

Theoretische Probleme und praktische Darstellung der akustischen Projektion der Artikulation

1. Theoretische Probleme

In Anlehnung an eine Erkenntnis von G. Fant (1960, S. 218) gibt O. A. Vértes (1982, S. 155) der Ansicht über die Beziehung zwischen der Artikulation und dem Klangprodukt Ausdruck, daß der Sprachschall immer seine artikulatorische Grundlage habe, nicht aber umgekehrt: die Artikulation manifestiere sich nicht immer in der akustischen Projektion. Auf die weitere Darlegung dieses Problems geht er nicht ein, nur die von ihm angeführten Beispiele zeugen dafür, daß er dabei nicht die Artikulation im allgemeinen, sondern nur gewisse Elemente und Momente der Artikulation meint. Für die Darstellung seiner Auffassung nennt er ungarische Beispiele, man kann aber leicht Analogien finden, die für das Deutsche zutreffen, das soll an den folgenden beiden Wörtern belegt werden: Die Verlagerung der Zunge bereits während der stummen Haltephase des Verschlußlautes [p] in „bipolar“ oder des intervokalischen [b] in „bebauen“ von der Artikulationsstelle des unmittelbar vorausgehenden Vokals zu der des nachfolgenden hinterläßt in der akustischen Struktur des Klangproduktes keine mit dem unbewaffneten Ohr unmittelbar wahrnehmbare Spur. Diese physiologische Bewegung der Zunge erfolgt aber offensichtlich nicht außerhalb der Artikulation.

Im Gegensatz zu wahrnehmbaren Sprechbewegungen, die aber keine akustische Projektion haben, erfolgen manche Sprechbewegungen unbemerkt, dementsprechend bleiben auch ihre akustischen Wirkungen unbemerkt: Neben der Assimilation der Frequenzwerte der Formanten benachbarter Laute mit Formantenstruktur tritt auch die Zunahme der Werte der Bandbreite des ersten Formanten (die Formanten im folgenden $F_0 \dots F_x$) der den Nasallauten [m] und [n] unmittelbar vorausgehenden Vokale vermutlich infolge der bereits während der letzten Bildungsphase des Vokals einsetzenden Lösung des Gaumensegels von der Rachenwand und somit infolge der allmählichen Einschaltung des nasalen Resonators ein, ohne daß man diese Veränderungen als typische Konstituenten der erzeugten Lautstruktur – im Gegensatz zu Merkmalen wie Verschlußbildung und -lösung, Enge, Reibung und Effekte der Turbulenz, Stimmhaftigkeit oder Stimmlosigkeit u.a. – wahrnimmt. Auch die Wahrnehmung des unterschiedlichen zeitlichen Ablaufs der Prägung von Vokalformanten nach den Nasalen erfolgt unbewußt: Die Entwicklung der

Vokalformanten nach [m] beansprucht bis zum Eintreten der sogenannten Klarphase im allgemeinen eine Zeitdauer von 20 Millisekunden (im folgenden ms), nach [n] dagegen eine von 80-100 ms, also eine vier- bis fünffache Zeitdauer, Untersuchungen von Olaszy (1985, S. 76, ferner 155 f.) sprechen dafür, daß man diese Erscheinung für eine mit der beschriebenen Artikulationsbewegung typisch korrelierende akustische Äquivalenz ansehen kann.

Neppert und Pétursson (1986, S. 69 ff.) nehmen die Kategorisierung von Sprachschallformen auf die Sprechlautbestimmung bezogen vor und kommen u.a. zur folgenden Feststellung über ihre Beziehung: „Die Schallformen entsprechen entweder direkt bestimmten Lautmerkmalen oder konstituieren in kombinierter Form solche Merkmale.“ Bei der Interpretation der Beziehung der Artikulation zu ihrer akustischen Projektion geht Lindner (1969, S. 81) noch einen Schritt weiter und schaltet den Perzipienten als schöpferischen Faktor ein, dessen Rolle in der Sprechkommunikation und deren Berücksichtigung das Verhältnis der physiologischen Erzeugung des Sprachschalls und seiner akustischen Projektion in ein neues Licht stellt. „Wörter, Silben und Laute sind Einheiten, die wohl vom Produzenten willentlich erzeugt und vom Perzipienten aus dem lautsprachlichen Zeichen entnommen werden, die aber als Ding an sich keinen Sinn haben. Nur vom Perzipienten aus gesehen besteht das akustische Zeichen aus Silben und Lauten, und nur von seiner Warte her, d.h. von einer sprachlich-gesellschaftlichen Grundlage, kann gesagt werden, welche Teile des lautsprachlichen Signals zu einer Einheit gehören ...“

Aus diesen Darlegungen ergibt sich unter anderem die Einsicht, daß die Konstituenten des Sprachschalls (die Kategorien der segmentalen Struktur, d.h. die Laute und ihre Übergänge, sowie die Elemente der suprasegmentalen Struktur, die die segmentale überlagern, d.h. die Dynamik – und Tonhöhenverhältnisse, die zeitliche Gliederung usw.) erst von der Akustik her – also durch den Perzipienten – der komplexen instrumentalphonetischen Untersuchung zuverlässig zugänglich sind, und daß die linguistisch – akustische Untersuchung des Sprachschalls ebenfalls nicht ohne die Kenntnis der physiologischen Sinnesdaten der Sprechbewegungen denkbar ist. Verschiedene Aspekte des Herangehens haben allerdings den Wechsel ihrer Priorität und Dominanz zur Folge.

Es wird eigentlich erst aus der zitierten Feststellung von Lindner ersichtlich, in welchem Sinne Bolla (1981, S. 9) die Sprech- oder Redelaute als akustische Korrelate der kinetischen Stereotypen, d.h. der Bewegungskomplexe der Sprechwerkzeuge betrachtet bzw. wieso Neppert und Pétursson (op. cit., S. 69-70) die akustische Phonetik für eine „Theorie der Beziehungen zwischen der Physiologie der Sprechmotorik und dem Schallsignal auf der einen Seite und auf der anderen Seite zwischen dem Signal und dessen Verarbeitung beim Hörer“ ansehen.

Modellhaft – und beinahe schon extrem vereinfacht – kann man sagen, daß sich die Klanggestalt des Sprachsignals aus vier Grundschalltypen als akusti-

schen Korrelaten der vier grundlegenden Artikulationsmodi konstituiert. Das sind:

1. die als stumme Schallform wahrnehmbare akustische Äquivalenz der Haltephase des Verschlusses von Explosivlauten, eine sich durch einen Nullwert der Frequenz und einen Nullpegel der Intensität, zugleich aber durch eine typisierbare Zeitdauer auszeichnende phonetisch und sprachlich relevante Konstituente, die sich rein akustisch betrachtet durch nichts von den kurzen Sprechpausen zwischen benachbarten Sinneinheiten der Rede unterscheidet. Eine Unterscheidung der Funktionen ist in solchen Fällen nur von der Artikulation (und im weitesten Sinne von der Kenntnis des Aufbaus des Lautkörpers des sprachlichen Zeichens, die Phonemfolge mit inbegriffen) möglich;
2. der durch den Luftdruckunterschied unter Mitwirkung der kortikalen Steuerung sehr schnell gelöste orale Verschluß von Klusilen und der dadurch erzeugte Sprachschall;
3. der bei der Überwindung von verschiedenen Verengungen des Ansatzrohrs erzeugte Frikationsschall (die Turbulenzeffekte der Initialbehauung und die der Aspiration von stimmlosen Explosivlauten mit inbegriffen);
4. der durch die Schwingung der Stimmbänder als ausschließlicher Schallgrundlage für die (nicht geflüsterten) Vokale und als sekundäre Schallquelle für stimmhafte Konsonanten an der Glottis erzeugte quasiperiodische Klang, der als primäres oder sekundäres Quellensignal eine Modulation durch verschiedene Querschnittsänderungen des Ansatzrohrs erfährt, bevor er als Schallwelle in den Raum abgestrahlt wird.

Diese Modellierbarkeit und ihre Ergebnisse könnten leicht den Eindruck erwecken, daß es – ein theoretisch gut überlegtes und technisch exaktes Forschungsverfahren vorausgesetzt – ziemlich einfach ist, die Konstituenten des Klangproduktes unter anderem den jeweiligen Querschnittsänderungen des Ansatzrohrs oder anderen Momenten der Sprechbewegung genau zuzuordnen, die akustische Projektion der Artikulation einwandfrei zu ermitteln und dadurch auch zu unanfechtbaren Rückschlüssen für die Artikulation zu kommen. In Wirklichkeit muß man aber vor allem die zu untersuchenden Abschnitte des Vorganges der Sprechkommunikation zuverlässig segmentieren können, und gerade das ist sowohl physiologisch als auch akustisch mit vielen Schwierigkeiten verbunden. Über die Problematik der Interpretation von Sonagrammen, die in der Regel als Grundlage der Segmentierung von Schallkonstituenten dienen, schreiben Neppert und Pétursson (op. cit. S. 85) folgendes: „Für Sonagramme gilt das gleiche wie für alle anderen Registrierungsverfahren der experimentellen Phonetik, daß es nämlich keine Regeln gibt, die es ermöglichen, die Zuordnungen zwischen den im Sonagramm iso-

liebaren Elementen der Schalldarstellung und den Sprachlauten (bzw. den Phonemen) oder den prosodischen Sprachformen (bzw. den Prosodemen) eindeutig anzugeben. Es ist nur möglich, gewisse Wahrscheinlichkeitsregeln über diese Entsprechungen zu formulieren. Das entscheidende Problem ist, daß im Sonagramm entsprechend der Artikulations- und Schallwirklichkeit kontinuierliche Schallbilder erscheinen, die prinzipiell keine exakten Grenzen haben und damit keine einzelnen diskreten Segmente darstellen können. Damit soll nicht geleugnet werden, daß es für den Sprachhörer im Sprachschall gewisse Strukturen und Markierungen als Grundlagen für sein sprachauditives Segmentieren gibt [...] Doch wie dieses sprachauditive Segmentieren im einzelnen funktioniert, ist bisher nicht exakt angebar." Da es in Ermangelung absolut zuverlässiger Methoden zur Segmentierung von diskreten Lauteinheiten im Sonagramm durchaus angebracht ist, die vorzunehmende bzw. die bereits vorgenommene Segmentierung zu rechtfertigen, müßte man immer die Aspekte und Argumente offenlegen, an die man sich bei der Durchführung solcher Analysen gerade gehalten hat.

Bei den mit den Instrumenten der Ungarischen Akademie der Wissenschaften in Budapest durchgeführten Untersuchungen wurden folgende Aspekte beachtet bzw. die folgenden Methoden und Verfahren angewandt:

1. Die von Lindner (a.a.O.) dargestellte Rolle des Perzipienten bei der sprachlichen Verarbeitung von Schallwellen erwies sich als guter Ansatz für die Segmentierung von Einheiten, die untersucht werden sollten. Es war bei der Durchführung der Untersuchungen klar, daß man sich von der Kombination einer möglichst zuverlässigen Meßtechnik und des sprachauditiven Segmentierungsvermögens von Perzipienten eine vertretbare Lösung der Aufgabe versprechen konnte. Die beiden Voraussetzungen mußten nur aufeinander abgestimmt werden.
2. Die Segmentierung der Laute wurde an Breitbandsonagrammen vorgenommen, weil diese die Veränderungen der Formanten von Lauten mit Formantenstruktur in ihrem zeitlichen Verlauf gut verfolgbar in der Form von relativ breiten und markanten Streifen darstellen. Die parallele Beachtung der an sichtbar gemachten harmonischen Teilschwingungen reichen Schmalbandsonagramme sowie auch der charakteristischen Schallpegel-Registrate hat die Segmentierung noch weiter präzisiert.
3. Zur Bestimmung der Zeitdauer von Segmenten ist die genaue Kenntnis von mindestens zwei Punkten unerläßlich. Das sind der Anfang und das Ende. Der Vorgang der Erzeugung des Sprachschalls aber zeichnet sich durch die kontinuierliche Bewegung der Sprechwerkzeuge und dementsprechend auch durch kontinuierliche Änderungen der Struktur der Klangprodukte aus. Das Mitartikulieren von Elementen der näheren und weiteren Umgebung ergibt ein Geflecht von Konstituenten, das eine mehrfache Analyse erfordert.

4. Bereits bei der elektroakustischen Analyse der Laute zeitlich vorausgehenden Bewertung der videolabiographischen und der videoradiographischen physiologischen Aufnahmen über die Artikulationsbewegungen der Sprechwerkzeuge wurde die erste Segmentierung der Laute mit Hilfe einer empirischen Kalibration am Tonkopf des Videorecorders vorgenommen.
5. Dem Zeitverlaufsspektrographen (der Firma Kay Elemetrics, Modell 700), mit dessen Hilfe die Breitband- und die Schmalband-Sonagramme (300 bzw. 45 Hz) sowie die Sektion in Breitband- bzw. Schmalbandregistrierung gewonnen wurden, wurde ein speicherungsfähiges Oszilloskop (Typ 5113) angeschlossen, das den Schwingungsverlauf des ganzen sprachlich sinnvollen Morphems, in dem auch das gesuchte Segment enthalten war, auf dem Bildschirm sichtbar machte. Unter wiederholtem Abspielen und Abhören des nun visuell dargestellten und im Sonagramm gespeicherten Morphems konnten dessen Konstituenten, die man gerade nicht brauchte, links und rechts von dem zu segmentierenden Laut mit Hilfe des eingebauten Segmentators unter ständiger auditiver und visueller Kontrolle elektronisch gelöscht werden, bis das gesuchte Segment von der Umgebung isoliert auf dem Bildschirm des Skops erschien. Beim Löschen der gerade überflüssigen Konstituenten des Schallkörpers mußte man besonders vorsichtig den noch visuell und auditiv identifizierbaren Rest beurteilen, denn sehr kurze Zeitabschnitte (schon 10–15 ms) konnten zu Täuschungen bei der Lautwahrnehmung führen. Es ist auch oft angebracht, die Ergebnisse solcher Analysen durch Synthese zu kontrollieren, wenn die technischen Voraussetzungen dazu gegeben sind. Auch die sorgfältig gewählten Lautnachbarschaften können die Zuverlässigkeit der Segmentierung erhöhen. Nur Logatome sollte man grundsätzlich vermeiden!

Nachdem die identifizierbaren Komponenten des gesuchten Segments restlos mit dem beschriebenen Verfahren, das noch durch eine Schwingungsanalyse mit dem Programm „BESZ-33“ des Computers PDP 11/34 ergänzt wurde, ermittelt worden waren, konnte man mit der Analyse der inneren Struktur des Lautes beginnen. Die prozentuale Umrechnung der in Millisekunden ermittelten Werte der Bildungsphasen und die diesen zuzuordnenden Frequenzwerte haben eine besser vertretbare Verallgemeinerung ermöglicht. Die Intensitätswerte werden in dB angegeben, F_0 steht für den Primärklang. Die prozentuell umgerechneten Werte des zeitlichen Verlaufs der Artikulation passen sich möglichst genau den signifikanten Entwicklungstendenzen der intensivsten Bereiche der Formanten an.

In der vorliegenden Studie werden die akzentuierten Vokale sowie die schwachtonige Realisierung des /e/ wegen ihrer Häufigkeit untersucht.

2. Die praktische Durchführung der Analyse und die Ergebnisse

Laut [a] in „Acker“			Laut [a:] in „Adam“		
Fo 120 Hz. Dauer 100 ms			Fo 120 Hz. Dauer 244 ms		
	Bildungs- phasen (%)	Frequenz- werte (Hz)		Bildungs- phasen (%)	Frequenz- werte (Hz)
F1	0-100	640- 790	F1	0-100	610-850
F2	0- 60	1000-1320	F2	0-20	1100-1130
	60-100	1320-1550		20-45	1130-1180
				45-75	1180-1400
				75-90	1400-1640
				90-100	1640-1900
F3	0- 30	2740-2680	F3	0-30	2980-3100
	30- 70	2680-2520		30-70	3100-3000
	70-100	2520-2380		70-100	3000-3100
F4	0- 50	3700-3630	F4	0-100	3750-4100
	50-100	3630-3550			
Intensität:		F2 36 DB	Intensität		F1 34 dB
		F1 -5 dB			F2 -0 dB
		F3 -7 dB			F3 -4 dB
		F4 -16 dB			F4 -18 dB
Laut [O] in „Ochse“			Laut [O:] in „Ofen“		
Fo 110 Hz. Dauer 62 ms			Fo 120 Hz. Dauer 173 ms		
	Bildungs- phasen (%)	Frequenz- werte (Hz)		Bildungs- phasen (%)	Frequenz- werte (Hz)
F1	0-100	470- 690	F1	0-100	230- 480
F2	0-100	920-1280	F2	0-100	640- 920
F3	0- 50	2650-2570	F3	0-100	2510-2710
	50-100	2570-2400			
F4	0-100	3660-3410	F4	0-100	nicht belegt
Intensität:		F1 41 dB	Intensität:		F1 45 dB
		F2 -0 dB			F2 -13 dB
		F3 -16 dB			F3 -32 dB
		F4 -27 dB			F4 keine Angabe
Laut [v:] in „Mutter“			Laut [u:] in „Uhr“		
Fo 120 Hz. Dauer 62 ms			Fo 120 Hz. Dauer 173 ms		
	Bildungs- phasen (%)	Frequenz- werte (Hz)		Bildungs- phasen (%)	Frequenz- werte (Hz)
F1	0-100	210- 590	F1	0-100	160- 400
F2	0- 35	840-1120	F2	0-100	500- 900
	35-70	1120-1420			
	70-100	1420-1800			
F3	0- 50	2800-2720	F3		nicht belegt
	50-100	2720-2480			
F4	0-100	3770-3300	F4		nicht belegt
Intensität:		F1 47 dB	Intensität:		F1 46 dB
		F2 -15 dB			F2 -17 dB
		F3 -30 dB			
		F4 -35 dB			

Laut [Y] in „üppig“
Fo 120 Hz. Dauer 86 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	190- 550
F2	0- 35	1710-1830
	35- 70	1830-1830
	70-100	1830-1650
F3	0-100	2350-2660
F4	0-100	3150-3530
Intensität:	F1	47 dB
	F2	-15 dB
	F3	-25 dB
	F4	-33 dB

Laut [y:] in „Übel“
Fo 120 Hz. Dauer 190 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	200- 580
F2	0- 35	2510-2450
	35- 70	2450-2320
	70-100	2320-2080
F3	0-100	3200-2750
F4	0-100	3850-3430
Intensität:	F1	40 dB
	F2	-10 dB
	F3	-21 dB
	F4	-31 dB

Laut [oe] in „öffnen“
Fo 120 Hz. Dauer 115 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	290- 550
F2	0-100	1500-1680
F3	0-100	2400-2600
F4	0-100	3470-3760
Intensität:	F1	38 dB
	F2	-5 dB
	F3	-11 dB
	F4	-21 dB

Laut [Ø:] in „böse“
Fo 120 Hz. Dauer 330 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	230- 560
F2	0- 20	1670-1540
	20- 60	1540-1670
	60- 80	1670-1720
F3	80-100	1720-1520
	0- 20	2450-2350
F4	20- 70	2350-2300
	70-100	2300-2650
	0-100	3380-3570
Intensität:	F1	42 dB
	F2	-12 dB
	F3	-20 dB
	F4	-31 dB

Laut [I] in „impfen“
Fo 120 Hz. Dauer 90 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	260-580
F2	0-35	2510-2450
	35-70	2450-2320
	70-100	2320-2080
F3	0-35	3200-3100
	35-70	3100-2900
	70-100	2900-2750

Laut [i:] in „Bibel“
Fo 120 Hz. Dauer 240 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	160-460
F2	0-20	2700-2550
	20-40	2550-2460
	40-60	2460-2400
F3	60-80	2400-2430
	80-100	2430-2470
F4	0-20	3130-3250
	20-40	3250-3400
	40-60	3400-3350
	60-80	3350-3300
	80-100	3300-3260

Laut [I] in „impfen“		
Fo 120 Hz. Dauer 90 ms		
	Bildungs-	Frequenz-
	phasen (%)	werte (Hz)
F4	0-100	3850-3430
Intensität:		
	F1	43 dB
	F2	-12 dB
	F3	-18 dB
	F4	-16 dB

Laut [i:] in „Bibel“		
Fo 120 Hz. Dauer 240 ms		
	Bildungs-	Frequenz-
	phasen (%)	werte (Hz)
F4	0- 20	3550-3650
	20- 40	3650-3720
	40- 60	3720-3680
	60- 80	3680-3600
	80-100	3600-3650
Intensität:		
	F1	46 dB
	F2	-24 dB
	F3	-21 dB
	F4	-27 dB

Laut [ɛ] in „Ebbe“		
Fo 120 Hz. Dauer 130 ms		
	Bildungs-	Frequenz-
	phasen (%)	werte (Hz)
F1	0-100	410- 770
F2		
	0-30	1600-1700
	30-50	1700-2000
	50-70	2000-1870
	70-100	1870-1820
F3		
	0-20	2500-2650
	20-50	2650-3000
	50-70	3000-2930
	70-100	2930-2800
F4		
	0-30	3850-3900
	30-60	3900-4020
	60-80	4020-3870
	80-100	3870-3800
Intensität:		
	F1	39 dB
	F2	-9 dB
	F3	-16 dB
	F4	-21 dB

Laut [ɛ:] in „wägen“		
Fo 120 Hz. Dauer 250 ms		
	Bildungs-	Frequenz-
	phasen (%)	werte (Hz)
F1	0- 10	290- 410
	10- 30	410- 650
	30- 80	650- 650
	80-100	650- 500
F2		
	0- 30	1700-1900
	30- 50	1900-2000
	50- 80	2000-2100
	80-100	2100-2280
F3		
	0- 20	2750-2680
	20- 30	2680-2770
	30- 50	2770-2870
	50- 80	2870-2870
	80-100	2870-2800
F4		
	0- 10	3450-3570
	10- 30	3570-3680
	30- 50	3680-3880
	50- 80	3880-3880
	80-100	3880-3840
Intensität:		
	F1	37 dB
	F2	-3 dB
	F3	-11 dB
	F4	-17 dB

Laut [e:] in „beben“			Laut [ø:] in „beben“		
Fo 120 Hz. Dauer 280 ms			Fo 120 Hz. Dauer cca 50 ms		
	Bildungsphasen (%)	Frequenzwerte (Hz)		Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-100	230-560	F1	0-100	340-480
F2	0-20	2300-2430	F2	0-100	1600-1770
	20-40	2430-2540			
	40-60	2540-2630			
	60-80	2630-2580			
	80-100	2580-2500			
F3	0-20	2800-3080	F3	0-100	2820-2920
	20-40	3080-3140			
	40-60	3140-2980			
	80-100	2980-2840			
F4	0-70	3600-3820	F4	0-100	3800-3900
	70-100	3820-3720			
Intensität:	F1	39 dB	Intensität:	F1	0 dB
	F2	-10 dB		F2	-18 dB
	F3	-14 dB		F3	-19 dB
	F4	-15 dB		F4	-25 dB

Das /e/ wird in vielen unbetonten Positionen schwachtonig realisiert. Die unterschiedlichen Realisierungen können unterschiedliche Parameter aufweisen. Zum Vergleich wird hier noch der schwachtonige Vokal im Präfix „be-“ im Wort „bebildert“ behandelt. Die spezifische Bildungsdauer beträgt 90 ms, die Frequenz des Primärklanges ist 120 Hz. Die Frequenzwerte der Formanten: F1 430-670 Hz, F2 1430-1760 Hz, F3 2500-2870 Hz, F4 3580-3730 Hz. Intensität: F1 43 dB, F2 -10 dB, F3 -13 dB, F4 -32 dB. Die eingehendere Untersuchung der Lautschwächung geht aber über den Rahmen dieser Arbeit hinaus.

Laut [ai] in „Ei“			Laut [oi:] in „Bäume“		
Fo 120 Hz. Dauer 385 ms			Fo 120 Hz. Dauer 310 ms		
	Bildungsphasen (%)	Frequenzwerte (Hz)		Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-50	780-740	F1	0-20	750-780
	50-60	740-280		20-40	780-760
	60-100	280-280		40-60	760-680
				60-80	680-480
				80-100	480-340
F2	0-20	1720-1600	F2	0-20	1000-1150
	20-40	1600-1780		20-40	1150-1300
	40-60	1870-2200		40-60	1300-1500
	60-80	2200-2350		60-80	1500-2100
	80-90	2350-2560		80-90	2100-2170
	90-100	2560-2520		90-100	2170-2270
F3	0-60	2900-2950	F3	0-35	2500-2550
	60-90	2950-3000		35-70	2500-2680
	90-100	3000-2950		70-100	2680-2640

Laut [ai] in „Ei“			Laut [ði:] in „Bäume“		
Fo 120 Hz. Dauer 385 ms			Fo 120 Hz. Dauer 310 ms		
	Bildungsphasen (%)	Frequenzwerte (Hz)		Bildungsphasen (%)	Frequenzwerte (Hz)
F4	0-40	3900-3900	F4	0-35	3300-3450
	40-60	3900-3720		35-70	3450-3570
	60-100	3720-3680		70-100	3570-3700
Intensität:	F1	35 dB	Intensität:	F1	39 dB
	F2	-1 dB		F2	-5 dB
	F3	-10 dB		F3	-25 dB
	F4	-12 dB		F4	-21 dB

F2 ist bei 100 Hz um 5 db schwächer als F1, bei 200 Hz gemessen ist er um das Vierfache schwächer: -19 dB.

Die akustische Projektion der Artikulation der Diphthonge ist zu den physiologischen Bewegungstendenzen umgekehrt proportional.

Der Diphthong [au] in „Auto“

Fo 120 Hz. Dauer 244 ms

	Bildungsphasen (%)	Frequenzwerte (Hz)		Bildungsphasen (%)	Frequenzwerte (Hz)
F1	0-35	770-660	F3	0-30	2350-2500
	35-50	660-400	F3	30-50	2500-2640
	50-70	400-280		50-80	2640-2550
	70-100	280-280		80-100	2550-2450
F2	0-35	1450-1320	F4	0-25	3430-3490
	35-50	1320-1200		25-50	3490-3550
	50-70	1200-1260		50-75	3550-3670
	70-100	1260-1350		75-100	3670-3600
Intensität:	F1 43 dB,	F2 -8 dB,	F3 -22 dB,	F4 -19 dB.	

Zusammenhänge zwischen sogenannten „charakteristischen Einstellungen der Sprechwerkzeuge“ und *einigen Faktoren* der akustischen Projektion *dieser physiologischen Eigenartigkeiten* der Artikulation sind bekannt: die Länge des Ansatzrohres steht zum zahlenmäßigen Wert der Formantenfrequenz der Vokale in umgekehrter Proportion; der Grad der Zungenhebung und parallel dazu der Öffnungsgrad des Kieferwinkels sind zum Frequenzwert des ersten Formanten ebenfalls umgekehrt proportional; der Frequenzwert des zweiten Formanten wird auch durch die horizontale Verlagerung der Zunge beeinflusst: der zweite Formant der (palatalen) Vorderzungenvokale liegt höher als

der der entsprechenden (velaren) Hinterzungenvokale; auch die zunehmende Vorstülpung und Rundung der Lippen – die Intensität der Labialisierung – ist zum Wert der ersten drei Vokalformanten umgekehrt proportional, noch dazu werden F2 und F3 stärker von der Intensität der Labialisierung beeinflusst als F1.

Unabhängig davon, ob man diese richtig erkannten Zusammenhänge zwischen den Bewegungsrichtungen der Sprechwerkzeuge und der Entwicklungstendenz der Vokalformanten für eine Art Modell der Artikulation und ihrer akustischen Wirkung ansieht oder nicht, springen mindestens zwei Mängel sofort ins Auge:

1. Wesentliche akustische Komponenten des artikulatorisch erzeugten Sprachschalls wie die Intensität und die Bandbreite der Formanten, die Angleichungen innerhalb der Frequenz- und der Intensitätsstruktur werden außer acht gelassen.
2. Sowohl die notgedrungen vielfrequente Eigenart der Teilschwingungen des Primärklanges und deren Konsequenzen bei ihrer Modulation im Ansatzrohr als auch die Tatsache, daß die Hohlraumform des Ansatzrohres gleichzeitig an mehreren Stellen durch Verengungen verändert werden kann, bleiben unberücksichtigt. Die möglichst genaue zeitliche Verfolgung der Entwicklung der Bildungsphasen der Laute und die darauf ebenfalls möglichst genau abgestimmte zeitliche Verfolgung der Entwicklung der Formanten u.a. ergibt genauere Informationen über die Artikulation und deren akustische Projektion. Für die angemessene Erschließung der Beziehungen der Artikulation zum erzeugten Sprachschall kann man sich mehr von der sogenannten Quelle-Filter-Theorie, wie sie von Neppert und Pétursson (op. cit. S. 113 ff.) formuliert wurde, versprechen.

Als Schallquelle dient ja der artikulatorisch (an den Stimmbändern) erzeugte Primärklang. Dieser zeichnet sich durch einen relativ großen Schallpegel und durch vielfrequente Teilschwingungen aus. Diese zusammengesetzten Primärschwingungen werden einem Trägermedium, dem Sekundärschwinger: der Luftsäule im Ansatzrohr übertragen. (Auch ein Tertiärschwinger kann noch eingeschaltet werden.) Der Primärklang erfährt im Ansatzrohr zwei entgegengesetzte Arten der Modulation: durch die erzwungene Mitschwingung des Trägermediums und durch weitere Faktoren wie die Schleimhaut usw. wird der Schalldruck erheblich gedämpft, und parallel dazu werden die auf die Eigenfrequenz abgestimmten Schwingungsbereiche hervorgehoben und verstärkt: der Klang wird gefiltert und schließlich als Schallwelle abgestrahlt.

In der vorliegenden Studie wurde diese Quelle-Filter-Theorie für die Darstellung von Beziehungen zwischen der Artikulation und ihrer akustischen Projektion angewandt.

Literatur

- KÁLMÁN BOLLA: Az amerikai angol beszédhangok atlasza. In: Magyar Fonetikai Füzetek 9, Budapest 1981.
- GUNNAR FANT: Acoustic theory of speech production with calculations based on x-ray studies of Russian Articulation. 's-Gravenhage 1960.
- GERHART LINDNER: Einführung in die experimentelle Phonetik. Berlin 1969.
- JOACHIM NEPPERT-MAGNUS PÉTURSSON: Elemente einer akustischen Phonetik. Hamburg 1986.
- GÁBOR OLASZY: A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise. A számítógépes beszédelőállítás néhány kérdése. Nyelvtudományi értekezések, 121. szám. Budapest, 1985.
- ANDRÁS O. VÉRTES: Az artikuláció akusztikus vetülete. In: Fejezetek a magyar leíró hangtanból. Hg. von K. Bolla. Budapest, 1982. S. 155–164.