. Sätze): 72 (Zweit-)Realisationen.

In Tabelle 22 sind die individuellen Leistungswerte für die zwölf Vpp, die sich allen fünf Tests unterzogen haben, zusammengestellt. Als Grundlage der Perzeptionsleistungswerte dienen die individuell ermittelten relativen Häufigkeiten erfolgreich diskriminierter bzw. identifizierter /ø/, während die individuelle Produktionsleistung durch Indexwerte erfaßt wird, die die als "gut" bzw. "nicht abweichend" beurteilten Einzelrealisationen von /ø/ vollständig und die als "(noch) akzeptabel" bezeichneten zur Hälfte berücksichtigen.²⁷

Faßt man die einzelnen Tests als verschiedene "Behandlungen" auf, denen die einzelnen Individuen sich unterzogen, kann mithilfe der Rangvarianzanalyse nach FRIEDMAN getestet werden, ob die einzelnen Behandlungen bei den untersuchten Individuen oder Gruppen unterschiedliche Effekte zeigen. Bei Annahme der Nullhypothese, daß die verschiedenen Bedingungen keinen Einfluß auf die Verteilung der Meßwerte nehmen, werden sich die Rangplätze der n Individuen nach Zufall auf die k Behandlungen verteilen, d.h. die Rangsummen für jede der k Behandlungen werden nicht oder nur zufällig voneinander abweichen. Üben jedoch einzelne Behandlungen einen systematischen Einfluß aus, so werden

²⁶Nur die Hörleistungen derjenigen Probanden, die sich nach der Serie von Hörtests auch allen Produktionstests unterzogen, finden hier Berücksichtigung.

²⁷Beispiele zur Errechnung dieser Indexwerte siehe oben, 4.6.0., S. 92. - Bei Test III wurden nur die vom Verf. in Übereinstimmung mit den Protokollunterlagen als "nicht abweichend" beurteilten Realisationen berücksichtigt (s.o., S.119).

die k Spalten unterschiedliche Rangsummen aufweisen.²⁸

Gruppe	Test Vp	I	II	III ²⁹	IV	V ³⁰
		"LR"	SA ₁	0.70	0.46	0.38
SS ₂	0.78		0.67	0.63	0.90	0.83
SS ₃	0.63		0.62	0.19	1.00	0.83
WB	0.81		0.71	0.06	0.40	0.25
"M"	BM	0.93	0.71	0.94	0.90	1.00
	DL	0.78	0.92	0.84	1.00	1.00
	DM	0.81	0.75	0.38	0.90	0.50
	HH	0.78	0.58	0.81	1.00	0.92
	HU	0.81	0.83	0.44	1.00	0.75
	JK	0.93	0.79	0.78	1.00	0.83
	RR	0.89	0.75	0.94	0.90	1.00
SM	0.85	0.75	0.69	1.00	0.75	
MW ³¹		0.81	0.71	0.59	0.92	0.81
StA		0.09	0.12	0.29	0.17	0.23

Tab. 22: Individuelle Perzeptions- und Produktionsleistungen: /ø/

Dieser Test erweist sich als besonders nützlich, wo individuelle Unterschiede zwischen einzelnen Vpp einer Testpopulation größer sind als die durch die verschiedenen Behandlungen verursachten Unterschiede in den Meßwerten für ein und dieselbe Vp: Der Test sieht von den großen, irrelevanten Unterschieden zwischen einzelnen Vpp ab und konzentriert sich auf die kleinen, relevanten, durch die unterschiedliche Be-

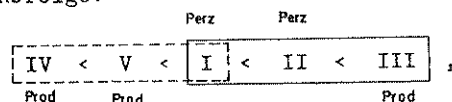
²⁸Siehe Sachs, Statistische Auswertungsmethoden, S. 422.

²⁹Erst- und Zweitrealisationen zusammengefaßt.

³⁰Nur Zweitrealisationen erfaßt.

³¹Abkürzungen: MW = Mittelwerte; StA = Standardabweichungen.

handlung verursachten Unterschiede.³² - Die Rangvarianzanalyse nach FRIEDMAN ergibt einen signifikanten Befund für den unterschiedlichen Einfluß der fünf "Behandlungen" (= Einzeltests) auf die Leistungen der Probanden ($\hat{\chi}_R^2 = 19.65$, $\chi_{4,0.05}^2 = 9.49$). Kann man von den individuell ermittelten Leistungsunterschieden unmittelbar auf die Schwierigkeit der in den einzelnen Tests gestellten Aufgaben schließen, so ergibt sich mithilfe der multiplen Vergleiche nach WILCOXON-WILCOX als Rangfolge für die Schwierigkeiten der einzelnen Aufgaben, die durch die Nummer des entsprechenden Tests bezeichnet werden, die folgende Abfolge:



wobei "<" als "leichter als" zu verstehen ist. Als signifikant leichter erweist sich damit die Produktionsaufgabe der beiden mit "echtem" englischen Wortmaterial durchgeführten Tests sowohl gegenüber der Identifikationsaufgabe (Test II) als auch gegenüber der Produktionsaufgabe des Nonsensewörter-Tests (vgl. auch Abb. 11, S. 119).³³ Das Ergebnis läßt eine deutliche Abhängigkeit der Leistungen von dem in den einzelnen Tests verwendeten sprachlichen Material erkennen, dessen Einfluß sich nicht nur in der Häufigkeit, sondern vor allem auch in der Art der verzeichneten Perzeptions- und Produktionsfehler zeigt. Überwiegen in beiden Perzeptionstests sowie in dem mit ähnlich strukturiertem Material veranstalteten ersten Produktionstest ganz eindeutig Fehler der Art, die auf eine auditive Verarbeitung der wahrgenommenen Stimuluswörter schließen lassen und im Fall des Nonsensewörter-Tests eine spontane Nachahmung des Wahrgenommenen mit dem Ziel der Produktion des gleichen auditiven Effekts nach sich ziehen, produzieren wenigstens einige der Probanden in den beiden letzten Produktionstests mit englischen

³²Vgl. J.V. Bradley, Distribution-free Statistical Tests, Englewood Cliffs 1968, S. 128.

³³Signifikante Befunde erhält man auch, wenn man den Test für jede Gruppe getrennt durchführt, wo allerdings nur für Test IV gegenüber Test III ein signifikanter Unterschied besteht und die übrigen drei Tests sich nicht signifikant voneinander unterscheiden.

Material die vertrauten [s]-Substitutionen deutscher Englischlernender, die auf eine ganz andere Verarbeitung der ebenfalls auditiv oder in orthographischer Form dargebotenen Stimuluswörter schließen lassen. Auf die weiteren Implikationen dieses beobachteten Fehlerverhaltens werden wir weiter unten noch zu sprechen kommen.

5.1.10. Perzeption:/ð/

Die einzelnen Testkontexte für die "stimmhafte" Entsprechung des stimmlosen interdentalen Frikativs, /ð/, sind aus Tabelle 6, S. 52 zu entnehmen. Als Testwörter, in denen /ð/ in initialer prä vokalischer Stellung vorkam, wurden nur "echte" englische Wörter aus der geschlossenen Klasse der sog. "Formwörter" verwendet, von denen diejenigen ausgewählt wurden, die den Vpp noch nicht bekannt waren (z.B. thence, thou, thus usf.).

5.1.11. Fehlervorgabe

Bei der Fehlervorgabe für beide Perzeptionstests wurde von den gleichen Voraussetzungen wie bei /ð/ ausgegangen. Fehlerbeobachtungen im Fremdsprachenunterricht legen eine Überprüfung der Frage nahe, inwieweit Substitution von [z] und gelegentlich auch [d] in der Lautproduktion bei deutschen Englischlernenden eine Entsprechung im perzeptorischen Bereich hat. [d] ist darüber hinaus der von englischsprachigen Kindern im Prozeß des Spracherwerbs am häufigsten für [ð] substituierte Laut, aber auch der labiodentale Frikativ [v] ist als Fehlerreflex im Erstspracherwerb ausreichend belegt.³⁴ [v], [z] und [d] kommen somit als mögliche Fehlerreflexe für [ð] in Frage und wurden in den beiden Perzeptionstests dem "kritischen" Laut [ð] gegenübergestellt.

5.1.12 Hördiskrimination

In Tabelle 23a und b sind die Ergebnisse von Test I zur Hördiskrimination von /ð/ für die beiden Gruppen von Vpp

³⁴Vgl. Ferguson (1973), 66-70; Olmsted (1971), S. 71 (Tab. 13).

zusammengestellt; Abb. 13 stellt die Fehlerdaten in Form von Flächendiagrammen für jeden der Testkontexte dar und ermöglicht einen unmittelbaren Vergleich der beiden Gruppen.

Kontext	TW	FZ	FR	Fehlerreflexe		
				/v/	/z/	/d/
*1 #_V	144	46	0.32	0.56	0.22	0.22
2 V_z#	32	7	0.22	0.57	-	0.43
3 V_d#	32	5	0.16	1.00	0.00	-
*4 V_#	144	29	0.20	0.59	0.31	0.10
Σ	352	87	0.25	0.60	0.22	0.18
gemeins. Kontexte	288	75	0.26	0.58	0.26	0.16
gewichtete Durchschn.	-	-	-	0.58	0.25	0.17

Tab. 23a: Fehlerraten und Fehlerreflexe: Hördiskrimination (/ð/); Gruppe "LR" (8 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe		
				/v/	/z/	/d/
*1	360	65	0.18	0.68	0.09	0.23
2	80	21	0.26	0.76	-	0.24
3	80	21	0.26	1.00	0.00	-
*4	360	36	0.10	0.75	0.14	0.11
Σ	880	143	0.16	0.75	0.08	0.17
gemeins. Kontexte	720	101	0.14	0.72	0.11	0.17
gewichtete Durchschn.	-	-	-	0.71	0.11	0.18

Tab. 23b: Fehlerraten und Fehlerreflexe: Hördiskrimination (/ð/); Gruppe "M" (20 Vpp)

Erneut erstaunt das über Erwarten gute Diskriminationsvermögen der Angehörigen beider Gruppen, die den kritischen Laut in 74 ("LR") bzw. 86% der Fälle ("M") erfolgreich diskriminieren; die Angehörigen der Kontrollgruppe weisen eine Erfolgsquote von 95% auf.

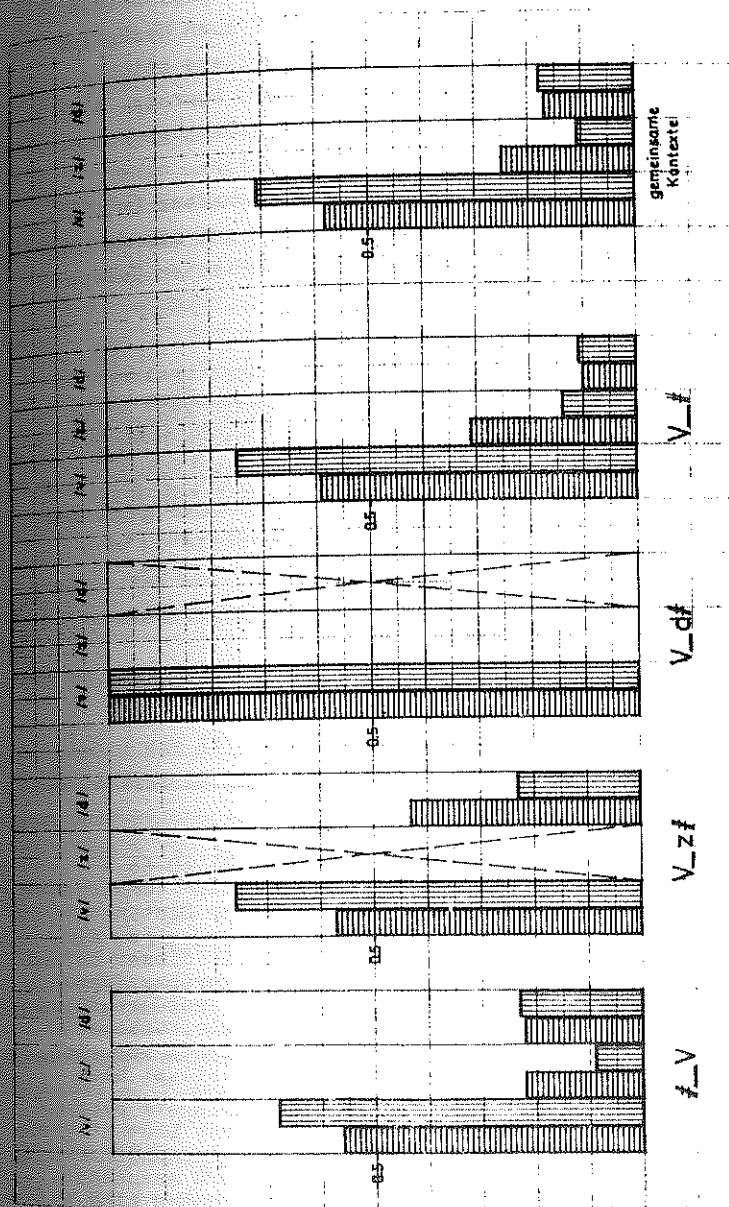


Abb. 13: Hördiskrimination: /ð/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten ("LR": ; "M":)

In sämtlichen getesteten Umgebungen erweist sich bei beiden Gruppen [v] als der häufigste Fehlerreflex;³⁵ auch bezüglich der Rangfolge der anderen beiden möglichen Reflexe stimmen die Befunde mit denen des /θ/-Tests weitgehend überein. Allgemein nimmt die Häufigkeit der [v]-Reflexe jedoch gegenüber der der [z]- und [d]-Reflexe ab, was als Anzeichen einer größeren perzeptorischen Ähnlichkeit der stimmhaften [v], [z] und [d] untereinander gewertet werden muß und gut mit den von Mohr u. Wang ermittelten allgemeinen perzeptorischen Gesetzmäßigkeiten erklärt werden kann.³⁶

Abb. 14 (s. S. 129) zeigt die Anfälligkeit der einzelnen phonetischen Kontexte für perzeptorische Verwechslungen, wobei sich kein eindeutiges Bild für die relativ größere Schwierigkeit eines Testkontextes gegenüber anderen ergibt. Allenfalls für die finale postvokalische Position von /ð/ sind gegenüber der initialen (prä vokalischen) gewisse Hörvorteile zu verzeichnen; da beide Kontexte die gleiche Anzahl von Testwörtern haben, kann der beobachtete Effekt kaum zufälliger Natur sein.

³⁵vgl. die Ergebnisse von Nemser (1971), S. 60-66.- Mohr u. Wang (1968) ermittelten bei ihren "naiven" (englischsprachigen) Hörern eine geringere perzeptorische Distanz (= größere auditive Ähnlichkeit) für das Paar [ð] - [d] als für das Paar [ð] - [v]; es fragt sich, inwieweit dieser Befund auf den sprachlichen Hintergrund dieser Hörer zurückzuführen ist oder aber auf spezifische Einzelheiten der Testdurchführung. So hatten die Vpp in manchen (?) Fällen eine schriftliche Repräsentation der auditiv zu beurteilenden Stimuli vor sich, die ihre Entscheidungen möglicherweise mit beeinflusst hat.

³⁶Mohr u. Wang (1968) ermittelten eine geringere perzeptorische Distanz für "markierte" Mitglieder einer Lautkategorie gegenüber unmarkierten. Bezüglich des (phonologischen) Merkmal "Stimmhaftigkeit" sind [ð], [v], [z] als markiert anzusehen, [θ], [f], [s] als unmarkiert. Die Untersuchungsdaten sind allerdings nicht ohne Widersprüche (vgl. die Distanzwerte für [θ] - [f] mit denen für [ð] - [v], S. 37).

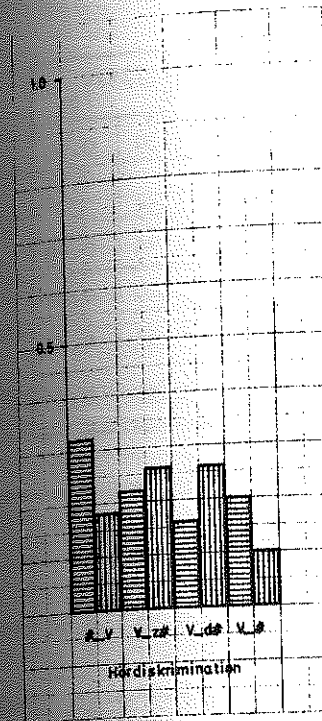


Abb. 14: Hördiskrimination: /ð/ Fehlerraten in den einzelnen Testkontexten ("LR": [horizontal lines]; "M": [vertical lines])

5.1.13. Höridentifikation

Alle der in Test I getesteten phonetischen Kontexte wurden auch in Test II (Höridentifikation) als Testkontexte für /ð/ zur Beurteilung angeboten. Die Ergebnisse sind in den Tabellen 24 a und b für beide Gruppen von Vpp getrennt zusammengestellt; Abb. 15 (s. S. 131) gibt Aufschluß über die Art und relative Häufigkeit der beobachteten Fehlidentifikationen.

Sieht man von den gegenüber Test I signifikant schlechteren Hörleistungen ab, die ihren Grund unmittelbar in dem verwendeten Testformat selber haben, bleibt für die Gruppe "LR" auf den generellen Zusammenfall der Hörerlebnisse für die beteiligten Verwechslungslaute zu verweisen. In einem der vier Testkontexte (V. f) überwiegen gar die [z]-Reflexe gegenüber den sonst üblichen [v]-Verwechslungen, während die Angehörigen der Gruppe "M" das auch in allen anderen Kontexten beobachtete Fehler-

Kontext	TW	FZ	FR	Fehlerreflexe		
				/v/	/z/	/d/
* 1 #_V	144	53	0.37	0.40	0.30	0.30
2 V_z#	32	21	0.66	0.67	-	0.33
3 V_d#	32	21	0.66	0.95	0.05	-
* 4 V_#	112	36	0.32	0.36	0.44	0.20
Σ	320	131	0.41	0.59	0.20	0.21
gemeins. Kontexte	256	89	0.35	0.38	0.37	0.25
gewichtete Durchschn.	-	-	-	0.39	0.36	0.25

Tab. 24a: Fehlerraten und Fehlerreflexe: Höridentifikation (/ð/); Gruppe "LR" (8 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe		
				/v/	/z/	/d/
* 1	360	89	0.25	0.78	0.12	0.10
2	80	34	0.43	0.76	-	0.24
3	80	37	0.46	1.00	0.00	-
* 4	280	49	0.18	0.74	0.16	0.10
Σ	800	209	0.26	0.80	0.09	0.11
gemeins. Kontexte	640	138	0.21	0.76	0.14	0.10
gewichtete Durchschn.	-	-	-	0.77	0.13	0.10

Tab. 24b: Fehlerraten und Fehlerreflexe: Höridentifikation (/ð/); Gruppe "M" (20 Vpp)

verhalten an den Tag legen.- Abb. 16 (s. S. 132) zeigt wiederum die unterschiedliche Anfälligkeit der einzelnen Testkontexte für Fehlidentifikationen. Für das Vorkommen von /ð/ in finaler postvokalischer Position scheint sich ähnlich wie auch in Test I ein leichter Perzeptionsvorteil gegenüber der initialen (prävo-kalischen) Position zu ergeben. Deutlich stärker fehlerverursachend sind die restlichen beiden Kontexte, in denen /ð/ Element einer auslautenden Konsonantenverbindung ist. Da /ð/ in diesen beiden Positionen gegenüber den beiden anderen nur

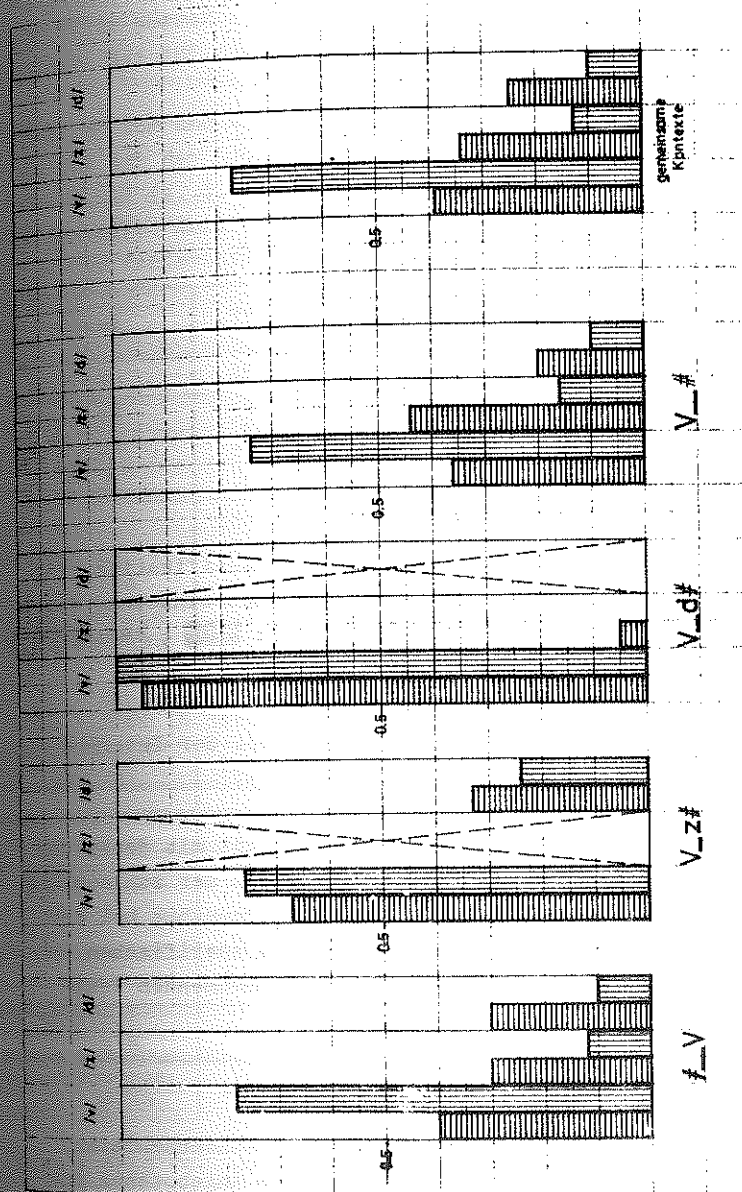


Abb. 15: Höridentifikation: /ð/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten ("LR": ; "M":)

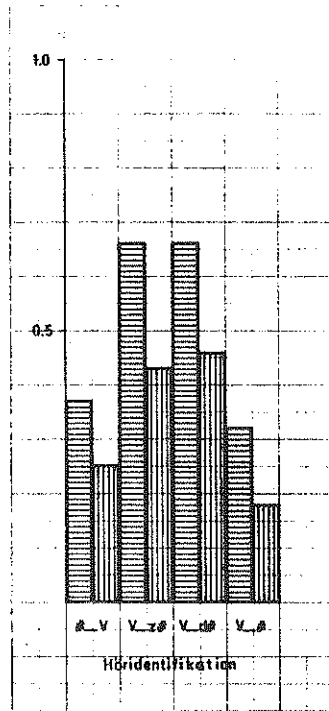




Abb. 16: Höridentifikation: /ð/ Fehlerraten in den einzelnen Testkontexten ("LR":  ; "M": )

selten zur Beurteilung angeboten wurde, ist bei der Interpretation dieses Befundes Vorsicht geboten.

Im Ganzen stimmen die Ergebnisse von Test II gut mit denen von Test I überein, zumindest was das Testverhalten der zahlenmäßig größeren Gruppe "M" anbelangt. Die hohe Fehlerzunahme bei der Gruppe "LR", die auch größere Schwankungen im spezifischen Fehlerverhalten zeigt, kann weitgehend der zunehmenden Testmüdigkeit dieser Gruppe zugeschrieben werden.

5.1.14. Perzeption - Produktion: Nonsensewörter

Auch Test III (Perzeption - Produktion: Nonsensewörter) enthielt sämtliche in beiden Perzeptionstests angebotenen phonetischen Kontexte für /ð/ und erweist sich ebenso wie bei /ð/

als eine schwierige Aufgabe, die bei der Gruppe "LR" nur in 19% der Fälle korrekte Realisationen des kritischen Lautes hervorbringt, bei "M" (immerhin) 51%.

Tabellen 25a und b verzeichnen die relativen Fehlerhäufigkeiten und die Art der Lautsubstitutionen in den vier Testkontexten; in Abb. 17 (s. S. 134) sind dieselben Fehlerdaten übersichtlich dargestellt. Ähnlich wie bei /θ/ erweisen sich diejenigen Kontexte am fehleranfälligsten, in denen /ð/ Element einer Konso-

Kontext	TW	FZ	FR	Fehlerreflexe			
				/v/	/z/	/d/	and.
*1 #_v	48	34	0.71	0.79	0.00	0.06	0.15
2 v_z#	8	8	1.00	(0.25)	0.00	(0.25)	0.50
3 v_d#	16	16	1.00	0.94	0.00	0.00	(0.06)
*4 v_#	31	28	0.90	0.68	(0.04)	0.07	0.21
Σ	103	86	0.83	0.73	(0.01)	0.07	0.19
gemeins. Kontexte	79	62	0.78	0.74	0.02	0.06	0.18
gewichtete Durchschn.	-	-	-	0.74	0.02	0.06	0.18

Tab. 25a: Fehlerraten und Fehlerreflexe: Perz - Prod (Nonsensewörter): /ð/; Gruppe "LR" (4 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe			
				/v/	/z/	/d/	and.
*1	96	26	0.27	0.69	0.08	0.15	0.08
2	16	16	1.00	0.31	0.00	0.56	0.13
3	32	29	0.91	1.00	0.00	0.00	0.00
*4	64	45	0.70	0.93	0.00	(0.02)	(0.05)
Σ	208	116	0.56	0.81	0.02	0.12	0.05
gemeins. Kontexte	160	71	0.44	0.81	0.04	0.09	0.06
gewichtete Durchschn.	-	-	-	0.85	0.03	0.07	0.05

Tab. 25b: Fehlerraten und Fehlerreflexe: Perz - Prod (Nonsensewörter): /ð/; Gruppe "M" (8 Vpp)

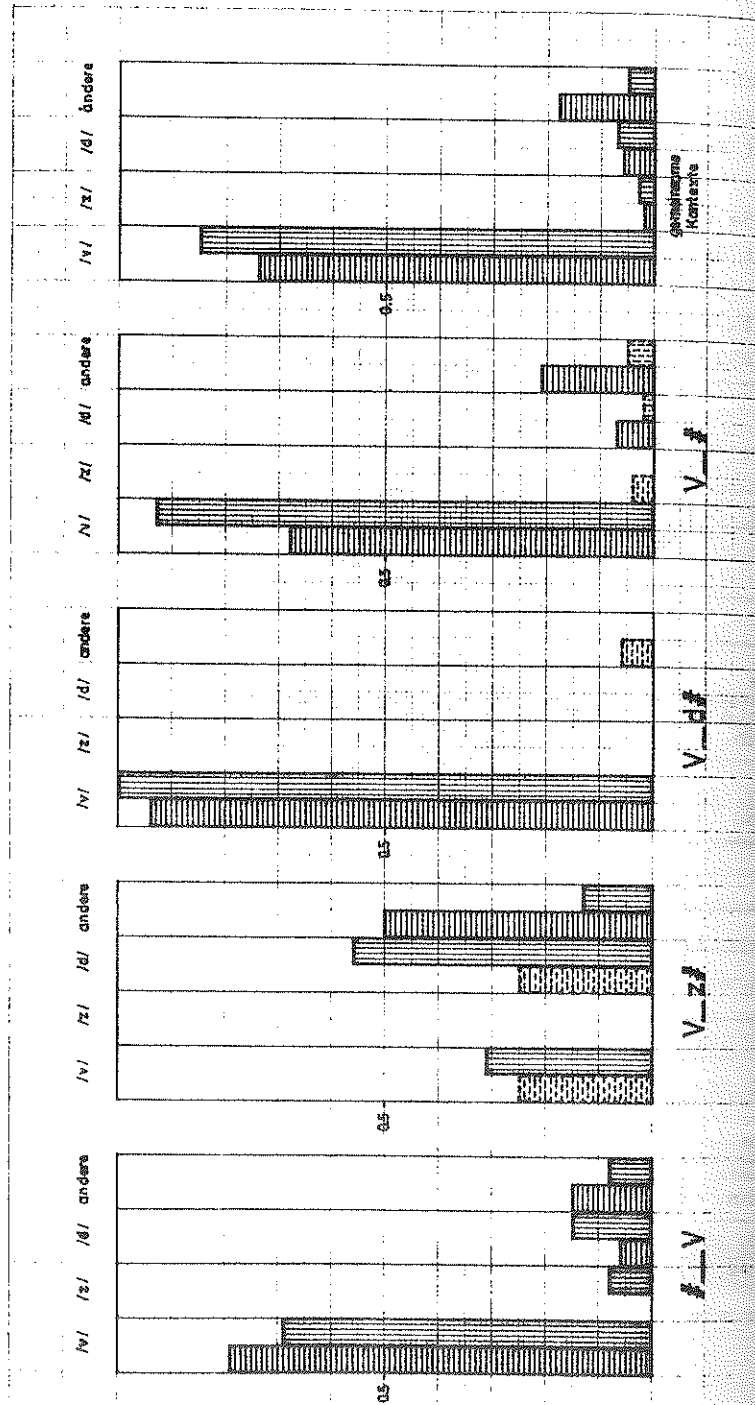


Abb. 17: Perzeption - Produktion (Nonsensewörter): /ð/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten (WB/LR: WB/LR:)

nantenverbindung ist. - Eine relativ geringe Anzahl von Fehlern, die Tabelle 26 verzeichnet, kann keiner der vorgegebenen Fehlerkategorien zugewiesen werden; in der Mehrzahl finden sie sich bei den Angehörigen der Gruppe "LR". In der initialen Position wurde nur einmal gegen die geforderte Stimmbeteiligungsart verstoßen, die entsprechende Realisation aber dennoch als korrekt eingestuft.

Kontext	Fehler	Häufigkeit
1 #_V	[j] ³⁷ [v] ³⁷ [w] [z]	je 2
2 V_z#	[bz] [ts] [θ]	4 je 1
3 V_d#	[fst]	1
4 V_f	[vø] [g] [vq] [z]	je 2 je 1

Tab. 26: Idiosynkratische Produktionsfehler
Perz - Prod (Nonsensewörter): /ð/; Fehlerdaten beider Gruppen

Abb. 18 schließlich (s. S. 136) stellt die Fehleranfälligkeit der einzelnen Kontexte für /ð/ in diesem Test dar. Die gegenüber allen anderen Kontexten deutlich besseren Leistungen für den Kontext #_V sind zweifellos darauf zurückzuführen, daß die Vpp mit dieser Position des Testlauts aufgrund des häufigen Vorkommens der mit /ð/ beginnenden Formwörter im Englischen besonders vertraut waren.

Bezüglich der Fehlertypik beobachten wir in diesem Test Erscheinungen, die dem Befund des /θ/-Tests in nahezu jeder Beziehung entsprechen. In beiden Perzeptions- wie auch in dem mit ähnlichem Testmaterial durchgeführten ersten Produktions-test wird am häufigsten der dem kritischen Laut [ð] auditiv ähnlichste Laut [v] substituiert, während [z] und [d] - die im Unterricht am häufigsten beobachteten Substitutionen - nur

³⁷ Alle Fehlrealisationen bei einem Sprecher (WB/LR).

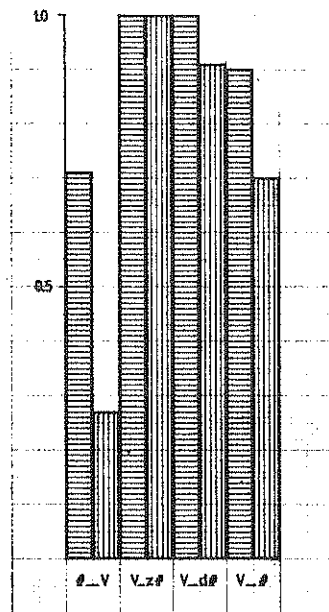




Abb. 18: Perz - Prod (Nonsensewörter): /ð/
Fehlraten in den einzelnen Testkontexten
("LR": ; "M": )

eine untergeordnete Rolle spielen. Die beobachteten Fehler sind dabei über die gesamte Population von 12 Vpp gestreut, obwohl selbstverständlich individuelle Leistungsunterschiede zu verzeichnen sind.

5.1.15 Perzeption-Produktion: englische Einzelwörter

Der zweite Produktionstest - der erste mit "echtem" englischen Wortmaterial - beseitigte, wie wir anhand der Fehlerdaten für /θ/ gesehen haben, schlagartig eine Reihe von Schwierigkeiten, die unmittelbar mit der Art des in Test I - III verwendeten Testmaterials gegeben sind.

Der gleiche Effekt ist auch für /ð/, wenngleich nicht in

derselben Deutlichkeit, zu beobachten. Tabelle 27 verzeichnet die Fehlerraten der um den Kontext V_d# verringerten Testkontexte; als neuer Kontext hinzugefügt wurde V_V, da für diesen Kontext eine relativ große Zahl von Testwörtern zur Verfügung stand.

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_V	16	6	0.38	32	4	0.13
2 V_z#	4	4	1.00	8	5	0.63
3 V_d#	4	-	0.00	8	-	0.00
4 V_V	12	-	0.00	24	-	0.00
Σ	36	10	0.28	72	9	0.13

Tab. 27: Perzeption-Produktion (engl. Einzelwörter):
787; Fehlerraten in den einzelnen Testkontexten

Tabelle 27 zeigt, daß fehlerhafte Realisationen von /ð/ nur in den ersten beiden Testkontexten beobachtet wurden, während in den anderen Positionen der Laut korrekt wiedergegeben wurde. Die Produktionsfehler selbst sind in Tabelle 28 verzeichnet.

Kontext	Fehler	Häufigkeit
1 #_V	[w], [d]	je 3
	[v]	2
	[β'], [s]	je 1
2 V_z#	[dθ]	3
	[d]	3
	[v], [gθ], [s]	je 1

Tab. 28: Produktionsfehler: Test III (/ð/);
Fehlerdaten beider Gruppen

Sie ergeben kein einheitliches Bild und erlauben vor allem nicht, von bevorzugten Substitutionen in einem oder gar beiden Testkontexten zu sprechen.

Identifikationsmängel mögen zu dem uneindeutigen Fehlerbild beigetragen haben. Ein solcher Schluß legt sich nahe, wenn

man Schülerrealisationen wie [dei] für [ðei] (dreimal) oder [weə] für [ðeə] (ebenfalls dreimal) betrachtet. Es muß wohl auch berücksichtigt werden, daß die #ð-Wörter dieses Tests in ihrer Eigenschaft als Formwörter nicht im gleichen Maße semantisierbar sind wie die ansonsten ausnahmslos lexikalischen Testwörter und somit eine Sonderstellung einnehmen, die eine korrekte Identifikation und spontane Nachahmung erschweren.³⁸

Die zahlreichen fehlerhaften Imitationen des Testwortes clothes schließlich müssen wohl auf mangelnde Vertrautheit der Schüler(innen) mit diesem Wort erklärt werden; solche Effekte lassen sich bei der Zusammenstellung von Testmaterial nur schwer kalkulieren.

5.1.16. Produktion-Perzeption-Produktion: englische Sätze

Die Fehlerdaten des letzten Produktionstests reproduzieren bezüglich der Häufigkeit und Art der Fehlrealisationen von /ð/ in den beiden beibehaltenen Testkontexten nahezu exakt den für /θ/ ermittelten Befund. Das Fehlerbild ist für die Erstrealisationen (beim Lesen der auf Karten in orthographischer Form vorliegenden Sätze) das gleiche wie für die Zweitrealisationen (nach Abhören der Modellrealisation); allerdings ist eine leichte Abnahme der Fehler in den zweiten gegenüber den ersten Realisationen festzustellen. Tabelle 29 gibt Aufschluß über die relative Fehlerhäufigkeit in den beiden Kontexten (nur die Zweitrealisationen wurden berücksichtigt).

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_V	43	12	0.28	85	3	0.04
2 V_V	16	5	0.31	32	4	0.12
Σ	59	17	0.29	117	7	0.06

Tab. 29: Produktion-Perzeption-Produktion (engl. Sätze): /ð/; Fehlerraten in den einzelnen Kontexten

³⁸ Die Gefahr von Identifikationsfehlern hätte vermieden werden können, indem man die #ð-Wörter eindeutigen Kontexten zugeordnet, also Stimuli wie that boy oder this house gebildet hätte anstelle der hier verwendeten that, this usw.

97% aller Fehlrealisationen sind Substitutionen von [z] und [ʒ]-artigen (meist dentalisierten) Lauten; einmal ist [d] für [ð] substituiert worden und zweimal - nach dem übereinstimmenden Urteil aller drei Beurteiler - [v].³⁹ Besondere Beachtung verdient die Tatsache, daß von den insgesamt zwölf Vpp, die sich diesem Test unterzogen, überhaupt nur vier - drei Angehörige von "LR", einer von "M" - die erwähnten Substitutionen vornehmen. Eine dieser Vpp produziert überhaupt nur drei korrekte /ð/, wobei es sich nicht um Formen des Artikels handelt, sondern um satzeinleitendes (im ersten Falle auch noch betontes) that bzw. they.⁴⁰

5.1.17. Beurteilung der /ð/-Realisationen nach Akzeptabilität

Ein Blick auf die Urteile über die Produktionsleistungen der 12 Schüler(innen) in allen drei Produktionstests (s.u., Abb. 19, S. 140) zeigt, daß die große Mehrzahl der Urteile über die Schülerrealisationen von /ð/ kategorischer Natur waren und nur eine geringe Anzahl von Urteilen (des zweiten und ggf. auch des dritten Beobachters) auf die Kategorie "(noch) akzeptabel" entfielen. Die Leistungsdiagramme lassen auch hier die für beide Gruppen gegenüber Test III deutlich verbesserten Produktionsleistungen in Test IV und V erkennen; die wichtige Information jedoch, die sie vorenthalten, ist die über das Testverhalten einzelner Individuen.

5.1.18. Vergleich der Perzeptions- und Produktionsleistungen (/ð/)

Abbildung 20 stellt die Hör- und Produktionsleistungen der beiden Gruppen für den allen fünf Tests gemeinsamen Kontext #, der wegen seines häufigen Vorkommens zweifellos der wichtigste

³⁹ Diese Substitution wurde bei der gleichen Vp (SM/M) beobachtet, die in dem gleichen Test für [θ] nach dem ebenfalls übereinstimmenden Urteil aller drei Beobachter einmal [f] substituierte.

⁴⁰ Es handelt sich bei dieser Vp um einen Angehörigen der Gruppe "LR" (WB), der von den vier zu Produktionstests herangezogenen Angehörigen dieser Gruppe die besten Perzeptionsleistungen (auch für /θ,ð/) vorzuweisen hat (siehe auch oben, Tab. 5, S. 43).

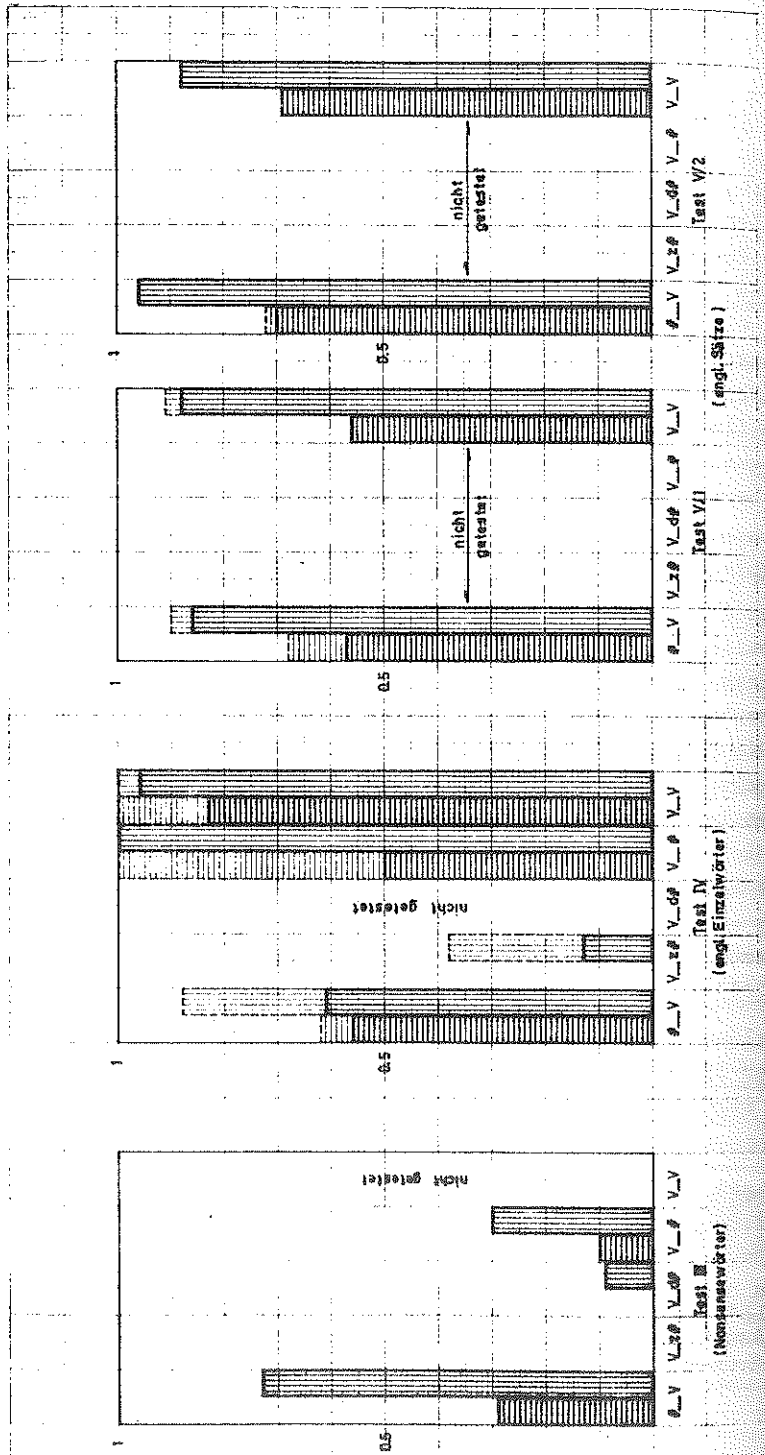


Abb. 19: Akzeptabilitätsbeurteilung der Produktionsleistungen: /ð/ ("LR":  ; "M": )

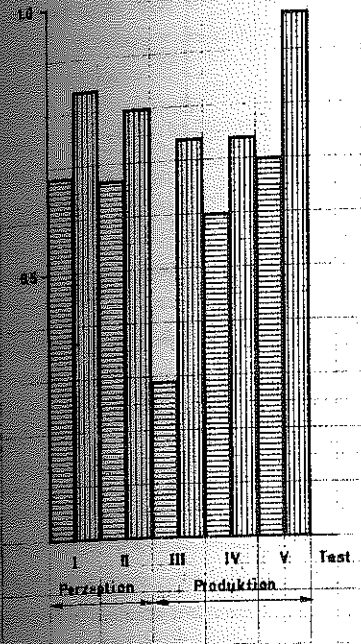




Abb. 20: Hör- und Produktionsleistungen: /ð/ ("LR":  ; "M": )

Testkontext für den hier untersuchten Laut ist, dar. Den individuell errechneten und über die Gruppe gemittelten Leistungsdiagrammen liegen zugrunde:

Test I:	216	(Hörer-)Urteile
Test II:	216	(Hörer-)Urteile
Test III:	72	Realisationspaare
Test IV:	48	Einzelrealisationen
Test V:	128	(Zweit-)Realisationen

Tabelle 30 stellt die individuellen Leistungswerte der zwölf Typ in jedem einzelnen der fünf Perzeptions- und Produktionstests zusammen. Hinzuweisen ist besonders auf das bereits erwähnte Verhalten von WB ("LR") sowie auf die Leistungen von SA ("LR"), der

Gruppe	Test	I	II	III	IV	V
	Vp					
"LR"	SA ₁	0.56	0.58	0.58	1.00	0.77
	SS ₂	0.72	0.63	0.25	0.13	0.77
	SS ₃	0.67	0.74	0.25	1.00	1.00
	WB	0.78	0.74	0.08	0.25	0.27
"M"	BM	0.94	0.74	0.92	0.80	1.00
	DL	0.89	0.89	0.50	0.75	1.00
	DM	0.67	0.63	0.88	0.75	0.73
	HH	0.78	0.74	0.66	0.63	1.00
	HU	0.89	0.84	0.50	0.63	1.00
	JK	0.89	0.79	0.67	0.88	1.00
	RR	0.83	0.95	1.00	0.75	1.00
	SM	0.83	0.79	0.83	0.75	1.00
MW		0.79	0.76	0.59	0.69	0.88
StA		0.11	0.11	0.29	0.26	0.22

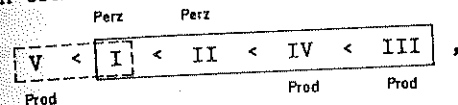
Tab. 30: Individuelle Perzeptions- und Produktionsleistungen: /ð/

in den beiden Perzeptionstests eindeutig schlechter als alle übrigen Gruppenmitglieder abschneidet, in allen drei Produktionstests hingegen in seiner Gruppe die besten Leistungen erbringt. Ein solches Verhalten läßt darauf schließen, daß für einen Problemlaut wie /ð/ (und auch für seine stimmlose Entsprechung /θ/, wo sich ein ähnliches Verhalten zeigt) mit einem einfachen Zusammenhang zwischen Perzeption und Produktion nicht gerechnet werden kann und die Perzeptionsleistungen - ob gut oder schlecht - keinerlei Voraussagen über die Produktionsleistungen erlauben.

Die wenigstens sporadisch korrekten Realisationen der entsprechenden Problemlaute auch von Seiten derjenigen Probanden die insgesamt häufig fehlerhaftes Verhalten an den Tag legen, weisen darauf hin, daß von einer mangelhaften Vertrautheit mit den artikulatorischen Bewegungen, die zur Hervorbringung des Lautes notwendig sind, keine Rede sein kann: Diese Lerner wissen, wie der entsprechende Laut hervorzubringen ist.

und wissen auch um den akustischen Effekt, den er produziert. Daß sie von ihrem Wissen keinen Gebrauch machen, muß seinen Grund in individualpsychischen Faktoren haben, die wir vorab nicht kennen.⁴¹

Trotz des mit dem Verhalten der übrigen Probanden nicht in Einklang zu bringenden "Ausreißer"-Verhaltens einzelner Vpp ergibt die Rangvarianzanalyse (FRIEDMAN-Test) einen signifikanten Befund für den unterschiedlichen Einfluß der fünf Einzeltests auf die Leistungen der 12 Probanden ($\chi^2 = 16.12 > \chi^2 = 9.49$); für die Reihenfolge der Schwierigkeiten der einzelnen Aufgaben erhält man nach WILCOXON-WILCOX



wobei sich die Produktionsaufgabe in Test V als signifikant leichter gegenüber beiden anderen Produktionstests wie auch gegenüber der Identifikationsaufgabe erweist.⁴² Die Sonderstellung von Test V, wie sie sich vor allem auch aufgrund der unterschiedlichen Fehlertypik (nahezu ausschließlich [z]-Reflexe anstelle der in beiden Perzeptionstests Überwiegenden [v]-Verwechslungen) ergibt, ist damit auch für /ð/ unter Beweis gestellt. Daß Test IV nicht in gleicher Weise wie bei /θ/ an dieser Sonderstellung teilhat, wird auf die Identifikationschwierigkeiten der Probanden bei den #ð-Wörtern von Test IV zurückgeführt, für die sich auch kein eindeutiges Fehlerbild (wie in dem gleichen Test für /θ/) ergab.

⁴¹Vgl. auch P. Ladefoged, "Units in the Perception and Production of Speech", in *dsb.*, Three Areas of Experimental Phonetics (= Language and Language Learning 15), London 1967, S. 169: "... students learning a foreign language often distinguish between sounds made by the instructor but do not use the distinctions in their replies to him. But this is usually not because they cannot make the distinction, but because they do not remember to make it. The claim that is being made here is that, in general, if people can hear a difference between a pair of similar sounds, then they can make the difference - but do not necessarily do so in their ordinary speech."

⁴²Bei getrennter Durchführung des Tests für beide Gruppen ergibt sich aufgrund des "Ausreißer"-Verhaltens einzelner Mitglieder für "LR" kein signifikanter Effekt; für "M" ist er jedoch in gleichem Maße vorhanden.

5.2.0. R-Test

5.2.1. Perzeption: /r/

Eine Übersicht über die phonetischen Kontexte, in denen /r/ als Testlaut in beiden Hörtests zur Beurteilung angeboten wurde, gibt Tabelle 6 (s.o., S. 52). Mit den unterschiedlichen phonetischen Kontexten ist zugleich eine hinreichend große Variationsbreite bezüglich verschiedener allophonischer Realisationen gegeben.⁴³ Auf die besondere Funktion des Kontextes V_V muß in diesem Zusammenhang hingewiesen werden: In beiden Hörtests wie auch in dem ersten der drei Produktionstests enthielt die hier zu beurteilende bzw. nachzuahmende Modellrealisation unserer Informanten das sog. "flapped r" ([ɾ]), niemals jedoch das in dieser Position ebenfalls mögliche Allophon [ɹ], das in allen anderen Kontexten (allerdings mit unterschiedlichen Stimmbeteiligungseigenschaften) realisiert wurde.⁴⁴ Auf die spezifischen Gründe für diese Realisationsart wird weiter unten eingegangen.

5.2.2. Fehlervorgabe

An den Prinzipien der Fehlervorgabe, von denen im TH-Test ausgegangen wurde, konnte bei der Vorgabe möglicher Fehlerreflexe für /f/ nicht festgehalten werden. Zwar ist gegen die Verwendung im Unterricht beobachteter Lautsubstitutionen (etwa eigensprachlicher - deutscher - r-Realisationen anstatt der korrekten englischen) in unseren Perzeptionstests grundsätzlich nichts einzuwenden, zumal die Perzeptionsleistungen unserer beiden Gruppen von Vpp mithilfe von Nonsensematerial untersucht wurden. Die spezifisch "englische" Strukturierung dieses Materials jedoch, was sowohl die lautlich-segmentellen wie auch die phonotaktischen Möglichkeiten betrifft, verbot eine Einbeziehung solcher lautlicher Erscheinungen, die im Englischen - zumindest in der hier dargebotenen Realisationsnorm - nicht vorkommen. Die gesamte Anlage der Hörtests und die Integration vier unterschiedlicher Ausspracheprobleme des Englischen für Deutsche in ein- und denselben

⁴³Einen Überblick gibt Gimson (1970), S. 205-210.

⁴⁴Eine Ausnahme bildet die Folge #or__, die ebenfalls das "flap"-Allophon enthält.

Test schrieb eine solche "englische" Strukturierung des Materials zwingend vor; die Notwendigkeit ergibt sich insbesondere dort, wo das Perzeptions- und Produktionsvermögen für spezifische Vokalphoneme des Englischen sowie für die ebenfalls spezifisch englischen "final voiced consonants" einschließlich aller hier beteiligten allophonischen Erscheinungen untersucht wird.

Auch ein Ausgehen von solchen fehlerhaften Substitutionen für /r/, wie sie sich in der Literatur zum Spracherwerb (englischer Kinder) belegen lassen, erweist sich als schwierig: Die in den meisten Fällen phonologisch bestimmte Fragestellung in der Literatur gibt häufig wenig Aufschluß über phonetische Fehler.⁴⁵ Die chronologischen Angaben über den sog. "Phonem"-Erwerb bei englischsprachigen Kindern deuten außerdem darauf hin, daß /r/ erst relativ spät erworben wird⁴⁶ - genauer gesagt zu einem Zeitpunkt, der außerhalb des Untersuchungsbereichs der meisten Studien liegt.

Bei der Fehlervorgabe für /r/ für die beiden Perzeptionstests wurde schließlich davon ausgegangen, diejenigen Laute einzusetzen, die in ihren allgemeinen Verteilungseigenschaften die größte Ähnlichkeit mit /r/ besitzen, also /l/, /w/ und /j/.⁴⁷ Mag eine solche Fehlervorgabe nicht völlig frei von Willkür sein, ist sie andererseits durch die spezifische Strukturierung

⁴⁵Siehe jedoch J. Kresheck, H. Fisher, D. Rutherford, "A Study of r-Phones in the Speech of Three-Year-Old Children", *Folia Phoniatrica*, 24 (1972), 301-312. 684 Einzelwortäußerungen englischer Kinder, in denen /r/ in verschiedenen phonetischen Kontexten auftrat, wurden von vier phonetisch geschulten Beobachtern eng transkribiert. Als häufigste Substitution für /r/ in sämtlichen getesteten Positionen mit Ausnahme der postvokalischen wurde /w/ verzeichnet.

⁴⁶Vgl. Olmsted (1971), S. 139, 204. Siehe auch P. Menyuk, *The Acquisition and Development of Language*, Englewood Cliffs 1971, bes. S. 34-91 ("The Acquisition and Development of Phonology").

⁴⁷Vgl. J.D. O'Connor u. J.L.M. Trim, "Vowel, Consonant and Syllable - A Phonological Approach", *Word*, 9 (1953), 103-122, bes. 116.

unseres Materials und die Notwendigkeit, die Diskriminations- und Identifikationsfähigkeit deutscher Englischlernender für den "Problemlaut" /r/ in den wichtigsten phonetischen Kontexten zu untersuchen, nachgerade geboten.⁴⁸

Für [r] in intervokalischer Position konnten /w/ und /j/ als mögliche Fehlerreflexe nicht verwendet werden; hier wurde neben /l/ als möglicher Verwechslungslaut /d/ eingesetzt.⁴⁹

5.2.3. Hördiskrimination

Die Fehlerdaten aus Test I (Hördiskrimination) sind in Tabelle 31a und b für beide Gruppen von Vpp zusammengestellt; Abbildung 21 stellt sie in Form von Flächendiagrammen dar, die Aufschluß über die Art und Häufigkeit der in den einzelnen phonetischen Kontexten beobachteten perzeptorischen Verwechslungen geben.

Von beiden Gruppen von Vpp wird der "kritische" Laut /r/ besser diskriminiert als /θ/ und /ð/: Auf die Gesamtzahl der zur Beurteilung angebotenen Teststimuli bezogen, wird /r/ in 79% ("LR") bzw. 92% ("M") der Fälle erfolgreich diskriminiert; für das getrennt behandelte zwischenvokalisches [r] ergeben sich Erfolgsquoten von 80 bzw. 79%.⁵⁰

Für die einzelnen Kontexte ergibt sich dabei, wie Abbildung 22 (s.u., S. 149) zeigt, eine unterschiedliche Fehleranfälligkeit: Während /r/ in initialer (prä vokalischer) Position ausgezeichnet definiert zu sein scheint und in 89 ("LR") bzw. sogar

⁴⁸So sind im Englischen in der Position #CCC nur /r, l, w, j/ phonotaktisch möglich, wobei /j/ zusätzlichen Beschränkungen bezüglich des Folgevokals unterliegt.

⁴⁹Diese Wahl ist nicht willkürlich getroffen: Ein dem [r] sehr ähnliches kurz geschlagenes [d] ist in der norddeutschen Aussprache häufig als Allophon des /t/ anzutreffen (z.B. [mødə] für Mutter, [hadə] für hatte); hier ergibt sich die Möglichkeit, die "phonologischen Perzeptionen" unserer (norddeutschen) Probanden zu testen.

⁵⁰Kontrollgruppe: /r/: 2% Fehler; [r]: 2%.

Kontext	TW	FZ	FR	Fehlerreflexe		
				/l/	/w/	/j/
1 #_V	168	18	0.11	0.22	0.56	0.22
2 #C_V	120	30	0.25	0.10	0.77	0.13
3 #CC_V	48	13	0.27	0.54	0.31	0.15
Σ	336	61	0.18	0.23	0.61	0.16
				/l/	/d/	
4 V_V	128	25	0.20	0.20	0.80	

Tab. 31a: Fehlerraten und Fehlerreflexe: Hördiskrimination (/r/); Gruppe "LR" (8 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe		
				/l/	/w/	/j/
1 #_V	420	6	0.01	0.00	0.50	0.50
2 #C_V	300	45	0.15	0.11	0.73	0.16
3 #CC_V	120	16	0.13	0.81	0.06	0.13
Σ	840	67	0.08	0.27	0.55	0.18
				/l/	/d/	
4 V_V	320	66	0.21	0.53	0.47	

Tab. 31b: Fehlerraten und Fehlerreflexe: Hördiskrimination (/r/); Gruppe "M" (20 Vpp)

99% ("M") der Fälle erfolgreich diskriminiert wird, ergeben sich für die anderen Testkontexte z.T. deutlich höhere Fehlerraten.

Kein einheitliches Bild ergibt sich von der Art der in den Gehörseindrücken der Probanden aufgetretenen Lautverwechslungen; vor allem kann keine der drei vorgegebenen Fehlermöglichkeiten völlig von der Hand gewiesen werden.⁵¹ Aller-

⁵¹Vgl. hiermit die Ergebnisse von P. Menyuk u. S. Anderson, "Children's Identification and Reproduction of /w/, /r/ and /l/", JSHR, 12 (1969), 39-52. Die Autorinnen fanden heraus, daß weder Kinder verschiedener Altersgruppen noch Erwachsene bei der ihnen gestellten Imitations- bzw.

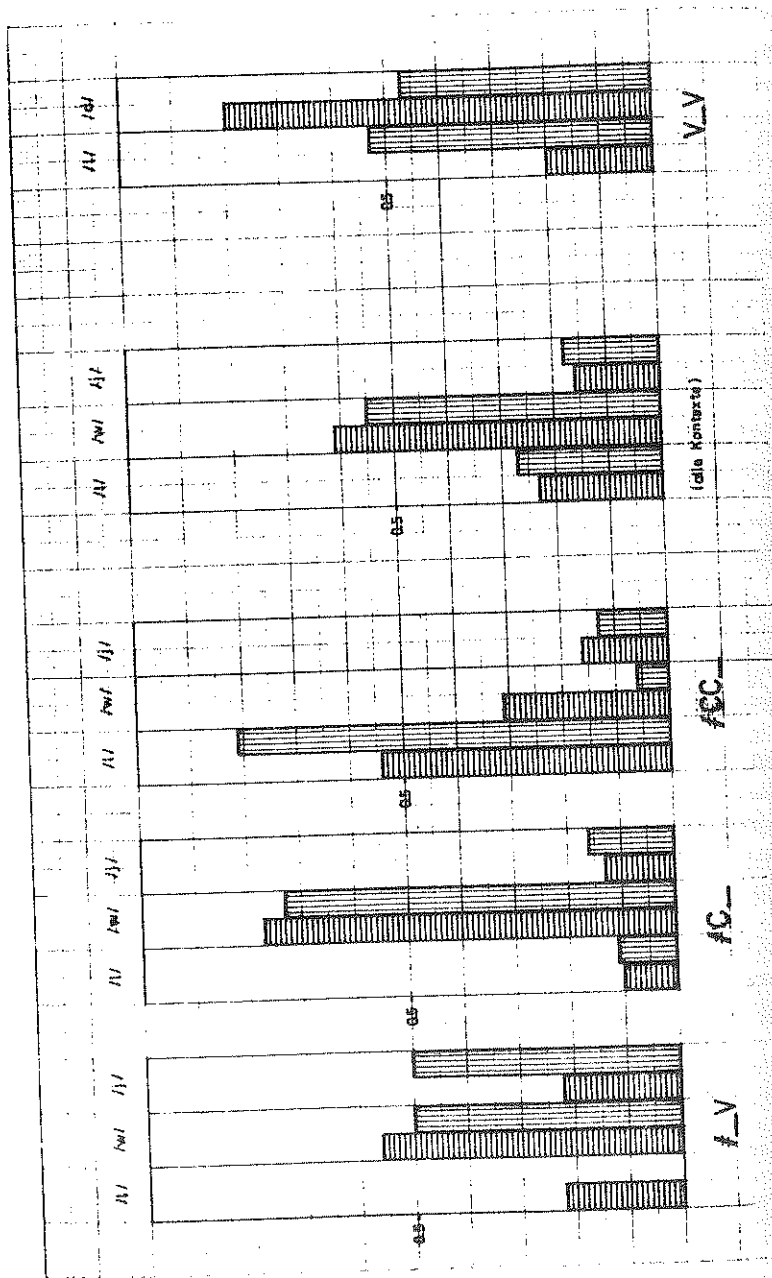


Abb. 21: Hördiskrimination: /r/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten ("LR": ■; "M": ▨)

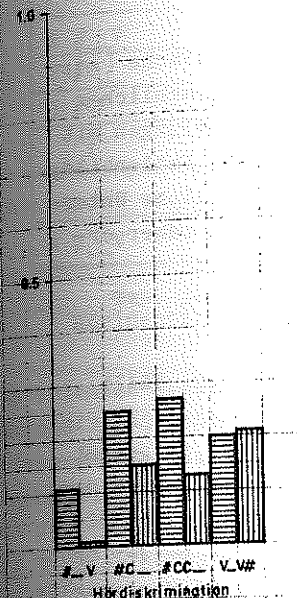


Abb. 22: Hördiskrimination: /r/; Fehlerraten in den einzelnen Testkontexten ("LR": ■; "M": ▨)

dings zeigen die individuellen Hörleistungen aller 28 Vpp, daß nur bei den /w/-Substitutionen für /r/ - und hier vor allem in der Umgebung #C_ - die Fehler ziemlich gleichmäßig gestreut sind: Nur eine einzige Vp hat diese Verwechslung niemals vorgenommen. Dies gilt nicht von den fehlerhaft substituierten /l/- und noch weniger von den /j/-Reflexen, obwohl

Identifikationsaufgabe in der Lage waren, synthetische CVC-Silben mit systematisch variierten Formantstrukturen (von light zu white, light zu w.ite und white zu write) kategorisch den unterschiedlichen Phonembereichen von /w/, /r/ und /l/ zuzuweisen. Als produktiver Reflex war bei den Kindern /w/ signifikant häufiger zu verzeichnen als /r/ und /l/; /w/ wurde allerdings in der Identifikationsaufgabe nicht häufiger identifiziert als die beiden anderen Laute.

sich die ersteren immerhin für 19 und die letzteren für 11 der 28 Vpp nachweisen lassen.

Auch für das gesondert zu behandelnde intervokalische [r], von dem erwartet wurde, daß es aufgrund eigensprachlicher phonologischer Gegebenheiten allenfalls Verwechslungen mit /d/ anheimfallen würde, ergibt sich kein klares Fehlerbild; die beiden Gruppen weisen hier zudem recht deutliche Unterschiede in ihrem Fehlerverhalten auf, die nicht näher erklärt werden können.

5.2.4. Höridentifikation

Das Ergebnis von Test II (Höridentifikation) fällt für /r/ um so unbefriedigender aus, als es - anders als im entsprechenden TH-Test - noch nicht einmal die insgesamt wenig aufschlußreichen Befunde von Test I erhärten und dort beobachtete Fehlertendenzen bestätigen kann. Tabellen 32a und b vereinigen die entsprechenden Fehlerdaten beider Gruppen, die in Abbildung 23 in der vertrauten Weise dargestellt sind. Zu beachten ist, daß sich - ebenfalls unerwartet - die Identifikationsaufgabe für diesen Problemlaut und seine möglichen Verwechslungslaute für die Angehörigen der Gruppe "M"

Kontext	TW	FZ	FR	Fehlerreflexe		
				/l/	/w/	/j/
1 #_V	144	54	0.37	0.20	0.65	0.15
2 #C_	128	31	0.24	0.64	0.26	0.10
3 #CC_	16	6	0.37	0.83	0.00	0.17
Σ	288	91	0.32	0.40	0.47	0.13
				/l/	/d/	
4 V__V	32	18	0.56	0.44	0.56	

Tab. 32a: Fehlerraten und Fehlerreflexe: Höridentifikation (/r/); Gruppe "LR" (8 Vpp)

Kontext	TW	FZ	FR	Fehlerreflexe		
				/l/	/w/	/j/
1 #_V	360	26	0.07	0.27	0.50	0.23
2 #C_	320	16	0.05	0.69	0.31	0.00
3 #CC_	40	1	0.02	1.00	0.00	0.00
Σ	720	43	0.06	0.44	0.42	0.14
				/l/	/d/	
4 V__V	160	25	0.16	0.28	0.72	

Tab. 32b: Fehlerraten und Fehlerreflexe: Höridentifikation (/r/); Gruppe "M" (20 Vpp)

als leichter erweist als die in Test I gestellte Aufgabe, während die Gruppe "LR" den bereits beim TH-Test gezeigten Leistungsabfall zeigt: Die Anwesenheit mehrerer der von uns angebotenen Verwechslungslaute in derselben Testwortreihe muß für die Angehörigen von "M" den besonderen Status von /r/ stärker herausgestellt und weniger Anlaß zu perzeptorischen Verwechslungen gegeben haben, als dies bei den nur eine Distinktion enthaltenden Testwortreihen von Test I der Fall war. Der einzige Effekt, der mit dem in Test I erhobenen Befund übereinstimmt, ist die geringe Streuung der Identifikationsfehler bei der Gruppe "M": 75% der bei den Angehörigen dieser Gruppe zu verzeichnenden Identifikationsfehler gehen auf das Konto von nur 4 Vpp (BT, DM, KS, WF₁). - Bezüglich der Fehleranfälligkeit der einzelnen Testkontexte, wie Abbildung 24 sie darstellt, ist ebenfalls keine Übereinstimmung mit dem Befund von Test I zu verzeichnen.

Die Ergebnisse der Perzeptionstests für /r/ sind demnach insgesamt unbefriedigend und insbesondere, was das nur als eratisch zu bezeichnende Fehlerverhalten der Vpp angeht, wenig aufschlußreich. Die einzige Schlußfolgerung, die sie erlauben, ist die, daß /r/ bei der Mehrzahl der Probanden einen auditiv gut definierten Status hat und insgesamt gesehen besser von den hier vorgegebenen Verwechslungslauten unterschieden ist,

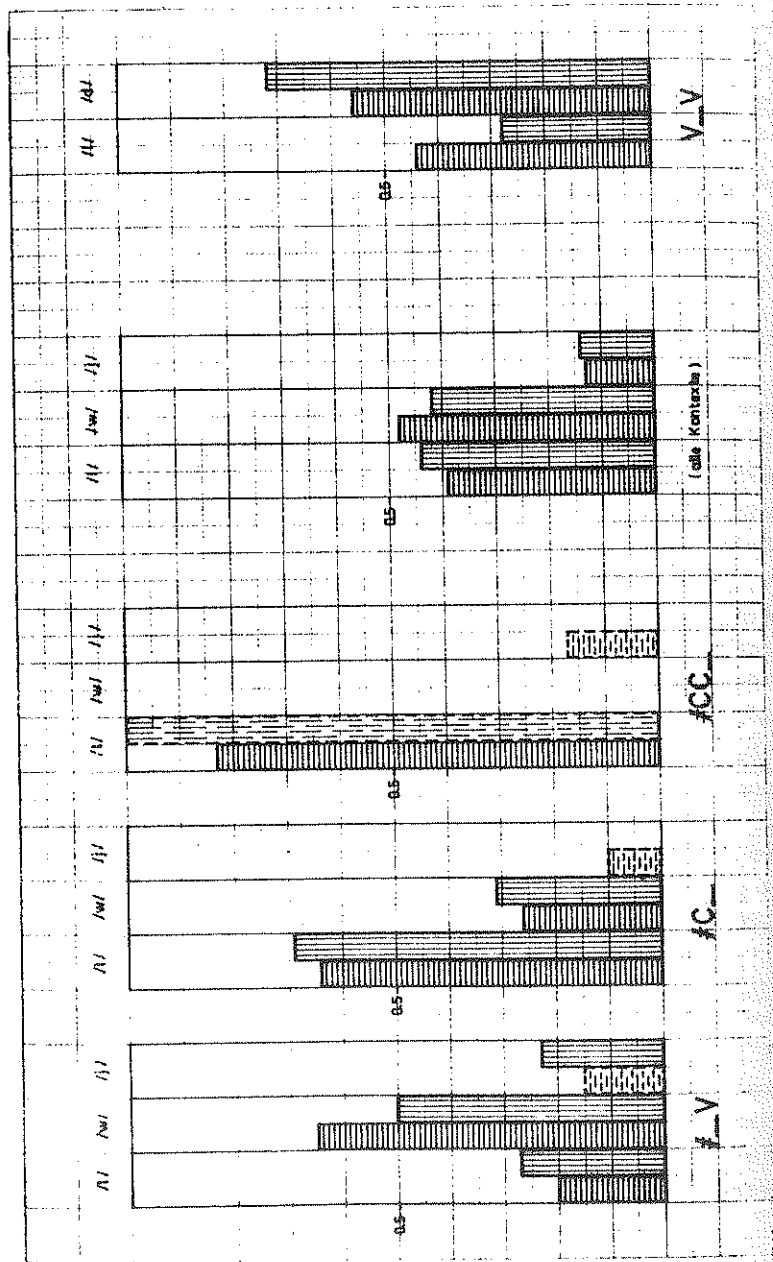


Abb. 23: Höridentifikation: /r/; Fehlerraten und Fehlerreflexe in den einzelnen Testkontexten ("LR": [diagonal lines]; "M": [horizontal lines])

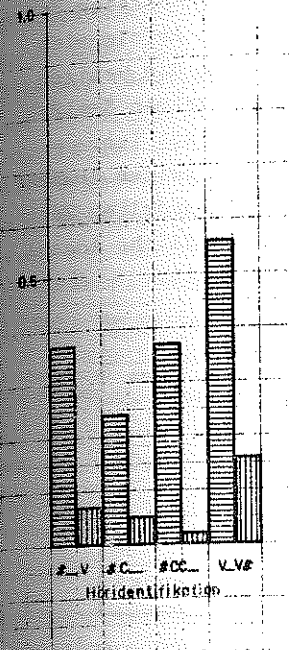


Abb. 24: Höridentifikation: /r/; Fehlerraten in den einzelnen Testkontexten ("LR": [diagonal lines]; "M": [horizontal lines])

als das bei /θ, ð/ und den dort vorgegebenen Konfusionslauten der Fall war. Nur das Verhalten einer relativ kleinen Anzahl von Vpp macht deutlich, daß die hier vorgegebenen Verwechslungsmöglichkeiten nicht völlig implausibel sind und einige Probanden bei der perzeptorischen Trennung von /r/, /l/ und /w/ (in geringerem Maße auch /j/) Schwierigkeiten haben.⁵²

5.2.5. Perzeption-Produktion: Nonsensewörter

Die Ergebnisse von Test III, bei dem die Nachahmung der den Probanden über Kopfhörer vorgespielten Nonsensewörter ge-

⁵² Bei der Kontrollgruppe wurde nur eine einzige Fehlidentifikation von [l] als [r] verzeichnet.

fordert wurde, bestätigen die Befunde beider Hörtests, was den wohldefinierten Status von /r/ in den Hörerlebnissen der Vpp angeht: /r/ wird in nahezu 100% aller Fälle erfolgreich identifiziert und spontan imitiert, wobei zudem die wenigen als fehlerhaft zu bezeichnenden Realisationen keinen Zweifel darüber zulassen, welches der "beabsichtigte" Laut ist. Tabelle 33 verzeichnet die Fehlerraten der Vpp beider Gruppen für die drei Testkontexte und den gesondert behandelten Testkontext V_V, in dem als einzigem eine beträchtliche Anzahl von fehlerhaften Realisationen zu verzeichnen ist.

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #__V	80	9	0.11	160	1	0.01
2 #C__	87	3	0.03	174	2	0.01
3 #CC__	48	6	0.12	96	4	0.04
Σ	215	18	0.08	430	7	0.02
4 V_V	16	9	0.56	32	20	0.63

Tab. 33: Perzeption-Produktion (Nonsensewörter): /r/; Fehlerraten in den einzelnen Testkontexten

Kontext	Fehler	Häufigkeit
1 #__V	[w ^r]	6
	[w]	3
	[ʁ]	1
2 #C__	[r]	2
	[l], [ʁ]	je 1
3 #CC__	[r]	6
	[w ^r]	2
	[ʁ], [ʁ]	je 1
4 V_V	[d]	26
	[t]	2
	[l]	1

Tab. 34: Produktionsfehler: /r/-Substitutionen in den einzelnen Kontexten (Perz -Prod: Nonsensewörter); Fehlerdaten beider Gruppen

Die in den einzelnen Testkontexten registrierten fehlerhaften Lautsubstitutionen für /r/ verzeichnet Tabelle 34 (s.o., S. 154). Die Aufstellung zeigt, daß Substitution eigensprachlicher (deutscher) r-Laute für die geforderten englischen /r/-Allophone nur in einer verschwindend geringen Zahl von Fällen zu verzeichnen ist: Substitution des dorsovelaren Frikativs [ʁ] wird nur dreimal vorgenommen; allenfalls die nach voraufgehendem stimmlosem #[t] bzw. #[st] mehrfach belegten, von unseren Beurteilern als "nicht akzeptabel" bezeichneten [r]-Realisationen können noch als regional-sprachliche deutsche r-Realisationen aufgefaßt werden. Alle anderen Produktionsfehler (mit Ausnahme der beiden [l]-Realisationen und der [d]-Realisationen im Kontext V_V) sind [w]-artige Laute mit starker Labialisierung und zusätzlicher velarer Artikulation, die auditiv dem englischen [ɹ]-Allophon nicht unähnlich sind; sie sind als Substitutionen für noch nicht erworbenes /r/ in der Kindersprache (englischsprachiger Kinder) reich belegt.⁵³

Die Perzeptionsdaten für den gesondert behandelten Kontext V_V haben kein eindeutiges Fehlerbild ergeben, so daß nicht davon ausgegangen werden kann, daß für die hohe Zahl von [d]-Realisationen in diesem Kontext allein die phonologischen Perzeptionen der Probanden verantwortlich sind, die [r] als allophonische Realisation eines Verschlusslautphonems interpretieren, wie es im norddeutschen Sprachgebiet häufig auftritt. Immerhin zeigen vereinzelte Imitationen des [r] als [l], daß die richtige phonologische Zuordnung vorgenommen worden ist.

5.2.6. Perzeption-Produktion: englische Einzelwörter

Die Ergebnisse des mit englischen Einzelwörtern durchgeführten Tests IV, die in Tabelle 35 für beide Gruppen von

⁵³Siehe Gimson (1970), S. 208; vgl. auch Smith (1973), Appendix C, S. 241-243; Kresheck et. al. (1972), Menyuk u. Anderson (1969).

Vpp zusammengestellt sind, ergeben keinen neuen Befund und weisen eine hohe Übereinstimmung mit den Ergebnissen des ersten Produktionstests auf: Der Übergang von Nonsense-Material zu "echtem" englischen Material produziert hier keinen besonderen Effekt, wie er bei /θ, ð/ zu verzeichnen war. Die wenigen Produktionsfehler in den drei Testkontexten

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_V	39	5	0.13	79	1	0.01
2 #C__	40	-	0.00	80	-	0.00
3 V__V	12	4	0.33	24	3	0.12
Σ	91	9	0.10	183	4	0.02

Tab. 35: Perzeption-Produktion (engl. Einzelwörter): /r/; Fehlerraten in den einzelnen Testkontexten

(in der Umgebung #CC__ wurde /r/ nicht mehr zur Nachahmung angeboten) führt Tabelle 36 auf. Der Kontext V__V verlangt bei diesem Test keine Sonderbehandlung mehr, weil die Modellrealisation des Informanten in diesem wie auch im folgenden Test nicht [r] enthielt, sondern das "normale" Allophon [ɹ], so daß mögliche Fehlidentifikationen ausgeschlossen werden konnten.

Kontext	Fehler	Häufigkeit
1 #_V	[w]	3
	[bɹ]	2
	[w]	1
2 #C__	-	-
3 V__V	[ʊ] [v] [ɣ]	je 2
	Ø 54	1

Tab. 36: Produktionsfehler: /r/-Substitutionen in den einzelnen Kontexten (Perz.-Prod. engl. Einzelwörter); Fehlerdaten beider Gruppen

In einigen Fällen können die zu verzeichnenden Produktionsfehler Identifikationsmängeln oder -fehlern seitens der Vpp zugeschrieben werden - so etwa, wenn Realisationen wie [bʊɔŋ] oder [bʊəɔz] die Vermutung nahelegen, daß die Stimuluswörter ([ɹɔŋ] bzw. [ɹəɔz]) nicht identifiziert oder (bei Imitation des Modellwortes [ɹʌŋ] als [wʌŋ]) falsch identifiziert wurden.

Die völlige Fehlerfreiheit des Kontextes #C__ sollte angesichts der ohnehin geringen Fehlerhaftigkeit der /r/-Realisationen in allen Kontexten nicht überinterpretiert werden.

5.2.7. Produktion-Perzeption-Produktion: englische Sätze

Auch für die /r/-Realisationen in unserem letzten Produktionstest ergibt sich, wie die Fehlerdaten in Tabelle 37 und die von den Probanden vorgenommenen Lautsubstitutionen zeigen, ein mit den Ergebnissen beider vorausgehender Produktionstests gut übereinstimmendes Bild. Für das geforderte

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 #_V	24	2	0,08	48	-	0.00
2 #C__	35	1	0.03	70	1	0.01
3 V__V	4	3	0.75	8	1	0.01
Σ	63	6	0.10	126	2	0.02

Tab. 37: Produktion-Perzeption-Produktion (engl. Sätze): /r/; Fehlerraten in den einzelnen Testkontexten

/r/ wurden substituiert: [w] (viermal), [s] (dreimal) und [v] (einmal). In den Erstrealisationen, die in die Fehlerwertung nicht eingingen, wurden darüber hinaus einige [ʊ]-Realisationen vorgefunden, die jedoch bei der zweiten Realisation des Stimulusatzes (nach dem Abhören der Modellrealisation) korrigiert wurden. Bei der geringen Fehlerzahl von nur acht verzeichneten Fehlern bei einer Gesamtzahl von 189 Äußerungen ergibt sich von selber, daß die Streuung der beobachteten Produktionsfehler gering ist.

In diesem Zusammenhang ist auf das Verhalten einer Vp der Gruppe "LR" zu verweisen, die ebenfalls zu allen Produktionstest herangezogen wurde, jedoch später keine Berücksichtigung fand und aus der Wertung ausschied.⁵⁵ Eine Auswertung der Produktionsleistungen dieser Probandin ergab, daß diese Vp in allen drei Produktionstests konsequent ihre eigensprachlichen r-Laute für die geforderten englischen r-Laute substituierte - auch in dem mit Nonsensematerial durchgeführten Test III, wo sie sich als durchaus imstande erwies, /r/ in den jeweiligen Testkontexten zu identifizieren. Auch die Hörtestergebnisse dieser Vp erlauben nicht den Schluß, daß Perzeptionsmängel zur Erklärung der konsequent abweichend produzierten r-Laute herangezogen werden können.

5.2.8. Beurteilung der /r/-Realisationen nach Akzeptabilität

Das insgesamt eindrucksvolle Bild von den Produktionsleistungen unserer zwölf Probanden bezüglich des "Problem- lautes" /r/ bzw. seiner verschiedenen Allophone wird modifiziert, wenn man die Akzeptabilitätsurteile des zweiten (ggf. auch des dritten) Beobachters betrachtet, wie sie in der graphischen Darstellung (s.u., Abbildung 25, S. 159) der Produktionsleistungen erfaßt worden sind. Aus dieser Darstellung geht hervor, daß die /r/-Realisationen unserer Probanden zwar in überwältigender Mehrzahl als "korrekt" einzustufen waren, was die Parameter Artikulationsart und -stelle betrifft, daß es sich also bei der großen Mehrzahl der /r/-Realisationen um Vorderzungenartikulationen (mit Approximation der leicht zurückgezogenen Zungenspitze an den Zahndamm) handelte. Die detaillierten Parameterurteile der Beurteiler lassen jedoch erkennen, daß insbesondere die zeitliche Koordination der Artikulationsbewegungen und die Gestaltung der

⁵⁵Es handelt sich um die Vp SA₇/LR, die die schlechtesten Hörtestergebnisse ihrer acht Personen umfassenden Gruppe aufzuweisen hatte (s.o., Tab. 5, S. 43). Ursprünglich war geplant, alle acht Angehörigen dieser Gruppe auch zu den Produktionstests heranzuziehen; das Vorhaben wurde jedoch später fallengelassen.

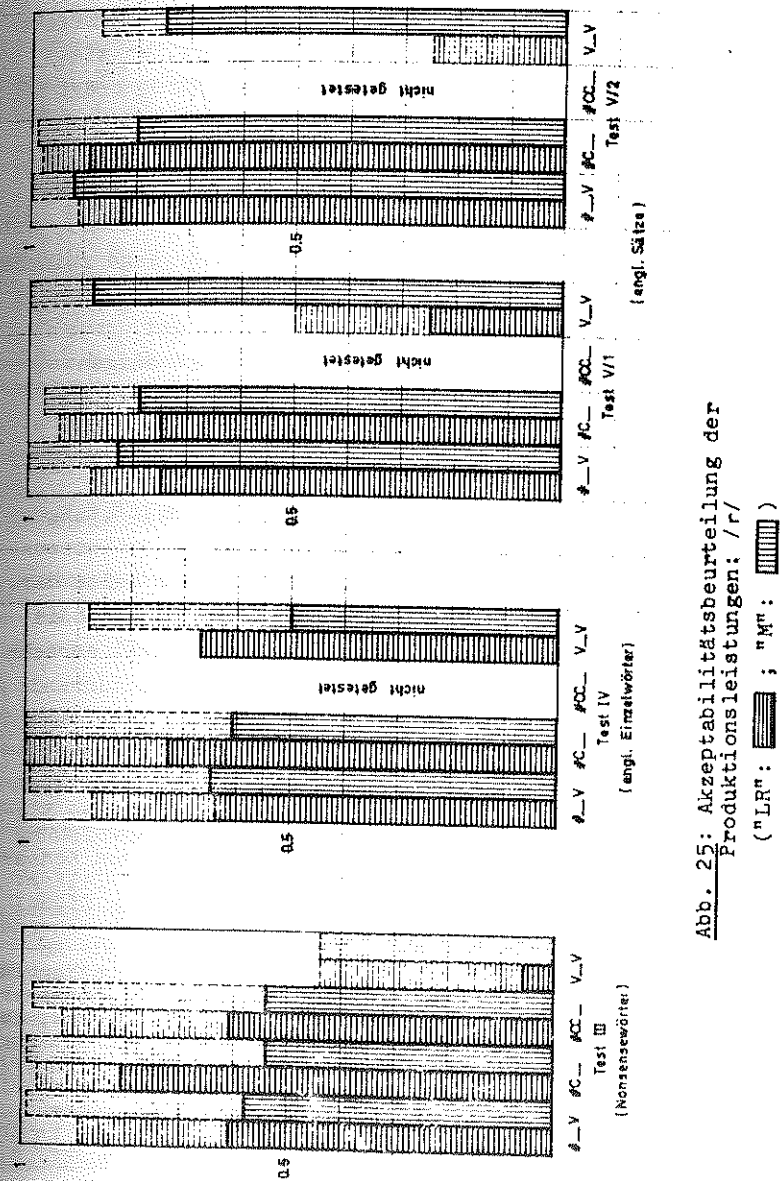


Abb. 25: Akzeptabilitätsbeurteilung der Produktionsleistungen: /r/ ("LR": [horizontal lines]; "M": [vertical lines])

Übergänge zu den Nachbarlauten in einer Weise geschehen, die die Bewußtheit der einzelnen motorischen Abläufe erkennen läßt und häufig zu einem unbefriedigenden akustischen Effekt führt.⁵⁶ Übermäßige Lippenrundung vor allem dort, wo sie durch den vokalischen Kontext nicht gefordert wird, trägt ebenfalls häufig zu der Einschätzung einer beträchtlichen Zahl von /r/-Realisationen als "(noch) akzeptabel" bei. Betrachtet man die als Produktionsfehler eingestuft /r/-Substitutionen, wie sie die Tabellen 34 und 36 (s.o., S.154 u. 156) verzeichnen, wird deutlich, daß die Übergänge zwischen als "(noch)akzeptabel" und "nicht akzeptabel" eingeschätzten /r/-Realisationen offensichtlich fließend sind und unsere deutschen Lerner bei der Produktion von /r/ übermäßigen Gebrauch von (zusätzlicher) Lippenartikulation machen.

5.2.9. Vergleich der Perzeptions- und Produktionsleistungen

Abbildung 26 stellt die Hör- und Produktionsleistungen unserer beiden Gruppen für die allen fünf Tests gemeinsamen Kontexte von /r/ (#_V und #C_) in der gewohnten Weise dar. Es muß nochmals darauf hingewiesen werden, daß den Stabdiagrammen für jede Gruppe sowie jeden Test individuell errechnete und über die Gruppe gemittelte Leistungswerte zugrunde

⁵⁶Eine akustische Analyse der von unseren Lernern produzierten /r/-Realisationen, insbesondere eine Analyse der Formantübergänge könnte Aufschluß über die spezifischen Signaleigenschaften dieser Laute und den Grad ihrer Abweichung von der Norm liefern.- Die zeitliche Koordination von Konsonantartikulationen in der Sprache englischer Kinder (Alter: 4-7 Jahre) hat S. Hawkins ausführlich am akustischen Signal studiert und mit den von M. Haggard ermittelten entsprechenden Daten der Erwachsenenaussprache verglichen. Bedeutende Abweichungen zwischen beiden Sprachmustern wurden vor allem für prä- oder postvokalisches /l/ als Element einer Konsonantverbindung gefunden, weniger deutlich auch für /s/. Für /r/, für welches bislang keine ausreichenden Daten vorliegen, ist möglichenfalls mit ähnlichen Abweichungen in der zeitlichen Organisation zu rechnen, die auf Integrationsschwierigkeiten bei der Koordination der motorischen Abläufe schließen lassen. Siehe S. Hawkins, "Temporal Coordination of Consonants in the Speech of Children: Preliminary Data", Journal of Phonetics, 1 (1973), 181-217.

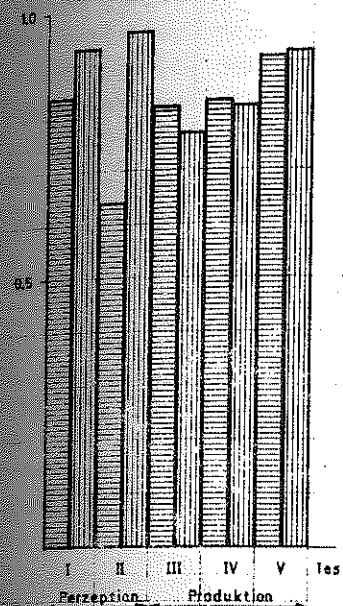


Abb. 26: Hör- und Produktionsleistungen: /r/ ("LR": ▨; "M": ▩)

liegen; die für die Produktionsleistung ermittelten Leistungswerte berücksichtigen dabei die als "gut" beurteilten Einzelrealisationen voll, die als "noch akzeptabel" beurteilten Realisationen jedoch nur zur Hälfte.

Den Leistungswerten liegen in diesem Falle zugrunde:

- Test I: 432 (Hörer-)Urteile
- Test II: 360 (Hörer-)Urteile
- Test III: 250 Realisationspaare
- Test IV: 238 Einzelrealisationen
- Test V: 177 Zweitrealisationen

Tabelle 38 stellt die individuellen Leistungswerte aller 12 Vp zusammen. Leistungsdaten wie die der Vpp RM, E' und EM (alle "M"), die sämtlich hervorragende Hörleistungen für /r/

Gruppe	Test	I	II	III	IV	V
	Vp					
"LR"	SA ₁	0.83	0.67	0.74	0.83	0.97
	SS ₂	0.75	0.53	0.86	0.82	0.90
	SS ₃	0.83	0.67	0.79	0.79	0.83
	WB	0.94	0.70	0.87	0.88	0.93
"M"	BM	1.00	0.97	0.57	0.67	0.86
	DL	0.97	1.00	0.93	0.89	0.97
	DM	0.86	0.73	0.66	0.90	0.93
	HH	0.89	0.97	0.76	0.83	1.00
	HU	0.92	1.00	0.73	0.73	0.87
	JK	0.94	0.97	0.81	0.90	0.77
	RR	0.94	1.00	0.96	0.98	1.00
	SM	0.92	1.00	0.77	0.63	0.93
MW		0.90	0.85	0.79	0.82	0.91
StA		0.07	0.18	0.11	0.10	0.07

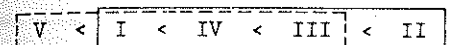
Tab. 38: Individuelle Perzeptions- und Produktionsleistungen: /r/

aufzuweisen haben, deren /r/-Realisationen in den Produktionstests jedoch häufig als fehlerhaft oder "(noch) akzeptabel" eingestuft wurden, legen ähnlich wie beim TH-Test nahe, daß ein unmittelbarer Zusammenhang zwischen Perzeption und Produktion für diesen Problemlaut nicht besteht und gute (oder auch schlechte: siehe das Verhalten von SS₃/LR) Hörleistungen bei manchen Lernern keinerlei Voraussagen über die Produktionsleistungen gestatten.

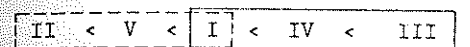
Die Rangvarianzanalyse (FRIEDMAN-Test) ergibt dennoch einen signifikanten Befund für den unterschiedlichen Einfluß der fünf Einzeltests auf die Leistungen der Probanden ($\chi^2_n = 12.44 > \chi^2_{n,0.05} = 9.49$). Mithilfe der multiplen Vergleiche nach WILCOXON-WILCOX ergibt sich als Rangfolge für die Schwierigkeit der einzelnen Aufgaben:



führt man den Test jedoch für jede der beiden Gruppen getrennt durch, was insbesondere angesichts des in Test II (Höridentifikation) gezeigten unterschiedlichen Testverhaltens angezeigt ist, ergeben sich für jede der beiden Gruppen abweichende Befunde, nämlich für "LR" ($\chi^2_n = 11.75 > \chi^2_{n,0.05} = 9.49$) die Reihenfolge



für "M" hingegen ($\chi^2_n = 15.8 > \chi^2_{n,0.05} = 9.49$) die Reihenfolge



Man wird hier aufgrund der Kenntnis des erraticen Testverhaltens der Gruppe "LR" in Test II (Höridentifikation) dem für "M" ermittelten Befund mehr Beachtung schenken; jedoch ergibt sich auch hier keine eindeutige Trennung von Perzeptionsaufgabe gegenüber Produktionsaufgabe, so daß man geneigt ist, den nachgewiesenen Effekt eher der spezifischen Art und Darbietungsform des Testmaterials in den einzelnen Tests zuzuschreiben als einer nicht nachzuweisenden wechselseitigen Abhängigkeit von Perzeption und Produktion.

5.3.0. V-Test

5.3.1. Perzeption: /i, e, æ/

Die Reihe der englischen palatalen Kurzvokale /i, e, æ/ ist der dritte Problembereich, zu dem in einer Serie von Hör- und Produktionstests empirische Befunde erhoben wurden. Diese Befunde verdienen um so größere Beachtung, als die Untersuchung dieses Bereichs nicht auf Schwierigkeiten bei der Materialauswahl und -darbietung stößt, wie das insbesondere bei der Untersuchung von /θ, ð/ und mit gewissen Einschränkungen auch von /r/ der Fall war.

5.3.2. Fehlervorgabe

Mit dem Untersuchungsbereich, der die gesamte Reihe der englischen Kurzvokale einschließt und sich nicht auf die gewöhnlich allein als problematisch gesehene /c/-/x/-Distinktion

beschränkt, waren die in beiden Perzeptionstests vorgegebenen Fehlermöglichkeiten zugleich gegeben: Untersucht wurde, inwieweit die 28 Vpp unserer beiden Gruppen in der Lage waren, die perzeptorische Unterscheidung der genannten drei Vokalphonem-Realisationen vorzunehmen.

Als Lernerfehler deutscher Englischlernender sind abweichende Realisationen wie [blæk] für black, [ʒɛd] für red, aber auch [mɪn] für men (pl.), die die Grenzen der als akzeptabel anzusehenden Realisationsnormen für /ɪ, e, ə/ ignorieren bzw. überschreiten, reich belegt. Als unproblematisch gilt allein ε /i/, dessen spezifische Ausprägungsform als mit ε /i/ übereinstimmend gesehen wird.⁵⁷

5.3.3. Hördiskrimination

Die drei Vokale wurden in Test I sowohl in initialer als auch in medialer Position zur Beurteilung angeboten. Die Hälfte der 84 Testwortreihen enthielt den Gegensatz /i/-/e/, die andere Hälfte den Gegensatz /e/-/ə/; die Opposition /i/-/ə/ war in keiner der Testwortreihen enthalten.

Tabelle 39 stellt die Fehlerdaten für Test I für beide Gruppen von Vpp zusammen, Abbildung 27 (S.166) stellt die Diskriminationsfehler in den beiden Testkontexten in Form von Flächen- diagrammen dar. Es zeigt sich, daß die Diskriminationsfähigkeit beider Gruppen für die drei Testvokale außerordentlich gut ist und die Probanden sich als fähig erwiesen, die geforderte perzeptorische Trennung in 92% ("LR") bzw. gar 97% ("M") der Fälle erfolgreich vorzunehmen.⁵⁸ Dies stellt eine signifikante Verbesserung der Hörleistungen sowohl gegenüber

⁵⁷ Die Realisationsnorm von ε /i/ ist - zumindest bei norddeutschen Sprechern - sowohl geschlossener als auch "palataler" als die von ε /i/; vgl. H.P. Jørgensen, "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen", *Phonetica*, 19 (1969), 217-245.

⁵⁸ Kontrollgruppe: 2 Fehler (<0.01%); beide Fehler bei Angehörigen der Untergruppe NSp.

Vokal	Kontext	"LR"			"M"		
		TW	FZ	FR	TW	FZ	FR
/i/	1 # C ₁ ³	64	3	0.05	160	5	0.03
	2 C ₁ ³ C ₁ ³	160	16	0.10	400	23	0.06
	Σ	224	19	0.08	560	28	0.05
/e/	1	64	4	0.06	160	3	0.02
	2	160	24	0.15	400	10	0.03
	Σ	224	28	0.12	560	13	0.02
/ə/	1	64	3	0.05	160	1 (<0.01)	
	2	160	5	0.03	400	1 (<0.01)	
	Σ	224	8	0.04	560	2 (<0.01)	
	$\Sigma \Sigma$	672	55	0.08	1680	43	0.03

Tab. 39: Hördiskrimination: /i, e, ə/; Fehlerraten in den beiden Testkontexten; Gruppe "LR" (8 Vpp); Gruppe "M" (20 Vpp)

dem TH-Test als auch gegenüber dem R-Test dar. /ə/ ist bei beiden Gruppen der am wenigsten fehlerhaft beurteilte Vokal; leichte Unterschiede zwischen beiden Gruppen ergeben sich für /i/ und /e/. Für die initiale Position scheinen gegenüber der medialen leichte Hörvorteile zu bestehen.

Eine andere Anordnung der Fehlerdaten ergibt für die beiden getesteten Distinktionen, daß die perzeptorische Trennung von /e/ und /ə/ beiden Gruppen besser gelingt als die Trennung von /i/ und /e/.⁵⁹ Zur Erklärung dieses Befundes können nicht

⁵⁹ Vgl. auch Mohr und Wang (1968), die in ihren Untersuchungen zur perzeptorischen Distanz von (qualitativ leider nicht näher bestimmten) Vokallauten mithilfe der Methode gepaarter Vergleiche herausfanden, daß [i] und [e] in den Ähnlichkeitsurteilen der naiven Hörer als einander ähnlicher eingestuft wurden als [e] und [ə]. Die entsprechenden numerischen, mithilfe des "semantic differential" ermittelten Distanzwerte sind für das Gegensatzpaar [i] - [e]: 1.81, für [e] - [ə]: 3.11. Zieht man phonologische Merkmale mit dem Ziel einer Verallgemeinerung des Befundes heran, ergibt sich allgemein für die Distanz zwischen einem "geschlossenen" und einem "mittleren" Vokal ein Wert von 2.26, für die zwischen

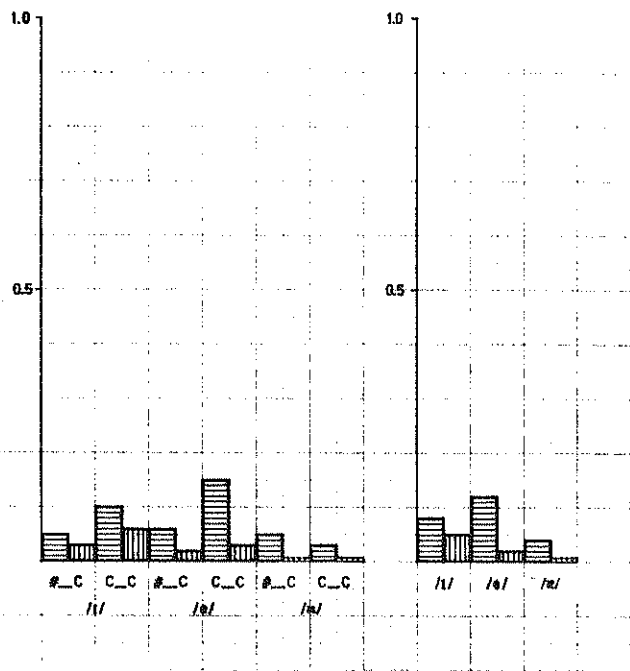


Abb. 27: Hördiskrimination: /i, e, a/;
Fehlerraten in den einzelnen Testkontexten
("LR": ■■■■ ; "M": ▨▨▨▨)

die phonologischen Verhältnisse im Deutschen herangezogen werden, das, wie Abbildung 28 zeigt, über die zusätzliche Distinktion des Englischen zwischen halboffenem und offenem palatalen Kurzvokal nicht verfügt; dieser Mangel ließe eher einen entgegengesetzten Befund erwarten.

einem "mittleren" und einem "offenen" Vokal ein Wert von 3.43. Die Autoren schließen aus ihren Befunden, daß "perceptual distances reflect physiological distances, in that a class whose members are articulatorily closer is judged to be perceptually closer." Bei fehlender Information über die spezifischen artikulatorischen und akustischen Eigenschaften der in der Untersuchung verwendeten Vokale muß diese Behauptung als unbewiesen angesehen werden.

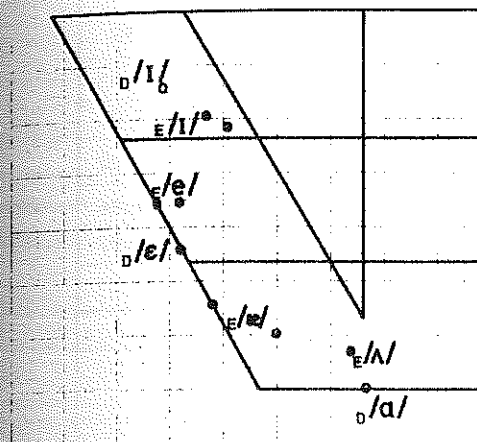


Abb. 28: Auditive Qualitäten⁶⁰
englischer und deutscher Kurzvokale

Die Realisationsnorm unseres Informanten für die drei Testvokale und insbesondere seine gegenüber der RP-Norm leicht geöffnete und "zurückgezogene" Realisation des /a/ liefert möglichenfalls eine Erklärung für den hier ermittelten Befund. Die insgesamt geringen Fehlerzahlen für diesen Test mahnen jedoch zur Vorsicht bei der Interpretation der Ergebnisse, und es muß abgewartet werden, welches Fehlerverhalten die Vpp in den weiteren Tests zu diesem Aussprachebereich an den Tag legen.

5.2.4. Höridentifikation

Tabelle 40 verzeichnet die Fehlervorkommnisse für den zweiten Perzeptionstest (Höridentifikation), bei dem durch

⁶⁰ Angabe der (auditiven) Qualitäten der englischen Kurzvokale nach Gimson (1970), S. 101-107; wir geben darüber hinaus unsere Einschätzung der Realisationsnorm unseres Informanten wieder, die durch * symbolisiert ist. - Für die deutschen Kurzvokale siehe Duden Aussprachewörterbuch, S. 29; vgl. auch Barry u. Gützknecht (1974), S. 45.

Vokal	Kontext	"LR"			"M"		
		TW	FZ	FR	TW	FZ	FR
/i/	1 # C ₁ ³	32	1	(0.03)	80	2	0.03
	2 C ₁ ³ C ₁ ³	96	13	0.14	240	4	0.02
	Σ	128	14	0.11	320	6	0.02
/e/	1	32	4	0.13	80	5	0.06
	2	96	7	0.07	240	4	0.02
	Σ	128	11	0.09	320	9	0.03
/a/	1	32	2	0.06	80	1	(0.01)
	2	96	9	0.09	240	8	0.03
	Σ	128	11	0.09	320	9	0.03
	ΣΣ	384	36	0.09	960	24	0.02

Tab. 40: Höridentifikation: /i,e,a/:
 Fehlerraten in den beiden Testkontexten;
 Gruppe "LR" (8 Vpp); Gruppe "M" (20 Vpp)

die Testkonstruktion gegeben war, daß jede einzelne, dem "Mottowort" mit dem zu identifizierenden "kritischen" Vokal folgende Vergleichswortreihe sämtliche drei Vokale enthielt, deren Reihenfolge jedoch systematisch variiert wurde. Die Anwesenheit einer dreifachen Distinktion in der Vergleichswortreihe ermöglichte demnach fehlerhafte Identifikationen von /i/ als /a/ (und auch umgekehrt), wie sie in Test I ausgeschlossen waren, wo Testwortreihen, die die einfache Distinktion /i/ - /a/ enthielten, nicht zur Beurteilung angeboten wurden.

Das sich aus Test II ergebende Fehlerbild⁶¹ für die einzelnen Vokale ist, wie Abbildung 29 zeigt, weitgehend mit dem Befund von Test I identisch, insofern /a/ sich erneut als der perzeptorisch am besten definierte Vokal erweist. Ein anderes Arrangement der Fehlerdaten zeigt jedoch, daß in diesem Test /e/ - /a/-Verwechslungen häufiger vorkamen als solche von /i/

⁶¹Kontrollgruppe: 1 Fehler (<0.01%).

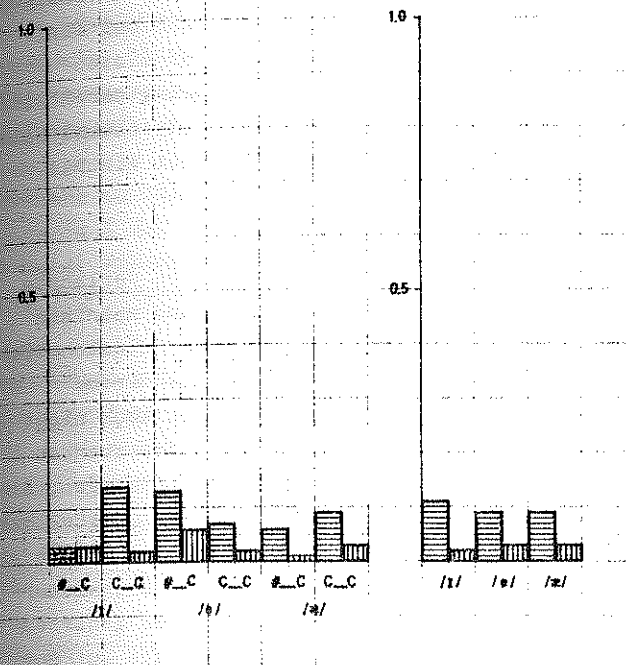


Abb. 29: Höridentifikation: /i,e,a/;
 Fehlerraten in den einzelnen Testkontexten
 ("LR": ■ ; "M": ▨)

und /e/, wobei jedoch in der relativen Häufigkeit dieser Identifikationsfehler kaum ein Unterschied besteht. Verwechslungen der nicht unmittelbar benachbarten Qualitäten /i/ und /e/ treten bei der Gruppe "M" nur bei einer einzigen Vp auf (WF₁/M: zwei Fehler),⁶² bei "LR" immerhin bei fünf von acht Vpp. Es ist aber festzuhalten, daß für diesen Problembereich eine signifikante Verschlechterung der Hörleistungen der Gruppe "LR" gegenüber Test I, wie sie sowohl beim TH- als auch beim R-Test beobachtet wurde, nicht eintrat.

5.3.5. Perzeption-Produktion: Nonsensewörter

Die in Test III gestellte Imitationsaufgabe verlangte von den Probanden eine schnelle Verarbeitung und spontane Nachahmung von Schallereignissen, die nicht nur hinsichtlich ihrer lautlich-segmentellen Zusammensetzung als "fremd" empfunden werden mußten, sondern darüber hinaus auch nicht semantisierbar waren. Fertige Artikulationsmuster standen ihnen somit nicht zur Verfügung und sie waren gezwungen, die ihnen über Kopfhörer zugespielten Stimuluswörter unmittelbar in Schallereignisse umzusetzen, die den gehörten gleich oder doch zumindest ähnlich waren. Von einem Prozeß des Wiedererkennens ("recognition"), der dem "matching" des gehörten Schallereignisses vorausgeht, kann hier allenfalls auf der Phonemebene gesprochen werden. Dabei gilt es zu bedenken, daß Phonembereiche und -grenzen streng genommen nur für Äußerungen definiert sind, die in einer spezifischen Sprache vorkommen und die sowohl phonetisch als auch semantisch unterschieden sind; sie sind jedoch nicht definiert für Nonsensewörter, die definitionsgemäß keiner Sprache zugehören. Es wird jedoch in Untersuchungen mit Nonsensematerial allgemein davon ausgegangen, daß dort, wo die Kategorisierung von phonetischen Ereignissen vorzunehmen ist, die als Phoneme bereits etabliert sind, die gleichen phonetischen Grenzen beachtet werden wie die, die sonst bedeutungstragende sprachliche Ein-

⁶²Für neun der 20 Vpp dieser Gruppe ist für Test II überhaupt kein Identifikationsfehler zu verzeichnen.

heiten unterscheiden.⁶³

Daß unsere Vpp in der Lage sind, eine solche Kategorisierung der Schallmerkmale vorzunehmen und sie den entsprechenden Phonembereichen der erlernten Fremdsprache zuzuweisen, erweisen die Ergebnisse unserer Perzeptionstests wie auch die der Produktionstests. Allerdings legen die Probanden insofern typisches Lernerverhalten an den Tag, als ihnen die phonematische Kategorisierung der relevanten Schalleigenschaften in der Fremdsprache nicht in jedem Falle gelingt, was im Fall der Hörtests zu Diskriminations- und Identifikationsfehlern führt, im Fall der Produktionstests zu Lautrealisationen, die z.T. anderen Phonembereichen zugeordnet werden müssen als die Modellrealisationen des entsprechenden Lautes.

Die Ergebnisse von Test III für die palatalen Kurzvokale des Englischen zeigen zwar, daß die Mehrzahl der von den Probanden produzierten Vokalrealisationen den Phonembereichen von /i, e, æ/ zuzuordnen sind, daß jedoch ein weitaus höheres Maß an Unsicherheit zu verzeichnen ist, als es die ausgezeichneten Hörleistungen in diesem Bereich erwarten ließen. Tabelle 41 stellt die Fehlerdaten für beide Gruppen von Vpp zusammen, wobei sich eine beträchtliche Zunahme der zu verzeichnenden Fehler gegenüber dem R-Test ergibt: Nur 80% der Vokalrealisationen der Gruppe "LR" und 79% der von den Ange-

⁶³Siehe L. Lisker, "Minimal Cues for Separating /w, r, l, y/ in Intervocalic Position", *Word*, 13 (1957), 256-267, bes. 259, Fn. 6. Vgl. auch die Untersuchungen von Scholes, in denen 69 synthetisch erzeugte Vokale mit systematisch variierten F₁/F₂-Kombinationen Vpp mit unterschiedlichem muttersprachlichen Hintergrund zur Beurteilung angeboten wurden. Die Aufgabe der Vpp bestand darin, die auditiv dargebotenen isolierten Vokalstimuli eigensprachlichen Phonemkategorien zuzuordnen oder aber als keiner solchen Kategorie entsprechend zurückzuweisen. Die Ergebnisse zeigten, daß die Vpp die synthetischen Vokale für solche Zwecke akzeptierten und eine Untermenge der Gesamtmenge von 69 Stimuli systematisch und konsistent Phonembereichen zuordnen konnten. Siehe R.J. Scholes, "Phoneme Categorization of Synthetic Vowel Stimuli by Speakers of Japanese, Spanish, Persian, and American English", *Lang. & Speech*, 10 (1967), 46-68; dsb., "Categorical Responses to Synthetic Vowel Stimuli by Speakers of Various Languages", *Lang. & Speech*, 10 (1967), 252-273.

Vokal	Kontext	"LR"			"M"		
		TW	FZ	FR	TW	FZ	FR
/i/	1 # C ₁ ³	16	-	0.00	32	-	0.00
	2 C ₁₋₁ ³ -C ₁ ³	32	11	0.34	64	13	0.20
	Σ	48	11	0.23	96	13	0.14
/e/	1	16	1	0.06	32	-	0.00
	2	31	5	0.16	64	8	0.13
	Σ	47	6	0.13	96	8	0.08
/a/	1	16	4	0.25	32	18	0.56
	2	32	7	0.22	64	21	0.33
	Σ	48	11	0.23	96	39	0.41
	ΣΣ	143	28	0.20	288	60	0.21

Tab. 41: Perzeption-Produktion (Nonsensewörter): /i, e, a/; Fehlerdaten in den einzelnen Testkontexten

hörigen der Gruppe "M" produzierten Vokale wurden von den Beurteilern als "gut" bzw. "(noch)akzeptabel" eingestuft; die fehlerhaften 20 bzw. 21% können nicht mehr den jeweiligen Phonembereichen von /i, e, a/ zugeordnet werden. Einen Überblick über die von den Probanden vorgenommenen Lautsubstitutionen gibt Tabelle 42;⁶⁴ die Art der verzeichneten Fehler zeigt deutlich, daß die Fehlrealisationen der Vpp sich auf den gesamten palatalen Bereich und auch große Teile des zentralen Bereichs erstrecken. Ehe jedoch aus diesen Fehlerdaten Schlüsse in bezug auf die von den Probanden bevorzugte Richtung der Abweichungen für die einzelnen Vokalphoneme gezogen werden können, ist es notwendig, einen näheren Blick auf die Akzeptabilitätsurteile unserer Beobachter zu werfen.

⁶⁴Nur Verstöße gegen die Vokalqualität werden verzeichnet; die meist in Zusammenhang mit den Stimmbeteiligungseigenschaften des auslautenden Konsonanten zusammenhängenden fehlerhaften Realisationen der Dauereigenschaften werden in Zusammenhang mit dem FVC-Test abgehandelt.

Zielvokal	Fehler	Fehlerbefund	Häufigkeit
/i/	[ɛ]	zu offen	11
	[i:] ⁶⁵	zu geschlossen	8
	[ë], [e]	zentralisiert	2
	[ÿ]	gerundet/zentral.	1
/e/	[i]	zu geschlossen	4
	[Y]	geschl./zentral.	4
	[ɛ]	zu offen	4
	[ë]	zentralisiert	1
	[e ¹] ⁶⁶	diphthongiert	1
/a/	[ä]-[ä]	zu offen/zurück	31 ⁶⁷
	[a ¹] ⁶⁸	diphthongiert	10
	[ɛ]	zu geschlossen	7
	[ɛ]	zu geschlossen	2

Tab. 42: Produktionsfehler: Vokalsubstitutionen für /i, e, a/ (Perz.-Prod.: Nonsensewörter); Fehlerdaten beider Gruppen

Abbildung 33 (s.u., S. 188) zeigt, daß von allen Vokalrealisationen unserer 12 Probanden überhaupt nur 33% als "nicht abweichend" bzw. "gut" beurteilt wurden und demnach die meisten Urteile - "LR": 47%; "M": 46% - auf die Kategorien "(noch) akzeptabel" entfielen. Da bei der Abgabe eines solchen Urteils die Art der festgestellten Abweichung unter Angabe des oder der beteiligten Parameter(s) vom jeweiligen Beobachter notiert wurde, machen wir uns die in den phonetischen Urteilen enthaltene Information zunutze, um ein genaueres Bild von der Richtung der vorgenommenen Abweichungen zu erhalten.

In Abbildung 30 sind die bevorzugten Richtungen der Vokalabweichungen für /i, e, a/ dargestellt, wie sie sich aus den Parameterurteilen für die als "(noch) akzeptabel" und die als

⁶⁵Alle Fehler im Testwort /trig/ mit "stimmhaftem" Endkonsonanten und allophonischer Länge des Vokals.

⁶⁶Testwort: /treg/.

⁶⁷Davon nur 2 bei "LR", 29 bei "M".

⁶⁸Testwort: /trag/.

"nicht akzeptabel" beurteilten Realisationen gewinnen lassen. Die auditiven Qualitäten der Modellrealisationen unseres Informanten geben den jeweiligen Bezugspunkt ab; wir stellen darüber hinaus die von Gimson für die RP angegebenen (auditiven) Qualitäten der jeweiligen Vokalphoneme dar, da sie den Bereich der von unseren Beobachtern als "gut" bzw. "(noch) akzeptabel" bezeichneten Vokalrealisationen ebenso definieren wie die etwas "extremere" Realisationen unseres Informanten.⁶⁹

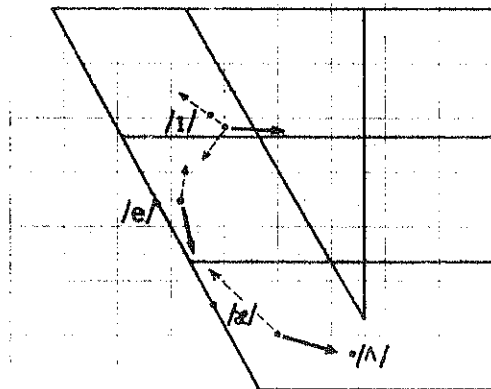


Abb. 30: Hauptrichtung der Abweichungen bei der Realisation von /i, e ə/ 70 (Perz - Prod: Nonsensewörter)

Die Analyse der Fehlrealisationen unserer Probanden, wobei wir in diesem Falle auch die beträchtliche Zahl der als

⁶⁹ Siehe weiter oben, 2.1.1., S. 14-16.

⁷⁰ • : Gimsions Angabe der auditiven Qualitäten der RP-Vokale (vgl. Gimson (1970), S. 101-107); ◦ : unsere Einschätzung der Realisationsnorm unseres Informanten für /i, e, ə/.

"(noch) akzeptabel" beurteilten Produktionen einschließen,⁷¹ ergibt einige unerwartete Befunde. Fehler sind für alle drei der untersuchten Vokalphoneme zu verzeichnen, wobei /ə/ die geringste Fehlerhäufigkeit aufzuweisen hat, gefolgt von /i/, das häufiger als erwartet abweichend realisiert wurde. Soweit ist der Fehlerbefund für beide Gruppen von Vpp übereinstimmend, deren Testverhalten insofern mit dem sonst beobachteten übereinstimmt, als die Angehörigen von "M" auch für diese beiden Problemlaute bessere Leistungen erzielen als die der Gruppe "LR". Dies trifft jedoch nicht auf die /ə/-Realisationen der Angehörigen von "M" zu, die häufiger als "nicht akzeptabel" beurteilt wurden als die der Mitglieder von "LR".

Mit einiger Sicherheit kann angenommen werden, daß diese Tatsache ihren Grund in der Realisationsnorm unseres Informanten für diesen Laut hat, dessen Kategorisierung den Probanden offensichtlich erhebliche Schwierigkeiten bereitete. Die gegenüber der RP-Norm (leicht) geöffneten und mit zurückgezogener Zunge ("weiter hinten") artikulierten /ə/-Laute unseres Informanten - phonetisch etwa [ə̠] - führten zu einer Vielzahl von (Fehl-)Realisationen, die bis in den Bereich des englischen /ʌ/-Phonems, in einigen Fällen auch des (ähnlichen) deutschen /a/-Phonems reichen. Ob eine solche phonematische Zuordnung jedoch tatsächlich von den Vpp vorgenommen wurde, ist im Lichte anderer Fehlerbefunde anzuzweifeln. Allenfalls der für die /e/-Realisationen unserer Probanden ermittelte Befund könnte ihn bestätigen, da die meisten verzeichneten Abweichungen in die Richtung größerer Öffnung gehen und in den Phonembereich von ə/e/ fallen. Immerhin sind jedoch auch hier Abweichungen in die Richtung geringerer Öffnung und/oder

⁷¹ Die Grenze zwischen als "(noch) akzeptabel" und "nicht akzeptabel" eingestuftem Vokalrealisationen kann nicht scharf gezogen werden. Um so notwendiger ist es, für die Beurteilung der Lernerrealisationen ein abgestuftes Kriterium zur Verfügung zu haben, das dem Übergangscharakter vieler Realisationen gerecht wird und aufzuzeigen vermag, welche Richtung die Abweichungen nehmen. Dies ist selbstverständlich nur bei genügender Spezifikation der phonetischen Urteile möglich, wie sie hier von geschulten Beobachtern gefällt wurden.

Zentralisierung zu verzeichnen. Die große Mehrzahl der als abweichend beurteilten /ɪ/-Realisationen jedoch nehmen Richtung auf einen (zentralen) Bereich, der durch kein deutsches oder englisches Vokalphonem spezifisch gedeckt ist, und Ähnliches gilt für die meisten der von den Vpp für /a/ substituierten Vokale, die weder als spezifisch deutsch noch als spezifisch englisch bezeichnet werden können.

Das hier beobachtete Verhalten ähnelt in entscheidenden Punkten dem Verhalten der Vpp im entsprechenden TH-Test (mit Nonsensewörtern), wo die Art der verzeichneten Produktionsfehler den Schluß nahelegte, daß die Vpp die ihnen vorgespielten Wörter zunächst *p e r z e p t o r i s c h* kategorisieren und ihre Produktion so organisieren, daß diese der perzeptorischen Kategorisierung entspricht. Dieses Verhalten vermag wenigstens zu einem Teil das Maß an Flexibilität zu erklären, das die Schüler(innen) bei dieser Imitationsaufgabe an den Tag legten.

Eben diese Flexibilität, die zu einer Vielzahl von "Überkompensationen" geführt hat, erklärt die hohe Anzahl von Fehlrealisationen von /a/ bei den Angehörigen von "M": 29 der 31 fehlerhaften (= "nicht akzeptablen") Realisationen von /a/ als [ä] bzw. [ü] gehen auf ihr Konto. Das von den Mitgliedern dieser Gruppe bewiesene hohe Anpassungsvermögen ist ihnen dabei jedoch vielfach nicht zustatten gekommen, da die eigenen Realisationen die Grenzen des Phonembereichs von /a/ überschreiten und von unseren Beurteilern, die mit der extremen Realisationsnorm des Informanten in diesem Bereich wohl vertraut sind, als nicht mehr akzeptabel bezeichnet werden mußten.⁷²

⁷²Allerdings ist für diesen Einzellaut ein hohes Maß an Nichtübereinstimmung beider Beobachter festzustellen gewesen. Es erwies sich dabei, daß der mit der Realisationsnorm des Informanten in den Stimuluswörtern besser vertraute erste Beobachter (Verf.) gerade in dem "hinteren" Realisationsbereich großzügiger urteilte als der zweite Beobachter (native speaker), dessen eigene Realisationsnorm von /a/ der RP-

5.3.6. Perzeption-Produktion: englische Einzelwörter

Es wurde bereits mehrfach darauf hingewiesen, daß die in Test IV gestellte Imitationsaufgabe sich insofern von der des Nonsensewörtertests unterscheidet, als sie den Probanden ermöglichte, die ihnen über Kopfhörer zugespielten Stimuluswörter nicht nur perzeptorisch zu kategorisieren und die einzelnen Schallmerkmale phonologischen Kategorien zuzuordnen, sondern darüber hinaus diese Stimuli mit bereits gespeicherten lexikalischen Einheiten zu identifizieren. Der Prozeß des "recognition and matching" spielt sich hierbei nicht nur auf der Ebene diskreter phonetischer und phonematischer Merkmale ab, sondern schließt die semantische Ebene mit ein und gestattet das Wiedererkennen ganzer Einheiten, die auch als solche abrufbar sind.

Tabelle 43 verzeichnet die Fehlerdaten für diesen Test für beide Gruppen von Vpp; auf eine Aufschlüsselung der phonetischen Kontexte wird dabei verzichtet, da nur zwei der 38 Testwörter den Vokal in initialer Position aufwiesen.

Vokal	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
/ɪ/	40	3	0.07	87	3	0.03
/e/	52	4	0.08	104	6	0.06
/a/	55	17	0.31	111	48	0.43
Σ	147	24	0.16	302	57	0.19

Tab. 43: Perzeption-Produktion (engl. Einzelwörter): /ɪ, e, a/; Fehlerraten für die einzelnen Testvokale

Die Abnahme der Fehler für /ɪ/ und /e/ bei beiden Gruppen gegenüber dem Nonsensewörtertest ist signifikant: Zweifellos

Norm eher entspricht. Dennoch ergab die Nachbeurteilung der zunächst nicht übereinstimmend beurteilten /a/-Realisationen keine systematische Bestätigung der Urteile des ersten Beobachters, die, wenn sie sich ergeben hätte, eine Nachbeurteilung aller /a/-Realisationen durch den dritten Beobachter notwendig gemacht hätte.

ist hier der gleiche Effekt wirksam, den wir in demselben Test für TH haben beobachten können. Ein Blick auf die Akzeptabilitätsurteile unserer Beobachter zeigt zudem, daß die Fehlerabnahme in diesem Bereich Hand in Hand geht mit einer deutlichen Zunahme der in der Kategorie "gut" bzw. "nicht abweichend" abgegebenen Urteile: 63% der /i, e/-Realisationen der Gruppe "LR" und 68% bei "M" wurden so klassifiziert. - Kaum verändert ist jedoch die hohe Anzahl als "nicht akzeptabel" bzw. "abweichend" beurteilter /a/-Realisationen ("LR": 31%; "M": 43%), und für diesen Laut sind auch gegenüber Test III nicht mehr Urteile in der Kategorie "gut" abgegeben worden (vgl. Abb.33, S. 188).

Ein Blick auf die verzeichneten fehlerhaften Vokalsubstitutionen für /i, e, a/ (siehe Tabelle 44) sowie auf Abbildung 31,

Zielvokal	Fehler	Fehlerbefund	Häufigkeit
/i/	[ə]	zu offen/zentral.	3
	[ä]	zentralisiert	2
	[ɜ]	zentralisiert	1
	[eʷ]	zentral./diphth.	1
/e/	[ɛ]	zu offen/zentral.	4
	[eʷ] ⁷³	diphthongiert	3
	[ɛ:]	zu geschlossen	1
	[eʷ]	zu offen/diphth.	1
	[ɛ̃]	zu offen/zentral.	1
/a/	[ɛ] _(,)	zu geschlossen	36 ⁷⁴
	[ä] - [ä]	zu offen/"zurück"	29 ⁷⁵

Tab. 44: Produktionsfehler: Vokalsubstitutionen für /i, e, a/ (Perz.-Prod.: engl. Einzelwörter); Fehlerdaten beider Gruppen

wo die Hauptrichtung der auditiv als "nicht akzeptabel" bzw.

⁷³Sämtlich vor "lenis"-Konsonant (Testwörter leg, head).

⁷⁴Davon 2 bei "LR", 34 bei "M".

⁷⁵"LR": 5, "M": 24.

"(noch) akzeptabel" beurteilten Abweichungen dargestellt ist, zeigt, daß trotz verringerter Fehlerzahl im Bereich der /i, e/-Realisationen die Hauptrichtung der Abweichungen die gleiche ist wie in Test III; für die abweichenden /a/-Realisationen ergibt sich jedoch ein zwiespältiges Bild, das erneut die große Unsicherheit der Probanden bezüglich der Kategorisierung dieses Problemlautes in seiner hier vorliegenden "extremen" Realisationsnorm zeigt. Berücksichtigt man neben

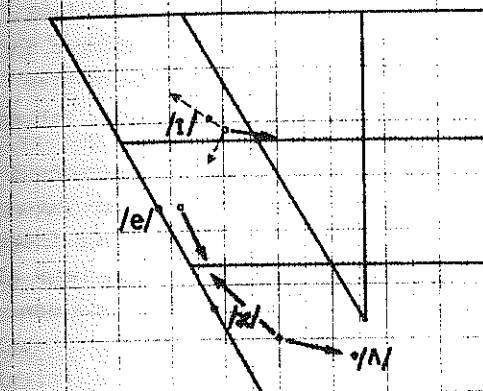


Abb. 31: Hauptrichtung der Abweichungen: /i, e, a/ (Perz - Prod: engl. Einzelwörter)

den Fehlerdaten der Tabelle 44 die Parameterurteile über die als "(noch) akzeptabel" eingestuftten Realisationen, ergibt sich zwar bei beiden Gruppen ein leichtes Übergewicht der Abweichungen in Richtung auf [ɛ]; eine hohe Zahl von Abweichungen nimmt jedoch Richtung auf den "hinteren" Bereich und gerät damit in gefährliche Nähe des Phonembereichs von ε/ʌ/.

Der beobachtete Effekt verlangt nach einer näheren Überprüfung der Fehlerdaten. - Ähnlich zweideutige Fehlerbefunde für

/θ,ð/ in Test IV wurden der spezifischen Darbietungsform dieses Tests angelastet und (besonders im Fall von /ð/) auf Identifikationsfehler bzw. -mängel der Probanden bei spezifischen Testwörtern zurückgeführt. Es kann nicht ausgeschlossen werden, daß ähnliche Faktoren auch bei unserem Vokaltest wirksam waren. Zuvor muß freilich sichergestellt sein, daß die beobachteten Schwankungen in den /æ/-Realisationen auch intraindividuell auftreten und Abweichungen in Richtung auf [ɛ] nicht etwa nur einer Untergruppe von Vpp vorbehalten sind, während eine andere Gruppe von Vpp ausschließlich Abweichungen in Richtung auf [ä] vornimmt. Bei jeder einzelnen Vp lassen sich jedoch Abweichungen in beide genannten Richtungen belegen, so daß individuelles Verhalten den beobachteten Effekt nicht zu erklären vermag.

Ein Blick auf besonders häufig fehlerhaft bzw. abweichend realisierte Testwörter zeigt, daß der Effekt in der Tat lexikalischer Natur ist. Tabelle 45 führt sieben der insgesamt 14 Testwörter auf, die /æ/ enthielten und von mehr als 10 der 12 Vpp abweichend realisiert wurden, d.h. von den Beurteilern als "nicht akzeptabel" bzw. "(noch) akzeptabel" eingestuft wurden. Nur eine genaue Kenntnis der Unterrichtsstoffe und -methoden der beiden Gruppen könnte Aufschluß darüber geben, ob die unterschiedliche Unterrichtsfrequenz spezifischer lexikalischer Einheiten gegenüber anderen ihre richtige Kategorisierung in unserem Test erleichtert oder erschwert hat bzw. ob hier der Grund für Realisationen liegt, die die als typisch zu bezeichnenden Lernerfehler deutscher Englischlernender in besonders hohem Maße aufweisen; eine solche Vermutung legt sich aber nahe angesichts der häufigen Fehlrealisationen bzw. Abweichungen bei Stimuli wie thank you, black, happy etc., die sicherlich leichter zu identifizieren waren als Testwörter wie bad und fat. Die Tabelle vermag darüber hinaus zu zeigen, daß nahezu die Hälfte aller als abweichend eingestuften /æ/-Realisationen für die Testwörter 1 - 5 (siehe Tab. 45) zu verzeichnen sind, auf die zugleich 87% der Abweichungen in Richtung auf [ɛ] entfallen.

Testwort	Abweichende Realisierung von /æ/ in Richtung	
	[ɛ]	[ä]
1. happy	8	2
2. hand	9	2
3. black	14	3
4. thank you	13	-
5. cat	9	1
6. bad	1	9
7. fat	2	8

Tab. 45: Abweichende Realisationen von /æ/ in den am häufigsten als abweichend beurteilten Testwörtern; Fehlerdaten beider Gruppen

Für die Abweichungen in Richtung auf [ä] muß hingegen festgestellt werden, daß sie konzentriert nur bei den Testwörtern 6 und 7 auftreten, sich ansonsten aber über das gesamte übrige Testmaterial gleichmäßig verteilen.

5.3.7. Produktion-Perzeption-Produktion: englische Sätze

Das Vorliegen der Testsätze in verschrifteter Form im letzten der von uns durchgeführten Produktionstests beseitigt diejenigen Schwierigkeiten der phonematischen und lexikalischen Zuordnung, die bei einer rein auditiven Darbietung der Stimuluswörter in Test III und IV sowohl für /θ,ð/ als auch für die palatalen Kurzvokale /i, e, æ/, jedoch nur in geringem Maße für /r/ zu verzeichnen waren. Schafft die verschriftete Form insofern Erleichterung, kann sie andererseits doch neue Schwierigkeiten beinhalten, die die Umsetzung des Geschriebenen in lautsprachliche Produktion betreffen. Wo die Graphem-Phonem-Korrespondenzen so eindeutig sind wie bei TH und R, werden sie in geringerem Maße eintreten als bei unseren Testvokalen, bei denen eine nicht-eindeutige Korrespondenz von Schriftzeichen und "Laut" zumindest im Falle einzelner Testwörter gegeben ist. So muß beispielsweise bei Wörtern wie pretty und many damit gerechnet werden, daß die insgesamt

geringen Kenntnisse unserer Probanden von Graphem-Phonem-Korrespondenzen im Englischen beim Lesen der Sätze Anlaß zu Fehlrealisationen geben - obwohl die Probanden mit der "korrekten" Aussprache der Testwörter vertraut sein müßten.

Solche Fehlrealisationen, die die Kenntnis einiger grundlegender Regeln über Graphem-Phonem-Entsprechungen voraussetzen und von falscher Generalisierung dieser Regeln zeugen, sind in den Erstäußerungen unserer Probanden in geringer Anzahl belegbar - so z.B. die Aussprache von lives (beim Lesen des Satzes "Billy lives in our lane") als [laɪvz], von reads (in "Billy reads a book") als [ɹɛdʒ], von friends als [fɹɪ:ndʒ] oder (am häufigsten) pretty als [pɹɛtɪ]. Die Fälle, in denen solche unmittelbar durch die Schrift verursachten Fehlaussprachen (im Bereich der hier den Gegenstand der Untersuchung bildenden Vokale) vorkommen konnten, sind damit bereits aufgezählt. Es leuchtet ein, daß solche Fehler anderer Natur (wenngleich von nicht geringem Interesse) sind als die von uns untersuchten systematischen Lernerfehler bei der Aussprache der englischen palatalen Kurzvokale; sie verdienen insofern eine gesonderte Behandlung.

Hier liegt denn auch der Grund dafür, daß den Vpp nach dem Ablesen der Sätze von Karten eine Modellrealisation des gelesenen Satzes über Tonband zugespielt wurde und sie denselben Satz unmittelbar danach noch einmal realisieren mußten: Die groben, unmittelbar auf den Einfluß der orthographischen Vorlage zurückzuführenden "phonemisch-graphemischer" Fehler konnten dadurch ausgeschaltet werden; wie sich zeigte, wurden sie in der Tat in nahezu allen Fällen korrigiert. Um so aufschlußreicher sind die in den Zweitrealisationen der Probanden verbleibenden abweichenden Realisationen unserer "kritischen" Vokale, in denen wir die systematischen Fehlertendenzen, wie sie bei deutschen Englischlernenden zu beobachten sind, aufzeigen können.

Tabelle 46 stellt die Fehlerdaten für Test IV zusammen, wobei nur die Zweitrealisationen unserer Probanden Berücksichtigung finden.⁷⁶

Vokal	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
/ɪ/	51	3	0.06	78	3	0.04
/e/	32	1	0.03	67	1	0.01
/æ/	83	57	0.69	165	78	0.47
ɛ	166	61	0.37	310	82	0.26

Tab. 46: Produktion-Perzeption-Produktion (engl. Sätze) : /ɪ, e, æ/; Fehlerraten für die einzelnen Testvokale

Die von den Probanden vorgenommenen Lautsubstitutionen für /ɪ, e, æ/ verzeichnet Tabelle 47.

Zielvokal	Fehler	Fehlerbefund	Häufigkeit
/ɪ/	[ɪ]	zu geschlossen	2
	[ɛ]	zentralisiert	1
/e/	[ɛ]	zu offen	2
/æ/	[ɛ]	zu geschlossen	113 ⁷⁷
	[ä]	zu offen/zurück	16 ⁷⁸
	[ə] [ɛə]	diphthongiert	je 1
	[ə]	zu geschl./zentral.	1

Tab. 47: Produktionsfehler: Vokalsubstitutionen für /ɪ, e, æ/ (Prod.-Perz.-Prod.: engl. Sätze); Fehlerdaten beider Gruppen

⁷⁶Die auf die einzelnen Kategorien entfallenden Urteile können den Flächendiagrammen von Abbildung 33 (S. 188) entnommen werden, wobei die für Erst- und Zweitrealisationen getrennte Darstellung unmittelbar die vorgenommenen Korrekturen erkennen läßt.

⁷⁷"LR": 55; "M": 58.

⁷⁸"LR": 1; "M": 15.

Die hier zusammengetragenen Fehlerdaten geben jedoch ein unvollständiges Bild von der Richtung der abweichenden Vokalrealisationen und es ist notwendig, die Akzeptabilitätsurteile unserer Beobachter, wie sie Abbildung 33 (unten S. 188) darstellt, mit in die Fehlerbetrachtung einzubeziehen. Hier zeigt sich, daß von den drei Testvokalen nur /i/ uneingeschränkt von der in Test V gewählten Darbietungsform profitiert und die Realisationen der Probanden nicht nur wenige Urteile der Kategorie "nicht akzeptabel", sondern auch insgesamt eine geringe Zahl von Urteilen der Kategorie "(noch) akzeptabel" auf sich ziehen: Die große Mehrzahl der Urteile für diesen Vokal fällt in die Kategorie "gut" bzw. "nicht abweichend".

Für /e/ ist ebenfalls nur eine geringe Anzahl von Fehlrealisationen zu verzeichnen (bei jeder Gruppe jeweils nur ein einziger Fehler), jedoch wurden 59% aller /e/-Realisationen von "LR" und 49% von "M" als "abweichend" bzw. "noch akzeptabel" beurteilt, was eine deutliche Verschlechterung gegenüber Test IV bedeutet. Die Hauptrichtung der Abweichungen, wie sie sich aus den phonetischen Urteilen unserer Beobachter ergibt, stimmt dabei jedoch mit der in Test III und IV ermittelten überein.

Anders die /a/-Realisationen unserer Vpp, die sowohl gegenüber Test III als auch Test IV eine höhere Zahl fehlerhafter bzw. abweichender Realisationen aufweisen und von denen überhaupt nur 4 ("LR") bzw. 8% ("M") als "gut" eingestuft wurden. Abbildung 32 (siehe S. 185) zeigt darüber hinaus, daß in diesem Test die Abweichungen von der korrekten Realisationsnorm für /a/ eindeutig in die Richtung einer "zu geschlossenen" Realisation weisen und nur in einer geringen Zahl von Fällen den "zurückgezogenen" Bereich erschließen.

Solche abweichenden Realisationen in Richtung auf [ä] sind überhaupt (mit einer einzigen Ausnahme) nur für die Angehörigen von "M" zu verzeichnen und betreffen nahezu ausschließlich dasselbe Testwort fat, dessen "zurückgezogene" Reali-

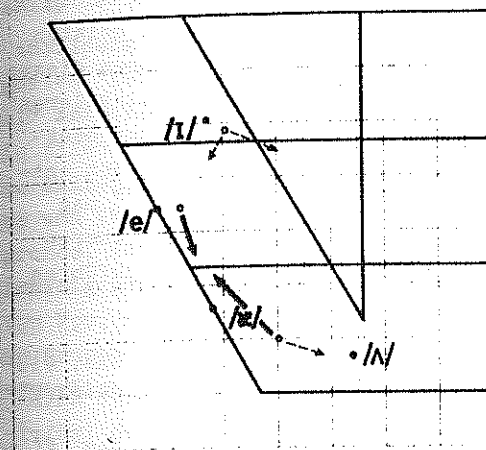


Abb. 32: Hauptrichtung der Abweichungen: /i, e, a/
(Prod - Perz - Prod: engl. Sätze)

sationsnorm von diesen Schülern im Verlauf der einzelnen Tests neu erworben wurde und eine neue Kategorisierung dieses Lautes zur Folge hatte. Einen solchen Schluß lassen Realisationen des Testsatzes "The dog is black and fat" als [ðə dɔg ɪz blæk ənɪ fæt] zu, wie sie bei einigen Vpp belegt werden können.

Der für /a/ erhobene widersprüchliche Fehlerbefund, wie er sich besonders aus Test III und IV, in geringem Maße auch aus Test V ergibt, muß unmittelbar auf die extremere Realisationsnorm dieses Lautes bei unserem Informanten zurückgeführt werden, mit der die Vpp vor Durchführung des Tests nicht vertraut waren. Wir sehen hierin einen besonders aufschlußreichen Befund insofern, als er darauf verweist, zu welcher bemerkenswerten Modifikationen und Kompensationen ihres phonetischen Verhaltens unsere Vpp in der Lage sind, wanngleich dieses Verhalten in unserem Falle zu neuen, vorher nicht einkalkulierten Reali-

sationsfehlern führt. - Auch der für /i/ erhobene Befund läßt ähnliche Schlüsse zu: Die Tatsache, daß die meisten der als abweichend einzustufenden Fehlrealisationen dieses Vokals nicht wie erwartet in den Bereich der "palataleren" und geschlosseneren Realisationsnorm des ähnlichen deutschen Kurzvokalphonems /i/ vorstoßen, sondern in den entgegengesetzten zentralen Bereich bis hin zu [e],⁷⁹ mag ebenfalls auf die geringfügig zentraleren und offeneren Realisationen unseres Informanten für /i/ zurückzuführen sein.

Ansonsten erlauben unsere Ergebnisse den Schluß, daß, wie fest auch immer die unterschiedliche perzeptorische Kategorisierung der entsprechenden Vokalphoneme in den individuellen Lerner-Systemen etabliert sein mag, wie es die Ergebnisse der Perzeptionstests nahelegen,⁸⁰ die lautsprachliche Produktion unserer Vpp darauf verweist, daß für /e/ und /a/ mit dem Vorhandensein unterschiedlicher Artikulationsmuster und Zielvorstellungen bei unseren Vpp vielfach nicht gerechnet werden kann. Die meisten Realisationen von /a/ (sieht man von dem besonderen oben genannten Effekt einmal ab) fallen in den Phonembereich von /e/ und weisen z.T. sogar einen geringeren Öffnungsgrad auf als die "offeneren" der ebenfalls als abweichend einzustufenden /e/-Realisationen.

⁷⁹Secondary cardinal vowel 19: vgl. D. Abercrombie, Elements of General Phonetics, Edinburgh 1967, S. 161.

⁸⁰Angesichts der in den Produktionstests zutage getretenen Unsicherheiten der Vpp bei der Kategorisierung von /a/ muß allerdings gefragt werden, ob in beiden Perzeptionstests /a/ überhaupt korrekt kategorisiert worden ist und welcher Fehlerbefund sich ergeben hätte, wenn in den individuellen Testwortreihen nicht /e/ und /a/, sondern /æ/ und /ʌ/ gegenübergestellt worden wären. - Mit der vollen Variationsbreite englischer Phonemrealisationen sind die Vpp nach viermonatigem Englischunterricht zweifellos noch nicht vertraut, auch wenn beide Gruppen außer mit der Realisationsnorm ihrer Lehrer vereinzelt (im Sprachlabor) mit Sprachmustern englischer Sprecher vertraut gemacht worden waren.

5.3.8. Beurteilung der Vokalrealisationen nach Akzeptabilität

Aus Abbildung 33 (s. S. 188) sind die auf die einzelnen Kategorien "gut", "(noch) akzeptabel" und "nicht akzeptabel" entfallenden Urteile für die Realisationen von /i, e, a/ in allen drei Produktionstests unmittelbar abzulesen; wie üblich liegen den Flächendiagrammen individuell ermittelte und über die jeweilige Gruppe gemittelte Leistungswerte zugrunde. Aufmerksamkeit verdienen insbesondere die verhältnismäßig geringen Leistungsunterschiede der beiden Gruppen von Vpp, sieht man von dem abweichenden Verhalten der beiden Gruppen für /a/ in Test III und IV ab, das bereits ausführlich diskutiert und begründet wurde.

Abbildung 33 legt nahe, daß für die "Schwierigkeit" der einzelnen Vokale /i/, /e/ und /a/ in Abhängigkeit von dem jeweiligen (Produktions-)Test mit Unterschieden zu rechnen ist. Anhand der individuellen Leistungswerte für die Realisationen von /i, e, a/ in den drei Tests, wie sie in den Tabellen 48 - 50 zusammengestellt sind, kann die Wirksamkeit dieses Effekts überprüft werden.

Signifikante Unterschiede in der "Schwierigkeit" von /i/, /e/ und /a/ lassen sich mit Hilfe der Rangvarianzanalyse (FRIEDMAN-Test) nur für Test IV und V nachweisen ($\hat{\chi}_n^2 = 19.39$ bzw. $21.5 > \chi_{2,0.05}^2 = 5.99$): In beiden Tests erweist sich /a/ als signifikant schwerer sowohl gegenüber /i/ als auch /e/, während für die letztgenannten keine signifikanten Unterschiede untereinander bestehen. In Test III hingegen besteht kein signifikanter Unterschied zwischen /i/, /e/ und /a/, die sich alle als ungefähr gleich schwer erweisen ($\hat{\chi}_n^2 = 4.07 < \chi_{2,0.05}^2 = 5.99$). Der statistische Nachweis unterstreicht damit erneut die Sonderstellung von Test III, der eine Kategorisierung der einzelnen Laute und Lautklassen in beträchtlichem Maße erschwert und "typische" Lernerfehler nicht in dem erwarteten Maße zutage fördert.

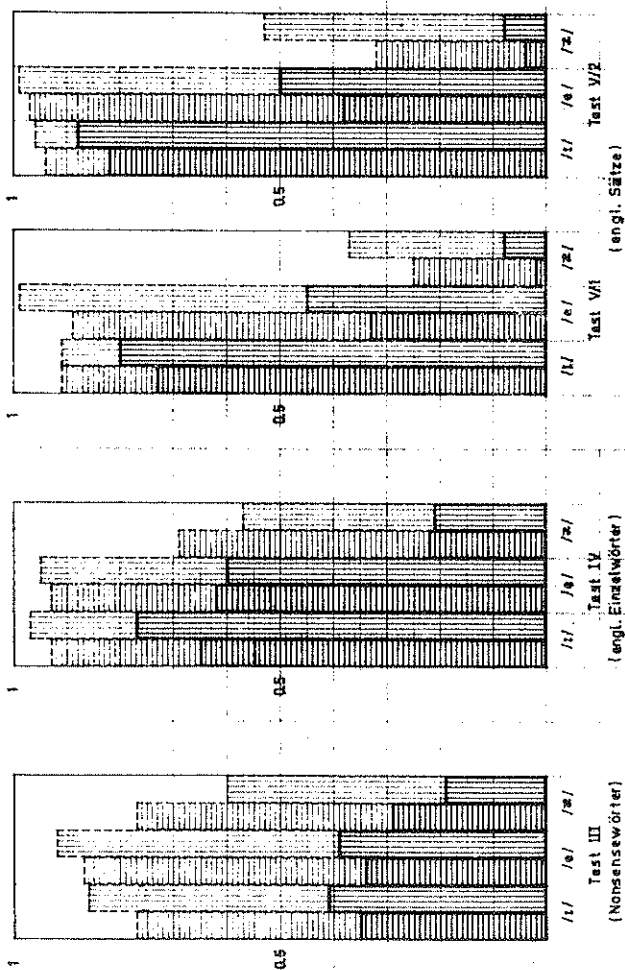


Abb. 33: Akzeptabilitätsbeurteilung der Produktionsleistungen: /i, e, a/ ("LR": ; "M":)

Gruppe	Vok.		/i/	/e/	/a/
	Vp				
"LR"	SA ₁		0.71	0.50	0.42
	SS ₂		0.42	0.38	0.54
	SS ₃		0.38	0.56	0.41
	WB		0.63	0.82	0.54
"M"	BM		0.33	0.50	0.38
	DL		0.67	0.54	0.00
	DM		0.58	0.67	0.33
	HH		0.58	0.58	0.58
	HU		0.58	0.71	0.54
	JK		1.00	0.92	0.46
MW			0.60	0.62	0.41
	StA		0.19	0.18	0.16

Tab. 48: Individuelle Leistungswerte: /i, e, a/ (Perz - Prod: Nonsenswörter)

Gruppe	Vok.		/i/	/e/	/a/
	Vp				
"LR"	SA ₁		0.80	0.65	0.50
	SS ₂		0.75	0.73	0.68
	SS ₃		0.75	0.85	0.29
	WB		0.85	0.85	0.35
"M"	BM		0.83	0.69	0.36
	DL		0.82	0.69	0.32
	DM		0.86	0.81	0.18
	HH		0.86	0.69	0.64
	HU		0.95	0.77	0.23
	JK		1.00	0.88	0.61
	RR		0.91	0.92	0.54
MW			0.85	0.77	0.41
	StA		0.08	0.09	0.18

Tab. 49: Individuelle Leistungswerte: /i, e, a/ (Perz - Prod: engl. Einzelwörter)

Gruppe	Vok.		/i/	/e/	/a/
	Vp				
"LR"	SA ₁		0.79	0.47	0.11
	SS ₂		0.90	0.64	0.31
	SS ₃		0.96	0.53	0.14
	WB		0.73	0.84	0.05
"M"	BM		0.95	0.59	0.21
	DL		0.70	0.69	0.24
	DM		0.95	0.58	0.29
	HH		0.85	0.42	0.27
	HU		0.90	0.58	0.24
	JK		0.97	0.85	0.21
	RR		0.84	0.82	0.56
	SM		0.95	0.89	0.04
MW		0.87	0.66	0.23	
StA		0.09	0.16	0.14	

Tab. 50: Individuelle Leistungswerte: /i,e,a/
(Prod - Perz - Prod: engl. Sätze)

5.3.9. Vergleich der Perzeptions- und Produktionsleistungen

Abbildung 34 stellt die Hör- und Produktionsleistungen unserer beiden Gruppen von Vpp für die Reihe der englischen palatalen Kurzvokale /i,e,a/ in der vertrauten Weise dar. Wir sehen bei dieser Darstellungsweise davon ab, daß die Schülerrealisationen von /i/, /e/ und /a/ in unterschiedlichem Maße (zumindest in Test IV und V) zu Zurückweisungen seitens der Urteiler geführt haben, da es uns hier allein auf eine Verdeutlichung des bei den Schüler(inne)n beobachteten erheblichen Leistungsunterschiedes zwischen Perzeptionsvermögen und der Produktion akzeptabler Muster von /i,e,a/ ankommt. Die Abbildung vermag diesen Unterschied eindrucksvoll darzustellen und macht vor allen Dingen deutlich, daß die Leistungen beider Gruppen in diesem Bereich nur wenig voneinander unterschieden sind. Genaueren Aufschluß vermögen hier allerdings erst die individuellen Leistungs-

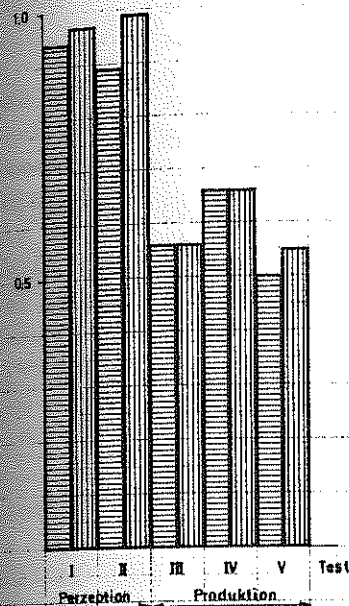


Abb. 34: Hör- und Produktionsleistungen: /i,e,a/
("LR": ■■■■, "M": ▨▨▨▨)

daten zu geben, die in Tabelle 51 zusammengestellt sind. Diesen Leistungswerten liegen zugrunde:

- Test I: 1008 (Hörer-)Urteile
- Test II: 576 (Hörer-)Urteile
- Test III: 216 Realisationspaare
- Test IV: 449 Einzelrealisationen
- Test V: 476 Zweitrealisationen

Ein Blick auf die Leistungswerte und vor allem auf die - besonders im Vergleich zum TH-Test - geringen Standardabweichungen, die einen Eindruck von den zwischen einzelnen Individuen beobachteten Leistungsunterschieden vermitteln,

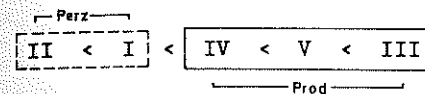
Gruppe	Test	I	II	III	IV	V
	Vp					
"LR"	SA ₁	0.90	0.90	0.54	0.64	0.46
	SS ₂	0.94	0.85	0.44	0.72	0.61
	SS ₃	0.98	0.87	0.60	0.61	0.50
	WB	0.92	0.94	0.66	0.67	0.43
"M"	BM	0.95	0.96	0.40	0.59	0.47
	DL	0.99	1.00	0.40	0.61	0.51
	DM	0.95	0.98	0.53	0.58	0.60
	HH	0.93	0.96	0.58	0.72	0.55
	HU	0.96	0.98	0.61	0.63	0.51
	JK	0.98	1.00	0.79	0.81	0.55
	RR	0.99	1.00	0.67	0.78	0.70
	SM	0.99	1.00	0.47	0.54	0.53
MW		0.96	0.95	0.56	0.66	0.54
StA		0.03	0.05	0.12	0.08	0.07

Tab. 51: Individuelle Perzeptions- und Produktionsleistungen: /i, e, ə/

zeigt deutlich, daß bei beiden Gruppen von Vpp nicht nur hohe Übereinstimmung bezüglich der Diskrepanz von Perzeption und Produktion besteht, sondern daß insbesondere die Produktion akzeptabler Realisationen von /i, e, ə/ a l l e n Schüler(inne)n in fast gleichem Maße Schwierigkeiten bereitet. Sonst klar zutage tretende Leistungsunterschiede zwischen beiden Gruppen von Vpp, wie wir sie bisher in fast allen Einzeltests nachweisen konnten, scheinen im Bereich der Vokalperzeption und -produktion keine Auswirkungen zu zeigen.

Die nichtparametrische Rangvarianzanalyse nach FRIEDMAN anhand der Leistungsdaten von Tabelle 51 ergibt die erwartete eindeutige Trennung zwischen den Perzeptions- und Produktionsaufgaben ($\hat{\chi}_r^2 = 42.62 > \chi_{4,0.05}^2 = 9.49$), wobei als Rangfolge der Schwierigkeiten unserer einzelnen "Behandlungen" die folgende Abfolge eintritt (multiple Vergleiche

nach WILCOXON-WILCOX):



Jede der drei Produktionsaufgaben unterscheidet sich somit signifikant von jeder der beiden Perzeptionsaufgaben,⁸¹ während jeweils innerhalb der Bereiche Perzeption und Produktion keine signifikanten Unterschiede in der Schwierigkeit der jeweiligen Aufgabe festzustellen sind. Die individuellen Leistungsunterschiede für /i/, /e/ und /ə/ in den drei Produktionstests, die wir bereits statistisch untersucht haben, lassen erkennen, daß der hier nachgewiesene signifikante Unterschied für die drei Vokale in unterschiedlichem Maße zu beobachten ist und /ə/ weitaus stärker betrifft als die beiden anderen Vokale. Die Produktion akzeptabler /ə/-Realisationen und die Etablierung des /e/ - /ə/-Gegensatzes in der lautsprachlichen Produktion sind damit als ein Lernproblem ersten Ranges ausgewiesen, dem um so mehr Bedeutung zukommt, als es alle der hier untersuchten Probanden in nahezu gleichem Maße betrifft und nicht davon ausgegangen werden kann, daß es sich hier (wie wir es bei TH und R annehmen) nur um ein Problem einzelner Individuen unserer Grundgesamtheit handelt.

5.4.0. FVC-Test

5.4.1. Perzeption

Das letzte unserer vier Ausspracheprobleme für deutsche Englischlernende ist das der sog. "final voiced consonants" in Englischen, das von besonderer Komplexität insofern ist, als es nicht einzelne Segmente oder Klassen von Segmenten betrifft, sondern syntagmatische, parallel enkodierte Einheiten.⁸²

⁸¹ Der Effekt ist auch bei getrennter Durchführung des Tests für jede der beiden Gruppen gesichert.

⁸² Die beteiligten phonetischen Parameter sind weiter oben, 4.2.4., S. 80-82, aufgeführt.

So häufig das hier behandelte Ausspracheproblem auch in der Literatur erwähnt wird, so unspezifisch sind durchweg die Angaben über die bei den "final voiced consonants" zu verzeichnenden Aussprachefehler, und nicht anders steht es um die didaktischen Ratschläge, die den Englischlehrern bezüglich dieser Schwierigkeit an die Hand gegeben werden.⁸³ Die vorwiegend phonologische Orientierung vieler Handbücher und Lehrwerke faßt das Problem meist als einfachen paradigmatischen Gegensatz von stimmlosen und stimmhaften Konsonanten auf und übersieht dabei die Notwendigkeit einer genauen phonetischen Spezifikation des abstrakten phonologischen Merkmals "Stimmhaftigkeit".⁸⁴

5.4.2. Fehlervorgabe

Paradigmen wie /reop/ - /reob/, /i:f/ - /i:v/, /hit/ - /hid/ liefern das Material für die beiden Perzeptions-tests zum Problem der "final voiced consonants". Die häufigen Lernerfehler deutscher Englischlernender - normalerweise ver-

⁸³ Allein in der Untersuchung von Barry (1974) werden (nach einer genauen Spezifikation der beteiligten phonetischen Parameter) spezifische Hypothesen bezüglich der bei deutschen Englischlernenden auftretenden Perzeptions- und Produktionsfehler aufgestellt und experimentell überprüft. Vgl. oben, S. 82, Anm. 30.

⁸⁴ Siehe z.B. Loggen (1972), S. 36-38, der den "elementaren Charakter" des Stimmtons betont und zur Einübung des Gegensatzes "stimmhaft-stimmlos" anhand von Summübungen und der sog. "Stimmprobe" rät: "Volle Stimmhaftigkeit verträgt sich mit dem verweilenden Rhythmus und dem Wohlklang der Kinderstimmen. Der Überschuß an Stimmtone wird später durch einen natürlichen Automatismus ausgeglichen; manipulatorischer Hilfen bedarf es zur Entsonorisierung nicht." Dem ist entgegenzuhalten, daß voll stimmhafte Realisation der (phonologisch) "stimmhaften" Endkonsonanten nicht der Realisationsnorm des Englischen entspricht; vgl. Gimson (1970), S. 152: "A phonemic transcription of rope, robe as /reop/, /reob/ is, therefore, to be interpreted as indicating that the words are distinguished not only or even primarily by difference of the final consonant, but rather by a complex of quantitative and qualitative contrasts extending over the greater part of the word." (Hervorhebung von mir).

einfachend als Nichtbeachtung dieser für das Englische wichtigen phonologischen (Stimmhaftigkeits-) Opposition bezeichnet - sind damit als Fehlervorgabe in die Diskriminations- bzw. Identifikationsaufgabe eingebaut. Als problematisch wird dabei der "markierte" (d.h. stimmhafte) Teil des Oppositionspaars angesehen, der in den einzelnen Testwortreihen das zu diskriminierende bzw. identifizierende Element abgibt.⁸⁵

5.4.3. Hördiskrimination

Die Fehlerdaten beider Gruppen von Vpp sind in Tabelle 52 zusammengestellt.⁸⁶ Die hohen Fehlerraten ("LR": 24%; "M": 15%) liegen deutlich über den für /r/ und die palatalen

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 VC ₁ ³	256	49	0.19	640	69	0.11
2 V:C ₁ ³	120	43	0.36	300	69	0.23
3 VVC ₁ ³	104	24	0.23	260	30	0.12
Σ	480	116	0.24	1200	168	0.15

Tab. 52: Hördiskrimination: FVC; Fehlerraten; Gruppe "LR" (8 Vpp); Gruppe "M" (20 Vpp)

Kurzvokale ermittelten Fehlerhäufigkeiten und liegen nahe den im TH-Test ermittelten Fehlerraten. Es ist daraus zu schließen, daß die getestete Distinktion zwischen "stimmhaften" und "stimmlosen" Endkonsonanten für die Angehörigen beider Gruppen perzeptorisch schlecht definiert ist.

Deutliche Unterschiede in der Fehleranfälligkeit sind, wie Abbildung 35 zeigt, für die einzelnen "Kontexte" der Fortis-Lenis-Distinktion zu verzeichnen. Bei beiden Gruppen erweist

⁸⁵ Siehe die Beispiele weiter oben, S. 30 und 37.

⁸⁶ Kontrollgruppe: 14 Fehler (= 4%).

sich die perzeptorische Trennung von Paaren wie $_{\epsilon}/sɪ:t/$ vs. $_{\epsilon}/sɪ:d/$ (mit phonologisch "langem" Vokal) als besonders schwierig, während Paradigmen wie $_{\epsilon}/hɪt/$ vs. $_{\epsilon}/hɪd/$ (mit phonologisch "kurzem" Vokal), aber auch wie $_{\epsilon}/rɒp/$ vs. $_{\epsilon}/rɒb/$ (mit diphthongischem Element) perzeptorisch besser unterschieden werden. Dies ist um so bemerkenswerter, als in den betreffenden Gegensatzpaaren die phonetischen Dauerunterschiede in dem dem Endkonsonanten voraufgehenden Vokal am größten sind, wenn es sich um einen Langvokal, aber am geringsten, wenn es sich um einen Kurzvokal handelt.⁸⁷ Da für die Ausprägungsart des finalen Konsonanten zwischen den beiden Kategorien "stimmhaft" und "stimmlos" nur geringe Unterschiede zu verzeichnen sind, ist eine Erklärung des Fehlerbefundes schwierig; es muß auch damit gerechnet werden,

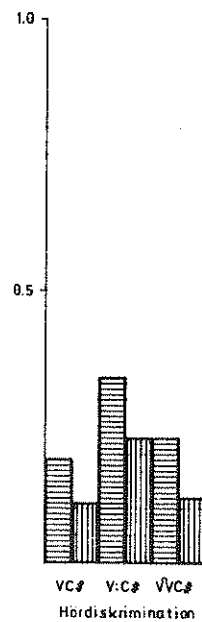


Abb. 35: Hördiskrimination: FVC; Fehlerraten in den einzelnen Testkontexten ("LR": ; "M":)

⁸⁷Vgl. Gimson (1970), S. 94-95.

daß im Falle der Kurzvokale die allophonisch bedingten Quantitätsunterschiede in den dem Endkonsonanten voraufgehenden Vokalen bei den Probanden zu einer unterschiedlichen phonematischen Kategorisierung der entsprechenden Paradigmen geführt haben in der Weise, daß die allophonisch bedingte Länge des Kurzvokals vor Lenis-Konsonant trotz unmodifizierter Qualität als (phonologische) Länge kategorisiert worden ist. Diese fehlerhafte Kategorisierung führt paradoxerweise zu (signifikant) besseren Diskriminationsleistungen als im Kontext V:C#, wo die allophonisch bedingte Variation der Quantität für die Vpp nur eine Einstufung der Vokalrealisationen als (phonologisch) lang zu ermöglichen scheint. Mit dieser Interpretation läßt sich jedoch der Fehlerbefund für den dritten Testkontext ($\sqrt{VVC\#}$) nur schwer in Einklang bringen, da Diphthonge sich bezüglich der Variation der Vokaldauer ähnlich wie die phonologisch langen Vokale verhalten.

5.4.4. Höridentifikation

Die Forderung nach Homogenität der einzelnen Blöcke eines Tests und nach gleicher Behandlung der vier untersuchten Ausspracheprobleme ist in Test II verletzt, insofern die Testwortreihen des FVC-Tests von anderer Art waren als die der übrigen Problembereiche: Anstatt der üblichen drei Distinktionen enthielten die Testwortreihen für dieses Problem nur die einfache Distinktion von "stimmhaftem" und "stimmlosem" Endkonsonant.⁸⁸ Damit ist zweifellos für dieses Problem eine Erleichterung gegenüber den anderen drei Gegenstandsbereichen gegeben, die bei den Angehörigen der Gruppe "M" zu einer deutlichen Leistungsverbesserung für diese kritische Distinktion gegenüber Test I geführt hat; die Angehörigen von "LR" weisen zwar auch hier eine Leistungsverschlechterung auf, die sich jedoch gegenüber dem sonst beobachteten Leistungsabfall eher gering ausnimmt.

⁸⁸Zur Begründung siehe weiter oben, S. 39-41.

Die Fehlerdaten beider Gruppen verzeichnet Tabelle 53.⁸⁹ Abbildung 36 zeigt, daß wie bei der Diskriminationsaufgabe Folgen von Langvokal + (auslautendem) Konsonant von beiden Gruppen häufiger fehlerhaft beurteilt wurden als $\check{V}C\#$ -Folgen;

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 $\check{V}C_1^3$	144	40	0.28	360	22	0.06
2 $V:C_1^3$	64	24	0.38	160	15	0.09
3 $\hat{V}C_1^3$	96	18	0.19	240	10	0.04
Σ	304	82	0.27	760	47	0.06

Tab.:53: Hördiskrimination: FVC; Fehlerraten; Gruppe "LR" (8 Vpp); Gruppe "M" (20 Vpp)

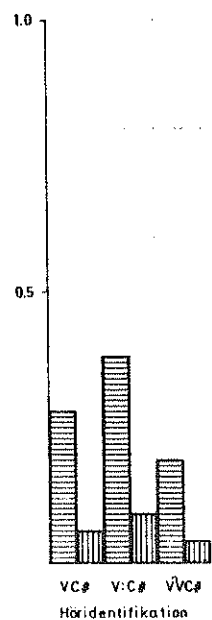


Abb. 36: Höridentifikation: FVC; Fehlerraten in den einzelnen Testkontexten ("LR": ; "M":)

⁸⁹Kontrollgruppe: kein Fehler.

noch geringer ist die relative Fehlerhäufigkeit dort, wo das dem Endkonsonanten vorausgehende Element diphthongisch ist. Der Fehlerbefund von Test I wird damit, zumindest was die höhere Fehleranfälligkeit von $V:C\#$ -Folgen angeht, erhärtet.⁹⁰ Wir stellen eine weitere Diskussion fürs erste zurück und wenden uns den Produktionsdaten zu.

5.4.5. Perzeption-Produktion: Nonsensewörter

Test III enthielt neben den "kritischen", mit stimmhaftem Konsonanten auslautenden Testwörtern auch die entsprechenden "Fortis"-Paradigmen,⁹¹ deren lautsprachliche Realisation durch die Vpp ebenfalls untersucht werden sollte; letztere werden in der Fehlerbewertung getrennt behandelt.

Tabelle 54 vereint die Fehlerdaten beider Gruppen für die "Lenis-Konsonant"-Realisationen in diesem ersten Produktionstest. Die hohen Fehlerzahlen beider Gruppen fallen unmittelbar ins Auge und demonstrieren die besonderen Schwierigkeiten,

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 $\check{V}C_1^3$	40	22	0.55	80	23	0.29
2 $V:C_1^3$	32	13	0.41	64	11	0.17
3 $\hat{V}C_1^3$	32	14	0.44	64	8	0.13
Σ	104	49	0.47	208	42	0.20

Tab. 54: Perz-Prod (Nonsensewörter): FVC; Fehlerraten in den einzelnen Testkontexten

die die Probanden bei der Kategorisierung und lautsprachlichen Realisation von stimmhaften Endkonsonanten im Englischen haben.

⁹⁰Der Befund ist auch bei Hinzuziehung der individuellen Fehlerdaten statistisch signifikant (Test I: $\chi^2 = 15.73 > 5.99$; Test II: $\chi^2 = 7.32 > 5.99$; Teststatistik: Rangvarianzanalyse nach FRIEDMAN).

⁹¹Einem Testwort /trig/ entsprach somit ein an anderer Stelle des Tests vorkommendes Testwort /trik/ usf.

Demgegenüber erweisen sich die entsprechenden Fortis-Konsonant-Realisationen zwar als weniger fehlerhaft (siehe Tabelle 55), jedoch keineswegs als so unproblematisch, wie es die phonologische Kontrastanalyse erwarten läßt.

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	PR
1 $\check{V}C_1^3$	32	6	0.19	63	2	0.03
2 $V:C_1^3$	32	4	0.13	63	3	0.05
3 $\hat{V}VC_1^3$	32	7	0.22	64	3	0.05
Σ	96	17	0.18	150	8	0.04

Tab. 55: Perz-Prod (Nonsensewörter): stimmlose Endkonsonanten; Fehlerraten in den einzelnen Testkontexten

Die in Tabelle 54 enthaltenen Fehlerdaten geben keinen vollständigen Aufschluß über das Ausmaß und die Art von Fehlern, die für den Bereich der "final voiced consonants" zu verzeichnen sind, und müssen durch eine Betrachtung der Akzeptabilitätsurteile der geschulten Beobachter ergänzt werden, so wie sie weiter unten in Abbildung 37 (S. 207) dargestellt sind. In diesem Zusammenhang ist es notwendig, die Urteilkriterien der Beobachter kritisch zu prüfen.

Die Kenntnis der spezifischen phonetischen Eigenschaften, die "stimmhafte" und "stimmlose" Endkonsonanten im Englischen voneinander unterscheiden und nicht auf den finalen Konsonanten beschränkt sind, sondern auch den vorausgehenden Vokal mit betreffen, läßt erwarten, daß Abweichungen sowohl bezüglich der Realisationsform des Endkonsonanten als auch der Dauereigenschaften des vorausgehenden Vokals eintreten können. Wie weiter oben gezeigt wurde, sind für englische Sprecher/Hörer die Dauereigenschaften des präfinalen Vokals (bzw. von V-Nasal- oder V-/l/-Verbindungen) nicht nur notwendige, sondern auch hinreichende Hörindizes für die Wahrnehmung des finalen Konsonanten. Allein Verstöße gegen die vokalischen

Dauereigenschaften können insofern bei englischsprachigen Hörern zu Verwechslungen beim finalen Konsonanten führen, sofern in der lautsprachlichen Produktion deutscher Englischlernender eine Unterscheidung von phonologischer und allophonischer Dauer nicht gegeben ist und (phonologisch) "lange" Vokale stets (phonetisch) lang, "kurze" Vokale hingegen stets kurz realisiert werden.

Die Wichtigkeit der im Vokal enkodierten Dauerinformation für die Wahrnehmung der Fortis-Lenis-Opposition steht außer Zweifel; das schließt jedoch nicht aus, daß bei der Beurteilung der FVC-Realisationen unserer Probanden auch die Realisationsform des finalen Konsonanten selber die Urteile der geschulten Beobachter beeinflusst hat: Die Parameterurteile geben Aufschluß darüber, daß in einer nicht geringen Anzahl von Fällen die vokalischen Dauereigenschaften korrekt realisiert wurden, daß jedoch eine zu starke Verschluslösung (hohe Intensität des Lösungsgeräuschs) und zu lange Aspirationsdauer zur Zurückweisung der entsprechenden Realisationen finaler "stimmhafter" Konsonanten geführt hat.

Bei einem derart komplexen Ineinandewirken verschiedener phonetischer Parameter und ihrer unterschiedlichen Bedeutung für die Perzeption ist nicht zu erwarten, daß die auditive Beurteilung letzte Aufschlüsse darüber zu geben vermag, welche Verstöße in welchen der beteiligten Parameter zur Einstufung von Äußerungen als "nicht akzeptabel" oder "(noch) akzeptabel" geführt haben: In der großen Mehrzahl dieser Fälle notierten die phonetisch geschulten Beobachter gleichzeitige Verstöße sowohl gegen die Vokaldauer als auch gegen die zu fordernde Realisationsnorm des Endkonsonanten; in allen anderen Fällen wurden Verstöße gegen nur einen dieser Parameter genannt. Das Maß an Beurteiler-Übereinstimmung ist allerdings nicht sehr hoch, was die genaue parametrische Spezifikation anbetrifft. Offensichtlich sind der auditiven Beurteilung hier Grenzen gesetzt und genauere Aufschlüsse nur von gezielten Untersuchungen sowohl im artikulatorischen wie auch im akustischen (Signal-) Bereich zu erwarten.

Die Ergebnisse des ersten Produktionstests (mit Nonsensewörtern) sind insofern aufschlußreich, als die Realisation "stimmhafter" Endkonsonanten in den drei Testkontexten, die durch die (phonologische) Spezifikation des präfinalen Vokals (bzw. Diphthongs) gegeben sind, in unterschiedlichem Maße Schwierigkeiten zu bereiten scheint. Der Befund der beiden Hörtests wird dabei in sein Gegenteil verkehrt, insofern hier bei beiden Gruppen niedrigere Fehlerraten für den Kontext V:C# zu verzeichnen sind als für VC#. Dies hat seinen Grund in einem Fehlerverhalten, das die oben gegebene Interpretation der Hörtestdaten in gewissem Umfang bestätigt.

Geht man davon aus, daß deutsche Sprecher bei der Produktion von Vokalen nur zwei undifferenzierte Quantitätskategorien zu realisieren haben und daß im Deutschen die Spezifikation der Vokalqualität die der Quantität in nahezu allen Fällen mit einschließt,⁹² muß mit Lernschwierigkeiten beim Erwerb neuer Aussprachegewohnheiten gerechnet werden, wo (phonologische) Quantität und (phonetische) Dauer sich, wie das bei der Stimmhaftigkeitsopposition im Auslaut englischer Wörter der Fall ist, in spezifischen Kontexten nicht entsprechen. Unsere Fehlerdaten entsprechen diesen Erwartungen: Die Vpp neigen dazu, die "kurzen" Vokale des Englischen auch dort kurz zu realisieren, wo sie im Kontext "stimmhafter" Endkonsonanten von größerer phonetischer Dauer sind, ohne daß ihre Qualität sich ändert. Solche Realisationen werden von englischen Hörern als Verletzung der Stimmbeteiligungseigenschaften des finalen Konsonanten interpretiert. Vereinzelt sind

⁹²Der Quantitätsunterschied "kurz" - "lang" ist nur bei *ɔ/ɛ/* gegenüber *ɔ/ɛ:/* und bei *ɔ/a/* gegenüber *ɔ/a:/* phonologisch relevant. Siehe G. Ungeheuer, "Das Phonemsystem der deutschen Hochlautung", in Siebs (1969), S. 27-42 (34). - Zur Phonologie des Deutschen allgemein und zur phonologischen Interpretation der Vokalquantität im Deutschen siehe auch O. Werner, Phonemik des Deutschen (= Sammlung Metzler, Bd. 109), Stuttgart 1972.

jedoch in den Realisationen unserer Vpp auch Dauerabweichungen in der anderen Richtung (größerer Dauer) belegt: Realisationen des Testwortes /trɪg/ als [tɹɪ:g]⁹³ lassen darauf schließen, daß hier die perzeptorische Kategorisierung anhand der phonetischen Dauereigenschaften vorgenommen worden ist und die Vpp nach ihren Gewohnheiten die vermeintliche phonologische Kategorie reproduzieren, wobei die Vokalqualität mißachtet wird. Solche Äußerungen stellen auch für die Auswertung ein besonderes Problem dar, weil die vokalischen Dauereigenschaften wie auch die Realisationsform der finalen Konsonanten eine Einstufung der Testwortrealisation als "gut" oder "(noch)akzeptabel" ermöglichen können. Selbst dort, wo die Stimmbeteiligungseigenschaften korrekt realisiert wurden, darf die von den Probanden vorgenommene fehlerhafte Kategorisierung der Testwörter in der Fehleraufstellung nicht unberücksichtigt bleiben: Realisationen wie [tɹɪ:g] (als Reflex des Stimuluswortes /trɪg/) mußten deshalb auch dann als fehlerhaft bezeichnet werden, wenn die vokalischen Dauereigenschaften wie auch die Intensität der konsonantischen Verschlusslösung und Aspirationsdauer voll akzeptabel waren.

Eine geringe Anzahl von Realisationen von Testwörtern mit finalelem Plosiv weist volle Stimmhaftigkeit des Endkonsonanten (mit stimmhafter Lösung des Verschlusses) auf, wie sie in keiner einzigen der Modellwortrealisationen unseres Informanten anzutreffen ist. Solche Realisationen,⁹⁴ die sich gehäuft bei einzelnen Vpp finden, lassen erkennen, daß die für die Wahrnehmung der Stimmbeteiligungseigenschaften des finalen Konsonanten relevanten Dauereigenschaften des präfinalen Vokals richtig dekodiert wurden; bei der Umsetzung in die

⁹³Ähnlich auch [tɹɛ'ɪg] für /trɛg/ und - besonders häufig - [tɹɛ'ɪg] für /trɛg/. Diese Wörter fungierten jedoch nicht als Testwörter im FVC-Test.

⁹⁴Im Einklang mit den weiter oben, 4.3.2., S. 84-85, dargelegten Beurteilungskriterien konnten diese Realisationen, die nicht der Realisationsnorm des Englischen entsprechen, nur als "(noch) akzeptabel" eingestuft werden.

lautsprachliche Produktion jedoch greifen die Vpp zu anderen artikulatorischen Mustern, die in den Modellrealisationen keinerlei Entsprechung haben und als im Unterricht erworben angesehen werden müssen. Hier wäre eine genaue Kenntnis des Lehrervorbildes erforderlich.

5.4.6. Perzeption-Produktion: englische Einzelwörter

Anders als bei sämtlichen bisher untersuchten Problemlauten liefert die Einführung neuen Testmaterials in Test IV, das aus den Vpp bekannten englischen Einzelwörtern besteht, im allgemeinen keine deutliche Verbesserung der Ergebnisse für die "stimmhaften" Endkonsonanten des Englischen. Die Fehlerdaten für diesen Test sind in Tabelle 56 zusammengetragen. Auffällig ist die signifikante Verbesserung der Produktionsleistungen für die "stimmhaften" Endkonsonanten nach

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 $\check{V}C_1^3$	56	17	0.30	114	15	0.13
2 $V:C_1^3$	11	5	0.45	24	7	0.29
3 $\check{V}VC_1^3$	24	9	0.38	48	11	0.23
Σ	91	31	0.34	186	33	0.18

Tab. 56: Perz-Prod (engl. Einzelwörter) : FVC; Fehlerraten in den einzelnen Testkontexten

(phonologisch) kurzem Vokal. Die auf der phonologischen Ebene kategorisierbaren auditiv dargebotenen Stimuluswörter werden deutlich besser reproduziert als die Nonsensewörter von Test III. Nur einige wenige Realisationen der Testwörter leg und head als [le'g] bzw. [he:g] (mit einem Langvokal wie in dt. leg'!) und [he'd] lassen darauf schließen, daß diese Stimuli nicht anhand der Vokalqualität kategorisiert wurden, sondern aufgrund ihrer (allophonisch bedingten) Dauereigenschaften: In diesen Fällen resultiert die falsche perzeptorische Kategorisierung auf der Produktionsseite in korrekten Dauerreali-

sationen bei gleichzeitigem Verstoß gegen die Vokalqualität. Solche Realisationen sind vielleicht solchen Identifikationsmängeln oder -fehlern vergleichbar, wie sie in diesem zweiten Produktionstest auch für die anderen Problemlaute häufiger beobachtet wurden.

Einzelne Vpp - die gleichen, die diese Eigenart auch in Test III aufwiesen - realisieren die (phonologisch) "stimmhaften" Endkonsonanten auch phonetisch voll stimmhaft, obwohl die Beispielrealisationen unseres Informanten in keinem einzigen Falle solche Realisationen enthalten.

5.4.7. Produktion-Perzeption-Produktion: englische Sätze

Die Fehlerdaten für den letzten Test mit bekanntem englischen Material, das zudem in verschrifteter Form vorlag, stellt Tabelle 57 zusammen. Beide Gruppen zeigen insofern ein übereinstimmendes Verhalten, als die Schwierigkeiten,

Kontext	"LR"			"M"		
	TW	FZ	FR	TW	FZ	FR
1 $\check{V}C_1^3$	82	26	0.32	159	19	0.12
2 $V:C_1^3$	11	7	0.64	24	3	0.13
3 $\check{V}VC_1^3$	32	9	0.28	64	7	0.11
Σ	125	42	0.34	247	29	0.12

Tab. 57: Prod-Perz-Prod (engl. Sätze): FVC; Fehlerraten in den einzelnen Testkontexten

akzeptable Realisationen des "stimmhaften" Endkonsonanten zu produzieren, trotz des unterschiedlichen Testmaterials und der Beseitigung aller Schwierigkeiten, die durch eine ausschließlich auditive Darbietung der Teststimuli gegeben sind, in fast vollem Maße weiterbestehen: Nur ein Viertel der Realisationen der Angehörigen von "LR" und etwas mehr als die Hälfte (55%) der Äußerungen von "M" wurden von den Beurteilern als "gut" bzw. "nicht abweichend" gewertet,

womit gegenüber Test III und IV nur eine geringe Leistungssteigerung zu verzeichnen ist (vgl. die Darstellung der Akzeptabilitätsurteile in Abbildung 37, S. 207): Die hier untersuchte Ausspracheschwierigkeit des Englischen für Deutsche ist durchgängig für alle der von uns durchgeführten Hör- und Produktionstests zu beobachten. Auch die Kenntnis der zugrundeliegenden Vokalphoneme schließt in keinem Falle eine Kenntnis jener allophonischen Prozesse mit ein, denen die Vokale im Kontext stimmhafter Endkonsonanten unterworfen sind.

5.4.8. Beurteilung nach Akzeptabilität

Abbildung 37 (s.S. 207) stellt die Akzeptabilitätsurteile der phonetisch geschulten Beobachter in Form von Flächen- und Balkendiagrammen dar. Es erweist sich erneut, wie wichtig es ist, für die Beurteilung von Lernerrealisationen in einem Bereich wie diesem, in dem nur die Hälfte (oder gar weniger) der Äußerungen kategorisch als "gut" oder "nicht akzeptabel" eingestuft werden können, über ein abgestuftes Beurteilungskriterium zu verfügen, das dem "Übergangscharakter" vieler Realisationen gerecht wird.

Eine graphische Darstellung wie in Abbildung 37 gibt jedoch allein keine verlässlichen Aufschlüsse über systematische Fehlertendenzen in den FVC-Realisationen unserer Probanden in den einzelnen Testkontexten; diese Tendenzen müssen stets im Vergleich der individuellen Leistungswerte der einzelnen Vpp untersucht und auf signifikante Unterschiede hin geprüft werden.

Die individuellen Leistungswerte unserer 12 Vpp für die drei Produktionstests sind in den Tabellen 58 - 60 zusammengestellt.

Die statistische Prüfung (Rangvarianzanalyse nach FRIEDMAN) ergibt nur für Test III (Nonsensewörter) einen signifikanten Befund ($\chi^2_n = 7.87 > \chi^2_{1,0.05} = 5.99$), wobei sich der Kontext VC gegenüber VVC als signifikant schwerer erweist (Multiple

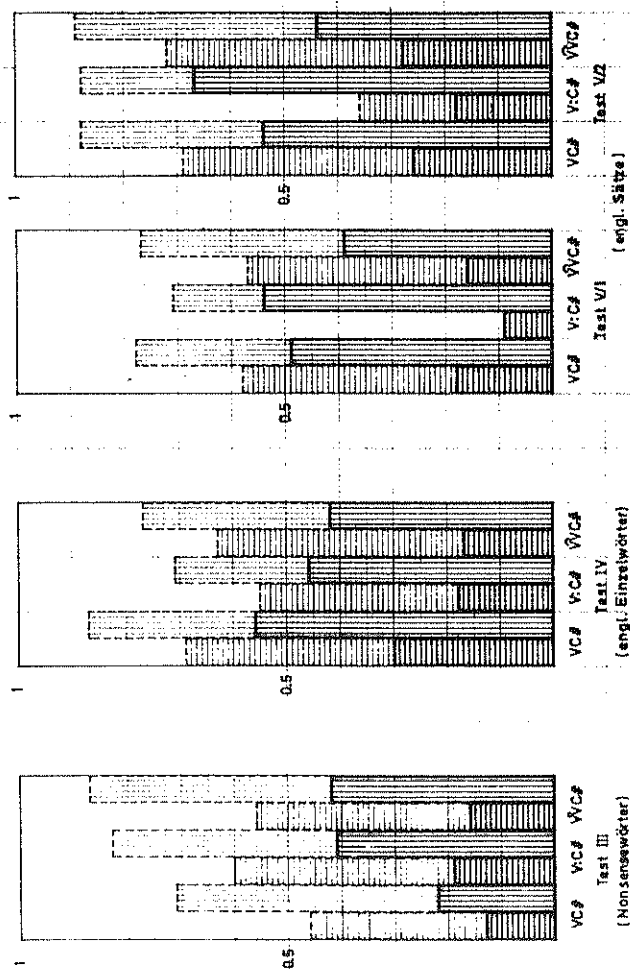


Abb. 37: Akzeptabilitätsbeurteilung der Produktionsleistungen: FVC ("LR": [diagonal lines]; "M": [horizontal lines])

Gruppe	Kontext Vp	$\bar{V}C$	V:C	$\hat{V}C$
"LR"	SA ₁	0.05	0.31	0.38
	SS ₂	0.30	0.44	0.44
	SS ₃	0.25	0.69	0.31
	WB	0.40	0.13	0.31
"M"	BM	0.50	0.38	0.44
	DL	0.60	0.69	0.75
	DM	0.25	0.25	0.50
	HH	0.50	0.75	0.63
	HU	0.55	0.81	0.81
	JK	0.50	0.69	0.81
	RR	0.40	0.63	0.88
SM	0.30	0.75	0.38	
MW		0.38	0.54	0.55
StA		0.16	0.23	0.21

Tab. 58: Individuelle Leistungswerte: FVC
(Perz - Prod: Nonsensewörter)

Gruppe	Kontext Vp	$\bar{V}C$	V:C	$\hat{V}C$
"LR"	SA ₁	0.39	0.00	0.50
	SS ₂	0.54	0.67	0.42
	SS ₃	0.61	0.50	0.42
	WB	0.46	0.50	0.25
"M"	BM	0.83	0.50	0.75
	DL	0.71	0.50	0.67
	DM	0.29	0.50	0.33
	HH	0.89	0.50	0.58
	HU	0.63	0.50	0.42
	JK	0.96	0.67	0.83
	RR	0.61	0.83	0.67
SM	0.79	0.67	0.75	
MW		0.64	0.53	0.55
StA		0.20	0.20	0.19

Tab. 59: Individuelle Leistungswerte: FVC
(Perz - Prod: engl. Einzelwörter)

Gruppe	Kontext Vp	$\bar{V}C$	V:C	$\hat{V}C$
"LR"	SA ₁	0.35	0.08	0.47
	SS ₂	0.46	0.17	0.56
	SS ₃	0.49	0.50	0.41
	WB	0.38	0.08	0.28
"M"	BM	0.61	0.75	0.53
	DL	0.08	0.92	0.75
	DM	0.33	0.58	0.31
	HH	0.86	0.75	0.88
	HU	0.68	0.67	0.84
	JK	0.89	0.75	0.72
	RR	0.71	0.67	0.50
SM	0.48	0.50	0.44	
MW		0.59	0.54	0.56
StA		0.20	0.28	0.20

Tab. 60: Individuelle Leistungswerte: FVC
(Prod - Perz - Prod: engl. Sätze)

Vergleiche nach WILCOXON-WILCOX).

Bei Test IV und V hingegen lassen sich keine signifikanten Unterschiede in der Schwierigkeit spezifischer Testkontexte der Fortis-Lenis-Distinktion nachweisen.⁹⁵

Aufmerksam zu machen ist auf die durchweg hohen Werte für die Standardabweichungen in Tab. 58 - 60, die auf beträchtliche interindividuelle Leistungsunterschiede verweisen. Sie gehen vor allem auf das Abschneiden der Angehörigen von "LR", aber auch einzelner Mitglieder von "M" (siehe die Daten für DM und SM) zurück.

Das Bild von der Akzeptabilität der Produktionsleistungen, wie Abbildung 37 es zeichnet, zeigt einen Sachverhalt, wie er für keines der bislang untersuchten Ausspracheprobleme in dieser Deutlichkeit zu verzeichnen war - nämlich durchgehend

⁹⁵Test IV: $\hat{\chi}_n^2 = 3.11 < 5.99$; Test V: $\hat{\chi}_n^2 = 1.44 < 5.99$.

schlechte Produktionsleistungen auch nach Einführung neuen Testmaterials aus dem unmittelbaren Erfahrungsbereich der Schüler, bei dem jedoch die typischen Kategorisierungsfehler deutscher Englischlernender für die "stimmhaften" Endkonsonanten in beträchtlichem Maße fortbestehen und auf der Ebene der lautsprachlichen Produktion die qualitätsunabhängigen Dauermodifikationen nicht in einem als akzeptabel zu bezeichnenden Maße vorgenommen werden.

5.4.9. Vergleich der Perzeptions- und Produktionsleistungen

Abbildung 38 stellt die Perzeptions- und Produktionsleistungen unserer 12 Vpp für die "stimmhaften" Endkonsonanten des Englischen in der üblichen Weise dar und führt die bestehenden Unterschiede deutlich vor Augen. Deutlich werden vor allem auch die beträchtlichen Leistungsunterschiede beider Gruppen für diesen Aussprachebereich; inwieweit sie auf Unterschiede in der "Güte" des Lehrermodells zurückgeführt werden können, kann nicht beantwortet werden.

Näheren Aufschluß über Unterschiede in der Schwierigkeit der in den einzelnen Tests gestellten Aufgaben erwarten wir von der statistischen Analyse der individuellen Leistungswerte beider Gruppen, die Tabelle 61 verzeichnet. Diesen über alle drei Testkontexte der Fortis-Lenis-Opposition gemittelten Leistungswerten liegen zugrunde:

- Test I: 1680 (Hörer-)Urteile
- Test II: 1064 (Hörer-)Urteile
- Test III: 156 Realisationspaare
- Test IV: 277 Einzelwortrealisationen
- Test V: 372 Zweitrealisationen

Die Anwendung des FRIEDMAN-Tests auf die Leistungsdaten von Tabelle 61 ergibt einen (hoch-)signifikanten Befund ($\chi^2 = 41.57 > \chi^2_{n,0.05} = 9.49$) für die unterschiedliche Wirksamkeit der einzelnen "Behandlungen" (= Tests) auf die Leistungen der Probanden,

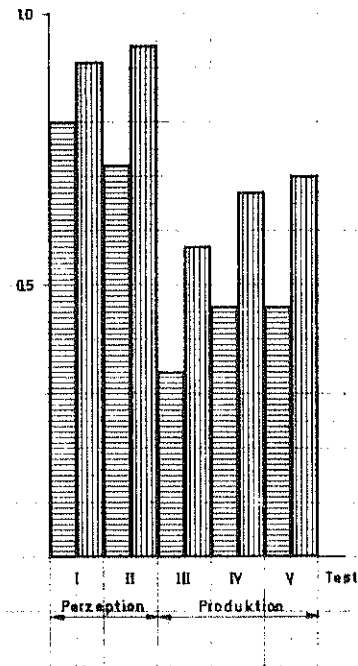
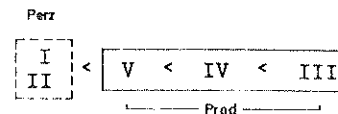


Abb. 38: Hör- und Produktionsleistungen: FVC ("LR": [diagonal lines]; "M": [vertical lines])

wobei sich eine genau so eindeutige Trennung von Perzeption und Produktion ergibt wie bei unserem Vokaltest. Mithilfe der multiplen Vergleiche nach WILCOXON-WILCOX erhält man als Rangfolge der Schwierigkeiten der einzelnen Tests die Abfolge:



Jede der drei Produktionsaufgaben unterscheidet sich somit signifikant von jeder der beiden Perzeptionsaufgaben, während jeweils innerhalb der Bereiche Perzeption und Produktion keine signifikanten Unterschiede in der Schwierigkeit der jeweiligen Aufgabe festzustellen sind. Die durchgängige Schwierigkeit einer akzeptablen lautsprachlichen Realisation "stimmhafter"

Gruppe	Test	Vp				
		I	II	III	IV	V
"LR"	SA ₁	0.72	0.61	0.23	0.37	0.42
	SS ₂	0.83	0.74	0.38	0.52	0.55
	SS ₃	0.82	0.68	0.40	0.55	0.50
	WB	0.82	0.84	0.29	0.41	0.37
"M"	BM	0.90	0.97	0.44	0.77	0.60
	DL	0.88	1.00	0.67	0.67	0.81
	DM	0.80	0.79	0.33	0.33	0.35
	HH	0.90	0.95	0.62	0.76	0.88
	HU	0.95	0.95	0.71	0.56	0.78
	JK	0.95	0.95	0.65	0.83	0.89
	RR	0.93	0.95	0.62	0.65	0.74
SM	0.93	0.92	0.46	0.76	0.57	
MW		0.87	0.86	0.48	0.60	0.62
StA		0.07	0.13	0.16	0.17	0.19

Tab. 61: Individuelle Perzeptions- und Produktionsleistungen: FVC

Endkonsonanten im Englischen für deutsche Lerner, die sich durch alle drei Produktionstests (unabhängig von der Art und spezifischen Darbietungsform des Testmaterials) hindurchzieht, ist damit auch statistisch nachgewiesen.

6.0.0. Hierarchie von Lernschwierigkeiten

Die in einer Batterie von Hör- und Produktionstests gesammelten Fehlerdaten, die im vorausgegangenen Kapitel für die einzelnen Problembereiche dieser Untersuchung getrennt dargestellt und analysiert wurden, werden in diesem Kapitel miteinander verglichen mit dem Zweck, eine empirisch gesicherte Hierarchie von Lernschwierigkeiten aufzustellen. Wir betrachten dabei die Bereiche Perzeption und Produktion zunächst getrennt.

6.1.0. Perzeptionsschwierigkeiten

Die Perzeptionsleistungen unserer zwölf Probanden, die an sämtlichen Perzeptions- und Produktionstests teilnahmen, stellt die im Anhang, S. 238, beigefügte Tabelle VI dar.¹ Der FRIEDMAN-Test ist ein geeignetes Verfahren, anhand dieser individuellen Fehlerdaten signifikante Unterschiede in der Schwierigkeit der einzelnen Probleme aufzudecken. Die Rangvarianzanalyse ergibt bei beiden Perzeptionstests (hoch-)signifikante Befunde (Test I: $\hat{\chi}_n^2 = 34.56 > \chi_{n,0.05}^2 = 9.49$; Test II: $\hat{\chi}_n^2 = 26.86 > \chi^2 = 9.49$).² Als Rangfolge der Perzeptionsschwierigkeiten ergibt sich mithilfe der multiplen Vergleiche nach WILCOXON-WILCOX für Test I die Abfolge

$$V < R < FVC < TH(\theta) < TH(\delta)$$

d.h. die Diskrimination der palatalen Kurzvokale /i, e, ə/ und des /r/ erweist sich als signifikant leichter als die Diskrimination der interdentalen Frikative und der "Stimmhaftigkeitsopposition" im Auslaut.

¹Zugrunde liegen die Leistungswerte für diejenigen phonetischen Kontexte der einzelnen Problemlaute, die allen Tests gemeinsam sind (siehe Tab. 6, S. 52): Nur so ist eine Vergleichbarkeit der Hör- und Produktionsleistungen gegeben.

²Der Effekt ist auch bei getrennter Durchführung der Tests für beide Gruppen vorhanden, ebenfalls bei Berücksichtigung aller 26 Vpp ("LP" und "M"), die an allen Hörtests teilgenommen haben.

Für Test II ergibt der gleiche Test die Abfolge

$$\boxed{V < R < FVC} < \boxed{TH(\delta) < TH(\theta)},$$

wobei der Wechsel von FVC in den Block der leicht zu identifizierenden Problemlaute eindeutig auf die Vereinfachung der Identifikationsaufgabe für diese Distinktion in Test II zurückzuführen ist.

Der Befund ist besonders eindeutig, was die besondere Schwierigkeit der Perzeption von TH und die Leichtigkeit der Vokalidentifikation und -diskrimination anbetrifft. Es ist aber darauf hinzuweisen, daß eine statistische Analyse wie die hier durchgeführte nur über die relativen Schwierigkeiten der einzelnen Problemlaute im Vergleich untereinander Auskunft gibt; in diesem Zusammenhang gilt es zu bedenken, daß die Perzeptionsleistungen unserer Probanden im allgemeinen angesichts der beträchtlichen Schwierigkeit der Aufgabe hervorragend genannt werden müssen. Hörleistungen, wie sie die "beste" Vp der Gruppe "M" aufzuweisen hat (WV/"M" mit einer Fehlerrate von nur 5%, womit diese Vp ein besseres Ergebnis erzielt als eine der Angehörigen der Kontrollgruppe: CJ/NSp), müssen als erstaunlich bezeichnet werden, und in dieses Urteil ist die große Mehrheit aller Vpp einzubeziehen.

6.2.0. Produktionsschwierigkeiten

Die Produktionsleistungen unserer Probanden in den drei Produktionstests sind für jedes einzelne Problem in der im Anhang, S. 239, beigefügten Tabelle VII zusammengestellt. Wir wenden den gleichen statistischen Test wie bei der Untersuchung der Perzeptionsschwierigkeiten an, um eine Hierarchie von Lernschwierigkeiten zu ermitteln, wie sie aus den individuellen Leistungsdaten empirisch gewonnen wurde. Die Rangvarianzanalyse (FRIEDMAN-Test) ergibt signifikante Befunde für jeden einzelnen der drei Tests (Test III: $\hat{\chi}_a^2 = 15.24 > 9.49$; Test IV: $\hat{\chi}_a^2 = 25.29 > 9.49$; Test V: $\hat{\chi}_a^2 = 29.98 > 9.49$); als Rangfolge

der Schwierigkeiten der getesteten Ausspracheprobleme erhält man:

$$\text{Test III: } \boxed{R < TH(\delta) < TH(\theta)} < \boxed{V < FVC},$$

$$\text{Test IV: } \boxed{TH(\theta) < R} < \boxed{TH(\delta) < V < FVC},$$

$$\text{Test V: } \boxed{TH(\delta) < R < TH(\theta)} < \boxed{FVC < V}.$$

In allen drei Produktionstests erweisen sich damit die palatalen Kurzvokale (und von diesen in ganz besonderem Maße /ø/) und die "stimmhaften" Endkonsonanten als signifikant schwerer als r/r ; der für die einzelnen Tests ermittelte Befund weist zudem darauf hin, daß auch TH in der Produktion keine echten Schwierigkeiten aufwirft. Dies gilt in besonderem Maße dort, wo eine eindeutige Kategorisierung der betreffenden /ø,δ/ möglich ist.

Für jeden einzelnen der vier untersuchten Problemlaute ergibt sich somit ein unterschiedliches Bild bezüglich des Zusammenhanges von Perzeptions- und Produktionsleistungen, wie die folgende schematische Abbildung zeigt:

		"leicht" ——— "schwer"
TH	Perz Prod	X X
R	Perz Prod	X X
V	Perz Prod	X X
FVC	Perz Prod	X X

Abb. 39: Zusammenhang von Perzeptions- und Produktionsleistungen: TH, R, V, FVC

Hörtests der von uns durchgeführten Art erweisen sich damit weitgehend als ungeeignete Instrumente, tatsächliche Lern-

schwierigkeiten, die allein in den lautsprachlichen Realisationen der Lerner diagnostiziert werden können, vorauszusagen. Allein im Falle der durch einen ganzen Komplex phonetischer Parameter charakterisierten Fortis-Lenis-Opposition, die perzeptorisch als "schwierig" (wenngleich nur in Relation zu den anderen Testproblemen) ausgewiesen wurde, entspricht dem Fehlerbefund der Hörtests ein ähnlicher Fehlerbefund im Bereich der lautsprachlichen Realisationen der betreffenden Distinktion. Für den Bereich der Vokale ist das genaue Gegenteil zu konstatieren: Die hervorragenden Perzeptionsleistungen aller Angehörigen der von uns getesteten Gruppen lassen die ebenfalls alle Schüler in nahezu gleichem Ausmaß betreffenden Produktionsschwierigkeiten nicht erwarten. - TH nimmt insofern eine Sonderstellung ein, als die den Probanden abverlangte Beurteilung nur auditiv dargebotener Stimuli (bei denen es sich in beiden Perzeptionstests auch noch um schwer kategorisierbare Nonsensewörter handelte) auch von geschulten Beobachtern und "native speakers" schwer vorzunehmen ist; die Testsituation enthält den Vpp visuelle Information über die Aktivität der äußeren Artikulatoren vor, die in natürlichen Kommunikationssituationen und auch in der Situation des Fremdsprachenunterrichts stets zugänglich ist. An diesen durch die Darbietungsform der Tests bedingten Schwierigkeiten gemessen, erscheinen die Perzeptionsleistungen der Probanden in einem weitaus besseren Licht, als es die relativ hohe Fehleranfälligkeit des "kritischen" TH in unseren Tests suggeriert.

6.3.0. Perzeption und Produktion:

Zusammenhänge auf individueller Ebene

Für den Bereich der vier von uns untersuchten Aussprache-probleme ist nachgewiesen, daß die Güte der Hörleistungen keinerlei Voraussagen gestattet über die Akzeptabilität der Produktionsleistungen, wenngleich für jeden der vier Test-bereiche Unterschiede zu verzeichnen sind. Allenfalls im Falle

der "final voiced consonants" entsprechen den (relativen) Diskriminations- und Identifikationsschwierigkeiten, wie sie in beiden Hörtests ermittelt wurden, Schwierigkeiten auf der Produktionsseite.

Es ist zu fragen, inwieweit auf individueller Ebene Hörtests der von uns durchgeführten Art als diagnostische Werkzeuge in Frage kommen, d.h. inwieweit individuelle Hörleistungen einzelner Vpp Voraussagen über die Akzeptabilität der lautsprachlichen Produktion dieser Vpp gestatten.

Eine Möglichkeit, diese Frage zu prüfen, besteht in der Anwendung der (nichtparametrischen) Rangvarianzanalyse (FRIEDMAN-Test) auf die Leistungsdaten der Tabellen VI und VII (siehe Anhang, S. 238-239), die in diesem Falle so angeordnet werden, daß die einzelnen Tests für die vier Problem-bereiche die Zeilen abgeben, die zwölf Vpp hingegen die Spalten. Die Spaltensummen ergeben sodann eine Rangfolge der Individuen und können auf Gemeinsamkeit geprüft werden.

Die Rangvarianzanalyse ergibt für die Perzeptionsdaten aller Tests (siehe Tab. VI, S. 238) einen signifikanten Befund ($\hat{\chi}_a^2 = 75.02 > \chi_{n,0.05}^2 = 19.7$); die Nullhypothese, daß alle 12 Vpp, die an sämtlichen Perzeptions- und Produktionstests teilgenommen haben, einer homogenen Grundgesamtheit entstammen, wird damit zugunsten der Alternativhypothese verworfen: Zwischen einzelnen Vpp bestehen signifikante Unterschiede in den Hörleistungen, wie sie bereits die anhand der rohen Fehlerdaten vorgenommene Cluster-Analyse (s.o., 3.7.0., S. 70-71) zutage förderte.

Mithilfe der Multiplen Vergleiche nach WILCOXON-WILCOX ergibt sich für die Rangfolge der einzelnen Vpp die Abfolge:

(Perz)

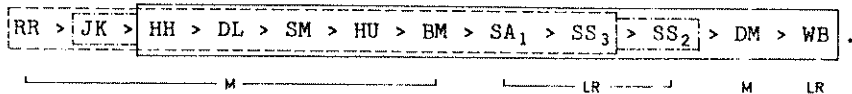


³Die Perzeptionsleistungen von BM und SS₃ unterscheiden sich nicht signifikant.

wobei sich die Perzeptionsleistungen der Probanden außerhalb der Überlappungszone der beiden "Blöcke" signifikant voneinander unterscheiden. Bemerkenswert ist dabei die nahezu saubere Trennung der beiden Gruppen ("LR" und "M"): Die Hörleistungen der Mehrheit der Gruppe "M" unterscheiden sich signifikant von den Hörleistungen der Mehrheit von "LR"; Ausnahmen bilden die Vpp WB/"LR" und DM/"M", die sich mit ihren Leistungen von den anderen Mitgliedern ihrer jeweiligen Gruppe entfernen. Jedoch sind in keinem Falle innerhalb einer der beiden Gruppen signifikante Leistungsunterschiede zu verzeichnen.

Die Durchführung des Tests anhand der für alle Problembereiche ermittelten Produktionsleistungswerte, wie sie in Tabelle VII (S. 239) zusammengestellt sind, ergibt ebenfalls einen signifikanten Befund ($\chi^2_n = 41.37 > \chi^2_{n,0.05} = 19.7$), der signifikante Leistungsunterschiede zwischen einzelnen Vpp erwarten läßt, die mithilfe der Multiplen Vergleiche nach WILCOXON-WILCOX näher zu untersuchen sind. Als Rangfolge erhält man die folgende Abfolge:

(Prod)



Nur die Produktionsleistungen der beiden "besten" Angehörigen von "M" (RR und JK) unterscheiden sich somit signifikant von den Leistungen zweier⁴ weiterer Mitglieder unserer Testpopulation. Bemerkenswerterweise handelt es sich bei den Vpp mit den schlechtesten Produktionsleistungen - DM/"M" und WB/"LR" - ausgerechnet um jene beiden, deren Perzeptionsleistungen sich nicht signifikant von denen der anderen Mitglieder der Gesamtheit unterscheiden. Damit treten anders als bei der Perzeptionsaufgabe signifikante Leistungsunter-

⁴Signifikant unterscheiden sich auch die Produktionsleistungen von RR und SS₂.

schiede auch innerhalb ein und derselben Gruppe von Vpp auf, während sich die Leistungen der großen Mehrheit der Gruppe "M" von denen der Mehrheit der Gruppe "LR" nicht signifikant voneinander unterscheiden: Für den Bereich der lautsprachlichen Produktion gilt, daß die Leistungen unserer zwölf Probanden im allgemeinen näher beieinander liegen, als dies für die Perzeptionsaufgabe festzustellen ist.

Vergleicht man darüber hinaus die Rangfolgen unserer zwölf Vpp, wie sie für die Bereiche Perzeption und Produktion getrennt ermittelt wurden, ist festzustellen, daß die Abfolge in beiden Tests eine unterschiedliche ist und daß für eine beträchtliche Anzahl von Vpp die Perzeptionsleistungen in allen untersuchten Aussprachebereichen keinerlei Rückschlüsse auf die Akzeptabilität ihrer lautsprachlichen Realisation der untersuchten Problemlaute erlauben. Besonders muß hier auf das Abschneiden von WB/"LR" und DM/"M" verwiesen werden, deren schlechte Produktionsleistungen (in allen vier Problembereichen) in krassem Widerspruch zu ihren guten Perzeptionsleistungen stehen; aufschlußreich ist in diesem Zusammenhang auch das Verhalten von SA₁/"LR" dessen Perzeptionsleistungen zu den schwächsten der Mitglieder seiner Gruppe⁵ gehören, dessen Produktionsleistungen ihm hingegen einen weitaus besseren Rangplatz sichern, als aufgrund der Hörtestleistungen hätte vorausgesagt werden können. - Da für die Mehrheit der Gruppe "M" in den Perzeptionsleistungen allgemein nur geringe Unterschiede bestehen,⁶ so daß die Zuweisung von unterschiedlichen Rangplätzen nicht völlig frei von Willkür ist, sollte den hier beobachteten Verschiebungen in der Rangfolge für die Produktionsleistungen nicht allzu viel Bedeutung beigemessen werden; signifikante Unterschiede sind für die Mehrheit dieser Gruppe in keinem der beiden Bereiche Perzeption und Produktion festzustellen.

⁵Gemeint ist hier die volle Gruppe "LR" mit acht Angehörigen, bei denen diese Vp den vorletzten Rang einnimmt: s.o., Tab. 5, S. 43.

⁶Vgl. auch oben, Tab. 5, S. 43.

6.4.0. Zusammenfassung und Schluß

Wir fassen unsere in einer Serie von Tests bei zwei Gruppen von Vpp gewonnenen Ergebnisse kurz zusammen:

1. Unsere Untersuchung hat gezeigt, daß ein allgemeiner Zusammenhang von Perzeptions- und Produktionsleistungen zumindest bei Vpp der hier untersuchten Altersstufe nicht besteht;⁷ dies gilt sowohl für die Ebene der einzelnen den Gegenstand dieser Untersuchung abgebenden Ausspracheprobleme des Englischen für Deutsche wie auch für die Ebene der individuellen Leistungen. Bei Betrachtung der Fehlertypik ist ebenfalls keine Übereinstimmung zwischen Perzeption und Produktion zu verzeichnen.
2. Über die Perzeptionsfähigkeiten der Vpp waren durch die Anwendung zweier unterschiedlicher Tests (Hördiskrimination und -identifikation) keine unterschiedlichen Aussagen zu gewinnen.
3. Hingegen ist deutlich zu erkennen, daß bei unterschiedlicher Kontextualisierung der Testlaute durch Einbettung a) in "englisch" strukturierte Nonsensewörter; b) in als bekannt vorauszusetzende englische Einzelwörter; c) in Sätze mit vertrautem Wortmaterial unterschiedliche Schwerpunkte des Produktionsfehlers gegeben sind. Bei der spontanen Nachahmung von Nonsensewörtern sind Produktionsfehler zu verzeichnen, wie sie aufgrund perceptorischer Verwechslungen zu erwarten sind. Bei Teststimuli hingegen, bei denen die korrekte Phonemidentifikation und Worterkennung gewährleistet ist, sind Produktionsfehler anderen Typs festzustellen, die auf andere Fehlerursachen hindeuten. Das Auftreten dieser Fehler in sämtlichen hier durchgeführten Produktionstests weist auf das Vorhandensein verfestigter

⁷Vgl. auch J. Berko und R. Brown, "Psycholinguistic Research Methods", in P.H. Mussen (Hrsg.), Handbook of Research in Child Development, New York 1960, S. 531.

Artikulationsmuster, die in diesem frühen Stadium des Fremdspracherwerbs die Notwendigkeit frühzeitigen korrektiven Eingreifens unterstreichen.

4. Die signifikanten Leistungsunterschiede (auf der Produktionsseite) zwischen TH und R einerseits und Vokalen und "stimmhaften" Endkonsonanten andererseits lassen erkennen, daß die Menge an leicht zugänglicher Information über die Art des fremden Lautes für die Aneignung wichtig ist. So sind für /θ,ð/ die eindeutige (sichtbare) Artikulationsstelle und für /r/ die Kombination von eindeutigem akustischem Bild und artikulatorischer Entfernung von dem (phonologisch und orthographisch) entsprechenden deutschen Laut maßgebend. Bei den "stimmhaften" Endkonsonanten sind solche deutlich hörbaren oder beschreibbaren Unterschiede nicht gegeben, und auch bei den Vokalen ist - trotz ausgezeichneter perceptorischer Trennung der Testvokale - ein Mangel an Bewußtheit von der Fremdheit der zu fordernden Aneignungsnorm zu verzeichnen. Als vielleicht wichtigster Faktor kommt hier hinzu, daß die zuletzt genannten Bereiche als Problembereiche noch nicht im selben Maße von allen Lehrern erkannt sind.

LITERATUR

- Abercrombie, D. "Phonetic Transcriptions." Le Maître Phonétique, 3^{me} série, No. 100 (1953), 32-34.
- . Elements of General Phonetics. Edinburgh 1967.
- Arnold, R. & Hansen, K. Phonetik der englischen Sprache; Eine Einführung. Leipzig 1968.
- Barry, W.J. Perzeption und Produktion im subphonemischen Bereich; Eine kontrastive Untersuchung an intersprachlichen Minimalpaaren des Deutschen und Englischen (= Linguistische Arbeiten, 15). Tübingen 1974.
- . "Language Background and the Perception of Foreign Accent." Journal of Phonetics, 2 (1974), 65-89.
- . "Parallel Encoding as a Source of Interference." Vortrag gehalten auf dem 4th Annual Phonetics Symposium, The Language Centre, University of Essex, January 1974 (erscheint in den Occasional Papers: Phonetics).
- & Gutknecht, Chr. Practical English Phonetics; A Programmed Course in English Pronunciation. 4. Aufl. Hamburg 1974.
- & Schindler, F. "Forschungstendenzen in der angewandten Phonetik." In Nickel, G. (Hrsg.). Kongreßbericht der 5. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) e.V. (= IRAL-Sonderband). Heidelberg 1974, S. 230-239.
- Berko, J. & Brown, R. "Psycholinguistic Research Methods." In Mussen, P.H. (Hrsg.). Handbook of Research in Child Development. New York 1960, S. 516-557.
- Bloch, B. "Phonemic Overlapping." In Makkai, V.B. (Hrsg.). Phonological Theory; Evolution and Current Practice. New York 1972, S. 66-70 (zuerst in American Speech, 16 (1941), 278-284).

- Bloch, B. "A Set of Postulates for Phonemic Analysis." In Makkai (1972), S. 167-199 (zuerst in Language, 24 (1948), 3-46).
- Bonnin, G.M. "Some Acoustic Aspects of Final Stop Allophones in Contemporary German." Phonetica, 11 (1964), 65-100.
- Bradley, J.V. Distribution-free Statistical Tests. Englewood Cliffs 1968.
- Brière, E.J. "Phonological Testing Reconsidered." Language Learning, 17 (1967), 163-171.
- . A Psycholinguistic Study of Phonological Interference (= Janua Linguarum, Ser. minor, 66). Den Haag 1968.
- Burgschmidt, E. & Götz, D. "Kontrastive Phonologie Deutsch-Englisch und Mundartinterferenz." Linguistik und Didaktik, 11 (1972), 209-225.
- . Kontrastive Linguistik Deutsch-Englisch; Theorie und Anwendung. München 1974.
- Corder, S.P. "The Significance of Learner's Errors." IRAL, 5 (1967), 161-170.
- . "Idiosyncratic Dialects and Error Analysis." IRAL, 9 (1971), 147-159.
- . "Zur Beschreibung der Sprache des Sprachlerner." In Nickel, G. (Hrsg.). Reader zur kontrastiven Linguistik. Frankfurt/Main 1972, S. 175-184.
- Delattre, P. "The Physiological Interpretation of Sound Spectrograms." PMLA, 66 (1959), 864-875.
- Denes, P. "Effect of Duration on the Perception of Voicing." JASA, 27 (1955), 761-764.
- Droescher, W.O. Rezension von Kufner (1971). Phonetica, 27 (1973), 111-114.
- Duden Aussprachewörterbuch. Bearbeitet von M. Mangold in Zusammenarbeit mit P. Grebe (= Der Große Duden, Bd. 6). Mannheim 1962.

- Ezawa, K. Die Opposition stimmhafter und stimmloser Verschlusslaute im Deutschen (= Tübinger Beiträge zur Linguistik 29). Tübingen 1972.
- Fant, G. "Analysis and Synthesis of Speech Processes." In Malmberg, B. (Hrsg.). Manual of Phonetics. Neudr. Amsterdam 1974, S. 173-277.
- Ferguson, Ch. "Fricatives in Child Language Acquisition." Papers and Reports on Child Language Development, Stanford University, 1973, 63-85.
- Friedrichs, H. Peter Pim and Billy Ball 1. Englisch Lehrwerk für Hauptschulen und sechsjährige Grundschulen. 26. Aufl. Berlin 1970.
- Gimson, A.C. An Introduction to the Pronunciation of English. 2nd ed. London 1970.
- Graf/Henning/Stange. Formeln und Tabellen der mathematischen Statistik. 2. Aufl. von K. Stange und H.-J. Henning. Berlin 1966.
- Greenberg, J.H. & Jenkins, J.J. "Studies in the Psychological Correlates of the Sound System of American English." Word, 20 (1964), 157-177.
- Harris, D.P. Testing English as a Second Language. New York 1969.
- Harris, K.S. "Cues for the Discrimination of American English Fricatives in Spoken Syllables." Lang. & Speech, 1 (1958), 1-7.
- Haugen, E. "The Phoneme in Bilingual Description." Language Learning, 7 (1956-57), 3-4; 17-23.
- Hawkins, S. "Temporal Coordination of Consonants in the Speech of Children: Preliminary Data." Journal of Phonetics, 1 (1973), 181-217.
- Heeschen, C. & Nehr, M. "Psycholinguistische Bemerkungen zur Stellung der generativen Transformationsgrammatik im Fremdsprachenunterricht." Praxis des neusprachlichen Unterrichts, 21 (1974), 3-11.

- Heike, G. Zur Phonologie der Stadtkölner Mundart; Eine experimentelle Untersuchung der akustischen Unterscheidungsmerkmale (= Deutsche Dialektgeographie, Bd. 57). Marburg 1964.
- "Die Rolle der Phonetik im Ausspracheunterricht." In Nickel, G. & Raasch, A. (Hrsgg.). Kongreßbericht der 3. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) e.V. (= IRAL-Sonderband). Heidelberg 1972, S. 1-13.
- Heinz, J. & Stevens, K. "On the Properties of Voiceless Fricative Consonants." JASA, 33 (1961), 589-596.
- Hildebrandt, B.F.O. & Hildebrandt, L.M. "Das deutsche R; Regelhaftigkeiten in der gegenwärtigen Reduktions-Entwicklung und Anwendung im Fremdsprachenunterricht." Linguistics, 11 (1965), 5-20.
- Hughes, G.W. & Halle, M. "Spectral Properties of Fricative Consonants." JASA, 28 (1956), 303-310.
- James, C. "Zur Rechtfertigung der kontrastiven Linguistik." In Nickel (1972), S. 21-38.
- Jespersen, O. Articulation of Speech Sounds, Represented by Means of Alphabetic Symbols. Marburg 1889.
- Phonetische Grundfragen. Leipzig 1904.
- Jones, D. The History and Meaning of the Term "Phoneme". Supplement zu Le Maître Phonétique, London 1957.
- An Outline of English Phonetics. 9th ed. repr. Cambridge (Engl.) 1969.
- Everyman's English Pronouncing Dictionary. 13th ed. edited by A.C. Gimson. Repr. London 1969.
- Joos, M. Acoustic Phonetics (= Language Monograph No. 23). Baltimore 1948.
- Jørgensen, H.P. "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen." Phonetica, 19 (1969), 217-245.

- Kohler, K. "On the Adequacy of Phonological Theories for Contrastive Studies." In Nickel, G. (Hrsg.). Papers in Contrastive Linguistics. Cambridge (Engl.) 1971, S. 83-88.
- . "Anwendungsorientierte Phonetik. Sinn und Unsinn der kontrastiven Phonologie für den Fremdsprachenunterricht." Ling. Berichte, 26 (1973), 70-78.
- . "Contrastive Sentence Phonology." In Proc. III. AILA Congress (Copenhagen 1972), Vol. I. Heidelberg 1974, S. 129ff.
- Koziol, H. Die Aussprache des Englischen. Wien 1959.
- Kresheck, J.; Fisher, H.; Rutherford, D. "A Study of r-Phones in the Speech of Three-Year-Old Children." Folia Phoniatrica, 24 (1972), 301-312.
- Kufner, H.L. Strukturelle Grammatik der Münchner Stadtmundart. München 1961.
- . Kontrastive Phonologie Deutsch-Englisch. Stuttgart 1971.
- Ladefoged, P. "The Value of Phonetic Statements." Language, 36 (1960), 387-396.
- . Three Areas of Experimental Phonetics (= Language and Language Learning 15). London 1967.
- . "The Nature of Vowel Quality." In Ladefoged (1967), S. 50-142.
- . "Units in the Perception and Production of Speech." In Ladefoged (1967), S. 143-172.
- Lado, R. Linguistics Across Cultures; Applied Linguistics for Language Teachers. Ann Arbor 1957.
- Lehiste, I. (Hrsg.). Readings in Acoustic Phonetics. Cambridge, Mass. 1967.
- & Peterson, G.E. "Transitions, Glides, and Diphthongs." JASA, 33 (1961), 268-277 (wieder abgedr. in Lehiste (1967), S. 228-237).

- Liberman, A.M. "Some Characteristics of Perception in the Speech Mode." Perception and Its Disorders, 48 (1970), 238-254.
- ; Cooper, F.S.; Shankweiler, D.P.; Studdert-Kennedy, M. "Perception of the Speech Code." Psychological Review, 74 (1967), 431-461.
- Lisker, L. "Minimal Cues for Separating /w,r,l,y/ in Intervocalic Position." Word, 13 (1957), 256-267.
- Loggen, K. Didaktische Phonetik des Englischen im Studium des Sprachlehrers. Bad Heilbrunn 1972.
- Lüdtke, H. "Zur Theoriebildung in der Phonetik." Folia Linguistica, 5 (1972), 333-354.
- Makkai, V.B. (Hrsg.). Phonological Theory; Evolution and Current Practice. New York 1972.
- Malécot, A. "The Lenis-Fortis Opposition: Its Physiological Parameters." JASA, 47 (1970), 1588-1592.
- Malmberg, B. (Hrsg.). Manual of Phonetics. Neudr. Amsterdam 1974.
- Melamed, J.T. An Experiment in Sound Discrimination. Ph.D. dissertation, University of Indiana 1961.
- Menyuk, P. The Acquisition and Development of Language. Englewood Cliffs 1971.
- & Anderson, S. "Children's Identification and Reproduction of /w/, /r/ and /l/." JSHR, 12 (1965), 39-52.
- Meyer, E.A. Englische Lautdauer; Eine experimentalphonetische Untersuchung (= Skrifter utgifna af K. Humanistiska Vetenskaps-Samfundet i Uppsala VIII.3). Uppsala und Leipzig 1903.
- Meyer-Eppler, W. "Zur Spektralstruktur der /r/-Allophone des Deutschen." Acustica, 9 (1959), 246-250.
- Mohr, B. & Wang, W.S.-Y. "Perceptual Distance and the Specification of Phonological Features." Phonetica, 18 (1968), 31-45.

Moulton, W.G. The Sounds of English and German (= Contrastive Structure Series). Chicago 1962.

Mussen, P.H. (Hrsg.). Handbook of Research in Child Development. New York 1960.

Nemser, W. An Experimental Study of Phonological Interference in the English of Hungarians (= Indiana University Publications, Uralic and Altaic Series, Vol. 105). Bloomington und Den Haag 1971.

----- . "Approximative Systems of Foreign Language Learners." IRAL, 9 (1971), 115-123.

----- . "The Predictability of Interference Phenomena in the English Speech of Native Speakers of Hungarian." In Nickel (1971), S. 89-96.

----- & Slama-Cazacu, T. "A Contribution to Contrastive Linguistics (A Psycholinguistic Approach: Contact Analysis)." Revue Roumaine de Linguistique, 15 (1970), 101-128.

Nickel, G. (Hrsg.). Papers in Contrastive Linguistics. Cambridge (Engl.) 1971.

----- . (Hrsg.). Reader zur kontrastiven Linguistik. Frankfurt/Main 1972.

----- . (Hrsg.). Fehlerkunde; Beiträge zur Fehleranalyse, Fehlerbewertung und Fehlertherapie. Berlin 1972.

----- . (Hrsg.). Kongreßbericht der 5. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) e.V. (= IRAL-Sonderband). Heidelberg 1974.

----- & Raasch, A. (Hrsg.). Kongreßbericht der 3. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) e.V. (= IRAL-Sonderband). Heidelberg 1972.

O'Connor, J.D. Better English Pronunciation. Cambridge (Engl.) 1967.

O'Connor, J.D. & Trim, J.L.M. "Vowel, Consonant and Syllable - A Phonological Approach." Word, 9 (1953), 103-122.

Olsted, D.L. Out of the Mouth of Babes (= Janua Linguarum, Ser. minor, 117). Den Haag 1971.

Peterson, G.E. & Shoup, J.E. "A Physiological Theory of Phonetics." JSHR, 9 (1966), 5-67.

Pike, K.L. Phonetics; A Critical Analysis of Phonetic Theory and a Technic for the Practical Description of Sounds. Neudr. Ann Arbor 1969.

----- . Phonemics; A Technique for Reducing Languages to Writing. Neudr. Ann Arbor 1971.

Pilch, H. "Das Lautsystem der hochdeutschen Umgangssprache." Zeitschr. f. Mundartforschung, 33 (1966), 247-266.

Pimsleur, P. & Quinn, T. (Hrsgg.). The Psychology of Second Language Learning. Cambridge (Engl.) 1971.

Platt, H. "A Comparative Study of the Phonetics of Australian English and German." Phonetica, 21 (1970), 1-30; 75-106.

Politzer, R.L. "Developmental Aspects of Auditory Discrimination." In Valdman (1972), S. 425-432.

The Principles of the International Phonetic Association. Neudr. London 1971.

Raphael, L.J. "Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English." JASA, 51 (1972), 1296-1303.

----- ; Dorman, M.F.; Freeman, F.; Tobin, C. "Vowel and Nasal Duration as Cues to the Perceptual Categorization of Word-Final Consonants in English." Paper given at the Eighth International Congress on Acoustics, London 1974. Contributed Papers Vol. 1, S. 318.

- Ruth, W. Experimentalphonetische Untersuchung über die Dehnung kurzer Vokale im Standard English (= Sitzungsberichte der Akademie der Wissenschaften in Wien, phil.-hist. Klasse, 222. Bd., 1. Abhdlg.). Wien und Leipzig 1943.
- Sachs, L. Statistische Auswertungsmethoden. 3. Aufl. Berlin 1972.
- Statistische Methoden; Ein Soforthelfer für Praktiker in Naturwissenschaft, Medizin, Technik, Wirtschaft, Psychologie und Soziologie. 2. Aufl. Berlin 1972.
- Sapon, S.M. & Carroll, J.B. "Discriminative Perception of Speech Sounds as a Function of Native Language." General Linguistics, 3 (1958), 62-72.
- Schane, S.A. Generative Phonology. Englewood Cliffs 1973.
- Scherer, G. & Wollmann, A. Englische Phonetik und Phonologie. (= Grundlagen der Anglistik und Amerikanistik, Bd. 6). Berlin 1972.
- Scholes, R.J. "Phoneme Categorization of Synthetic Vowel Stimuli by Speakers of Japanese, Spanish, Persian, and American English." Lang. & Speech, 10 (1967), 46-68.
- "Categorial Responses to Synthetic Vowel Stimuli by Speakers of Various Languages." Lang. & Speech, 10 (1967), 252-273.
- Selinker, L. "The Psychologically Relevant Data of Second-Language Learning." In Pimsleur & Quinn (1971), S. 35-43.
- "Interlanguage." IRAL, 10 (1972), 209-231.
- Sharf, D.J. "Distinctiveness of 'Defective' Fricative Sounds." Lang. & Speech, 11 (1968), 38-45.
- Siebs, Th. Deutsche Aussprache; Reine und gemässigte Hochlautung mit Aussprachewörterbuch. 19. Aufl. hrsgg. v. H. de Boor, H. Moser u. Chr. Winkler. Berlin 1969.
- Smith, N.V. The Acquisition of Phonology. Cambridge (Engl.) 1973.

- Stockwell, R.P. & Bowen, J.D. The Sounds of English and Spanish (= Contrastive Structure Series). Chicago 1965.
- Strevens, P. "Spectra of Fricative Noise in Human Speech." Lang. & Speech, 3 (1960), 32-49.
- Sweet, H. Handbook of Phonetics. 1877.
- Trim, J.L.M. "Some Suggestions for the Phonetic Notation of Sounds in Defective Speech." Speech, 17 (1953), 21ff.
- Ulbrich, H. Instrumentalphonetisch-auditive R-Untersuchungen im Deutschen (= Schriften zur Phonetik, Sprachwissenschaft und Kommunikationsforschung, 13). Berlin 1972.
- Ungeheuer, G. "Das Phonemsystem der deutschen Hochlautung." in Siebs (1969), S. 27-42.
- Valdman, A. (Hrsg.). Studies in Linguistics and Phonetics to the Memory of Pierre Delattre (= Janua Linguarum, Ser. maior, 54). Den Haag 1972.
- Weinreich, U. "On the Description of Phonic Interference." Word, 13 (1957), 1-11.
- Wells, J.C. A Study of the Formants of the Pure Vowels of British English. M.A. thesis, University of London. 1962.
- Werner, O. Phonemik des Deutschen (= Slg. Metzler, Bd. 109). Stuttgart 1972.
- Winer, B.J. Statistical Principles in Experimental Design. 2. Aufl. New York 1971.
- Winitz, H. & Lawrence, M. "Children's Articulation and Sound Learning Ability." JSHR, 4 (1961), 259-268.
- Wise, C.M. Applied Phonetics. Englewood Cliffs 1957.
- Wede, P. "Englische Satzintonation." Phonetica, 15 (1966), 129-218.
- "Natürliche Zweitsprachigkeit: Probleme, Aufgaben, Perspektiven." Ling. Berichte, 62 (1974), 15-36.
- Wörterbuch der deutschen Aussprache. Hgg. v. H. Krahn u.a. 2. Aufl. Leipzig und München 1969.

ANHANG:
Tab. I - VII
Testmaterial

Gruppe	Vp	Hördiskrimination				Höridentifikation		
		Bl. 1	Bl. 2	Bl. 3	Bl. 4	Bl. 1	Bl. 2	Bl. 3
"DR"	OK	19	11	16	11	11	21	18
	PA	12	6	6	8	10	15	22
	SS ₁	25	9	16	13	17	18	27
	SA ₁	11	20	11	22	17	27	26
	SA ₂	21	19	31	21	21	19	34
	SS ₂	16	14	9	13	17	31	17
	SS ₃	21	11	12	15	12	27	29
	WB	10	9	17	7	16	15	15
"E"	BM	7	6	4	2	8	9	10
	BS	14	14	6	7	4	7	11
	BT	17	14	14	12	11	14	21
	DL	11	7	7	9	7	4	7
	DM	14	15	15	13	15	13	15
	DK	9	11	7	7	13	9	8
	FT	20	18	7	11	11	14	12
	GK	12	12	4	4	4	3	6
	HH	3	11	10	6	14	6	14
	HU	7	2	9	5	9	7	4
	JK	8	3	6	4	8	10	7
	KS	16	4	11	5	9	13	5
	KB	4	4	2	8	10	10	14
	RR	6	5	8	4	5	6	6
	ST	17	10	5	8	9	11	11
	SM	10	9	6	8	11	6	7
	WJ	16	9	6	8	9	10	4
	WF ₁	25	14	10	11	11	12	15
WV	7	3	2	3	6	1	4	
WF ₂	10	9	3	5	11	5	4	
"K"	BB	5	1	2	-	-	-	-
	HE	1	1	-	2	1	-	1
	CC	6	1	-	1	1	-	-
	KH	3	6	1	2	2	2	-
	CJ	7	3	3	4	7	5	2
	ST	4	3	-	-	-	-	1
Σ		394	294	266	259	317	350	377

Tab. I: Absolute Fehlerzahlen (Perzeptionstests)

Gruppe	Vp	Test I				Test II			
		1	2	3	4	1	2	3	
"LR"	♀	OK	25.3	14.7	21.3	14.9	16.9	31.8	25.7
		PA	16.0	8.0	8.0	10.8	15.4	22.7	31.4
		SA ₂	28.0	25.3	41.3	28.4	32.3	28.8	48.6
		SS ₂	21.3	18.7	12.0	17.6	26.2	47.0	24.3
		SS ₃	28.0	14.7	16.0	20.3	18.5	40.9	41.2
	♂	SS ₁	33.3	12.0	21.3	17.6	26.2	27.3	38.6
		SA ₁	14.7	26.7	14.7	29.7	26.2	40.9	37.1
		WB	13.3	12.0	22.7	9.5	24.6	22.7	21.4
"M"	♀	BM	9.3	5.0	5.3	2.7	12.3	13.6	14.3
		DK	12.0	14.7	9.3	9.5	20.0	13.6	11.4
		GK	16.0	16.0	5.3	5.4	6.2	4.5	8.6
		HH	4.0	14.7	13.3	8.5	21.3	9.1	20.0
		HU	9.3	2.7	12.0	6.8	13.8	10.6	5.7
		JK	10.7	4.0	8.0	5.4	12.3	15.2	10.0
		KB	5.3	5.3	2.7	10.8	15.4	15.2	20.0
		WV	9.3	4.0	2.7	4.1	9.2	1.5	5.7
	♂	BS	18.7	18.7	8.0	9.5	6.2	10.6	15.7
		BT	22.7	18.7	18.7	16.2	16.9	21.2	30.0
		DL	14.7	9.3	9.3	12.2	10.8	6.1	10.0
		DM	18.7	20.0	20.0	17.6	23.1	19.7	21.4
		FT	26.7	24.0	9.3	14.9	16.9	21.2	17.1
		KS	21.3	5.3	14.7	6.8	13.8	19.7	7.1
		RR	8.0	6.7	10.7	5.4	7.7	9.1	8.6
		ST	22.7	13.3	6.7	10.8	13.8	16.7	19.7
		SM	13.3	12.0	8.0	10.8	16.9	9.1	10.0
		WJ	21.3	12.0	8.0	10.8	13.8	15.2	5.7
WF ₁	33.3	18.7	13.3	14.9	16.9	18.7	21.4		
WF ₂	13.3	12.0	4.0	6.8	16.9	7.6	5.7		
"K"	Ph(E)	BB	6.7	1.3	2.7	0.0	0.0	0.0	0.0
		HB	1.3	1.3	0.0	2.7	1.5	0.0	1.4
	Ph(D)	GC	8.0	1.3	0.0	1.4	1.5	0.0	0.0
		KH	4.0	8.0	1.3	2.7	3.1	3.0	0.0
	NSp	CJ	9.3	4.0	4.0	5.4	10.8	7.6	2.9
ST		5.3	4.0	0.0	0.0	0.0	0.0	1.4	

Tab. II: Fehler in Prozent (Perzeptionstests)

Gruppe	Vp	Test I				Test II		
		P 1	P 2	P 3	P 4	P 1	P 2	P 3
"LR"	OK	23	13	14	7	24	12	14
	PA	11	9	4	8	18	13	16
	SS ₁	24	20	13	6	27	19	16
	SA ₁	14	14	15	11	23	27	20
	SA ₂	13	35	27	17	27	22	25
	SS ₂	16	17	10	9	19	29	17
	SS ₃	17	10	17	15	27	21	20
	WB	20	14	6	3	16	11	19
"M"	BM	10	3	1	5	10	7	10
	BS	12	9	8	12	9	8	5
	BT	21	13	11	12	18	14	14
	DL	11	8	8	7	10	5	3
	DM	22	16	10	9	15	12	16
	DK	12	4	5	13	10	6	14
	FT	13	11	19	13	14	7	16
	GK	13	6	5	8	5	1	7
	HH	17	5	4	4	12	12	10
	HU	14	6	1	2	8	5	7
	JK	10	7	3	1	7	8	10
	KS	12	12	6	6	9	6	12
	KB	7	3	5	3	9	13	12
	RR	10	6	4	3	2	13	2
	ST	14	9	9	8	4	13	14
	SM	8	14	7	4	4	10	10
	WJ	11	14	4	10	5	9	9
	WF ₁	16	17	13	14	11	14	13
WV	3	4	4	4	2	6	3	
WF ₂	11	6	3	7	5	7	8	
"K"	BB	3	2	2	1	-	-	-
	HB	1	1	1	1	1	-	1
	GC	5	2	-	1	-	1	-
	KH	3	3	3	3	1	1	2
	CJ	10	6	-	1	6	4	4
ST	3	1	1	2	1	-	-	
		420	320	243	230	359	336	349

Tab. III: Positionsfehler

Gruppe	Vp	Test I			Test II		
		1.	2.	3.	1.	2.	3.
"LR"	OK	22	17	23	13	18	21
	PA	17	9	13	13	18	16
	SS ₁	20	20	26	7	24	32
	SA ₁	23	18	29	16	17	39
	SA ₂	32	38	31	16	28	31
	SS ₂	22	18	15	22	17	29
	SS ₃	21	23	19	17	24	28
	WB	14	12	20	12	13	22
"M"	BM	12	4	6	6	7	16
	BS	15	14	12	3	11	8
	BT	18	18	24	7	20	20
	DL	9	11	14	1	10	7
	DM	16	21	22	15	13	16
	DK	16	11	12	8	10	12
	FT	25	14	24	9	16	13
	GK	15	10	13	4	4	5
	HH	8	7	18	4	15	17
	HU	9	4	11	2	9	9
	JK	9	5	8	1	10	14
	KS	11	14	13	7	9	12
	KB	9	1	10	5	12	19
	RR	12	5	10	1	4	12
	ST	21	9	17	7	10	14
	SM	11	11	13	5	9	10
	WJ	19	14	10	6	7	11
	WF ₁	26	22	20	4	15	19
	WV	8	4	6	4	4	5
	WF ₂	12	9	10	9	6	6
Σ		452	364	449	224	360	463

Tab. IV: Fehlerverhalten in Abhängigkeit von der Testdauer

Gruppe	Vp	Test I		Test II	
		Beginn	Ende	Beginn	Ende
"LR"	OK	5	7	2	1
	PA	7	6	-	1
	SS ₁	3	7	1	3
	SA ₁	6	9	2	3
	SA ₂	9	10	1	1
	SS ₂	3	5	3	5
	SS ₃	4	4	1	2
	WB	3	4	1	4
"M"	BM	3	1	2	3
	BS	-	6	-	1
	BT	3	7	1	2
	DL	-	3	-	1
	DM	2	6	1	-
	DK	5	5	-	1
	FT	7	7	1	-
	GK	6	5	-	1
	HH	3	2	2	2
	HU	1	1	-	1
	JK	1	1	-	1
	KS	2	2	1	3
	KB	2	3	2	2
	RR	4	2	-	2
	ST	7	7	-	-
SM	2	3	-	-	
WJ	4	5	1	2	
WF ₁	8	5	-	1	
WV	3	2	2	1	
WF ₂	4	3	1	-	
Σ		107	128	25	44

Tab. V: Fehler in identischen Testwortreihen am Beginn und Ende jedes Einzeltests

Gruppe	Vp		/e/		/ø/		R		V		FVC	
	D	Id	D	Id	D	Id	D	Id	D	Id	D	Id
"LR"	SA1	0.70	0.46	0.56	0.58	0.67	0.83	0.67	0.90	0.90	0.72	0.61
	SS2	0.78	0.67	0.72	0.63	0.75	0.53	0.94	0.85	0.83	0.74	
	SS3	0.63	0.62	0.67	0.74	0.83	0.67	0.98	0.87	0.82	0.68	
	WB	0.81	0.71	0.78	0.74	0.94	0.70	0.92	0.94	0.82	0.84	
"M"	BM	0.93	0.71	0.94	0.74	1.00	0.97	0.95	0.96	0.90	0.97	
	DL	0.78	0.92	0.89	0.89	0.97	1.00	0.99	1.00	0.88	1.00	
	DM	0.81	0.75	0.67	0.63	0.86	0.73	0.95	0.98	0.80	0.79	
	HH	0.78	0.58	0.78	0.74	0.89	0.97	0.93	0.96	0.90	0.95	
	HU	0.81	0.83	0.89	0.84	0.92	1.00	0.96	0.98	0.95	0.95	
	JK	0.93	0.79	0.89	0.79	0.94	0.97	0.98	1.00	0.95	0.95	
	RR	0.89	0.75	0.83	0.95	0.94	1.00	0.99	1.00	0.93	0.95	
	SM	0.85	0.75	0.83	0.79	0.92	1.00	0.99	1.00	0.93	0.92	
MW		0.81	0.71	0.79	0.76	0.90	0.85	0.96	0.95	0.87	0.86	
StA		0.09	0.12	0.11	0.11	0.07	0.18	0.03	0.05	0.07	0.13	

Tab. VI: Perzeptionsleistungen in gemeinsamen Kontexten

Gruppe	Vp		/e/		/ø/		R		V		FVC					
	III	IV	V	III	IV	V	III	IV	V	III	IV	V				
"LR"	SA1	0.38	1.00	1.00	0.58	1.00	0.77	0.74	0.83	0.97	0.54	0.64	0.46	0.25	0.37	0.42
	SS2	0.63	0.90	0.83	0.25	0.13	0.77	0.86	0.82	0.90	0.44	0.72	0.61	0.38	0.52	0.55
	SS3	0.19	1.00	0.83	0.25	1.00	1.00	0.79	0.79	0.83	0.60	0.61	0.50	0.44	0.55	0.50
	WB	0.06	0.40	0.25	0.08	0.25	0.27	0.87	0.88	0.93	0.66	0.67	0.43	0.29	0.41	0.37
"M"	BM	0.94	0.90	1.00	0.92	0.80	1.00	0.57	0.67	0.86	0.40	0.59	0.47	0.44	0.77	0.60
	DL	0.84	1.00	1.00	0.50	0.75	1.00	0.93	0.89	0.97	0.40	0.61	0.51	0.67	0.67	0.81
	DM	0.38	0.90	0.50	0.88	0.75	0.73	0.66	0.90	0.93	0.53	0.58	0.60	0.33	0.33	0.35
	HH	0.81	1.00	0.92	0.60	0.63	1.00	0.76	0.83	1.00	0.58	0.72	0.55	0.62	0.76	0.88
	HU	0.44	1.00	0.75	0.50	0.63	1.00	0.73	0.73	0.87	0.61	0.63	0.51	0.71	0.56	0.78
	JK	0.78	1.00	0.83	0.67	0.88	1.00	0.81	0.90	0.77	0.79	0.81	0.55	0.65	0.83	0.89
	RR	0.94	0.90	1.00	1.00	0.75	1.00	0.96	0.98	1.00	0.67	0.78	0.70	0.62	0.65	0.74
	SM	0.69	1.00	0.75	0.83	0.75	1.00	0.77	0.63	0.93	0.47	0.54	0.53	0.48	0.76	0.57
MW		0.59	0.92	0.81	0.59	0.69	0.88	0.79	0.82	0.91	0.56	0.66	0.54	0.49	0.60	0.62
StA		0.29	0.17	0.23	0.29	0.26	0.22	0.11	0.10	0.07	0.12	0.08	0.07	0.16	0.17	0.19

Tab. VII: Produktionsleistungen in gemeinsamen Kontexten

Testwörter: Test III
Perzeption-Produktion (Nonsensewörter)

- | | | |
|---------------|--------------|---------------|
| 1. /trɪg/ | 26. /ɛft/ | 51. /saɪðd/ |
| 2. /grʌb/ | 27. /bru:s/ | 52. /rɔɪl/ |
| 3. /nu:fs/ | 28. /rəʊv/ | 53. /strɛmps/ |
| 4. /sl:ðz/ | 29. /meɪð/ | 54. /rɛp/ |
| 5. /ɛndz/ | 30. /sl:ð/ | 55. /trɪg/ |
| 6. /snɔ:d/ | 31. /θrɪf/ | 56. /θræt/ |
| 7. /sprɛɪn/ | 32. /ɛnts/ | 57. /klɛps/ |
| 8. /strɪmps/ | 33. /mɛθt/ | 58. /rɛp/ |
| 9. /ɛnts/ | 34. /lɔ:rl/ | 59. /ðɛns/ |
| 10. /klɛpθ/ | 35. /θɑ:ft/ | 60. /ðaɪn/ |
| 11. /rɪb/ | 36. /rɛŋks/ | 61. /θeɪnd/ |
| 12. /ðɛns/ | 37. /θeɪd/ | 62. /stɛɪvz/ |
| 13. /mɪnzð/ | 38. /rɑ:θ/ | 63. /nʌnθ/ |
| 14. /fɛrl/ | 39. /su:ð/ | 64. /ðɪ:/ |
| 15. /sɪlθ/ | 40. /nu:vz/ | 65. /snɔ:t/ |
| 16. /trɛg/ | 41. /ɛft/ | 66. /skraɪt/ |
| 17. /strɛmps/ | 42. /skraɪd/ | 67. /θɛtʃ/ |
| 18. /klɪbz/ | 43. /grɑ:θs/ | 68. /sɔ:θ/ |
| 19. /fɛɪθ/ | 44. /rəʊf/ | 69. /rɪp/ |
| 20. /ɪft/ | 45. /ɪnts/ | 70. /tɛɪnt/ |
| 21. /trɛg/ | 46. /ðɛə/ | 71. /su:s/ |
| 22. /θɔ:n/ | 47. /meɪðd/ | 72. /klɪps/ |
| 23. /ðʌs/ | 48. /rɪst/ | 73. /bru:z/ |
| 24. /θrɛɪm/ | 49. /gɑɪð/ | 74. /su:z/ |
| 25. /klɛps/ | 50. /stɛɪfs/ | 75. /tɛɪnd/ |

Testwörter: Test IV
Perzeption-Produktion: englische Einzelwörter

- | | | |
|------------|-------------|---------------|
| 1. run | 23. tree | 45. thick |
| 2. very | 24. clothes | 46. trousers |
| 3. six | 25. black | 47. this |
| 4. leg | 26. mother | 48. thank you |
| 5. cold | 27. mouth | 49. rain |
| 6. get | 28. page | 50. live |
| 7. head | 29. wrong | 51. thing |
| 8. dog | 30. map | 52. fat |
| 9. happy | 31. give | 53. read |
| 10. apple | 32. pencil | 54. ink |
| 11. hand | 33. bad | 55. seven |
| 12. that | 34. bring | 56. nose |
| 13. three | 35. write | 57. five |
| 14. there | 36. please | 58. friend |
| 15. with | 37. father | 59. green |
| 16. great | 38. sit | 60. bag |
| 17. good | 39. twelve | 61. teeth |
| 18. many | 40. story | 62. month |
| 19. simple | 41. milk | 63. letter |
| 20. they | 42. red | 64. cat |
| 21. sister | 43. finger | 65. rose |
| 22. train | 44. catch | 66. back |

Testwörter: Test V

Produktion-Perzeption-Produktion: englische Sätze

1. Billy is my friend.
2. Billy is a fat boy.
3. Are you fat or thin?
4. Thank you very much.
5. Billy's train isn't black; it's red.
6. What's in your bag?
7. Where is my apple, please?
8. Billy lives in our lane.
9. Roses are red.
10. Father, mother, sister, brother,
Hand in hand with one another.
11. He is fat.
12. You are thin.
13. She is a pretty girl.
14. The dog is black and fat.
15. The cat is thin and pretty.
16. Pat is Peter's dog.
17. He is black and white.
18. Dogs and cats are animals.
19. Billy has two hands with ten fingers.
20. Billy's face has two eyes, a nose, and a mouth.
21. His teeth are white.
22. Billy's hair is black.
23. His ears are red.
24. There are five knives on the table.
25. One, two, three, four, five,
Go and catch a hare alive.
26. The door is there.
27. Billy reads a book.
28. He writes in his exercise-book.
29. There is ink in his face and on his hands, too.
30. They are my friends.
31. We have many friends.

32. He is short and fat, indeed.
33. That boy is Billy.
34. The fat boy is Billy.
35. The pretty girl is Betty.