

Koartikulation

Zwei klassische Arbeiten von Öhman

1. Öhman (1966)

“Coarticulation in VCV Utterances: Spectrographic Measurements”

Akustische Eigenschaften von V2 sind bereits am Ende von V1 sichtbar

und umgekehrt:

Akustische Eigenschaften von V1 sind noch am Anfang von V2 sichtbar

Typische Begriffe in der Phonetik für diese zwei Koartikulationsrichtungen?

Akustische Eigenschaften von V2 sind bereits am Ende von V1 sichtbar

====> anticipatory

Akustische Eigenschaften von V1 sind noch am Anfang von V2 sichtbar

====> carryover

Illustration:

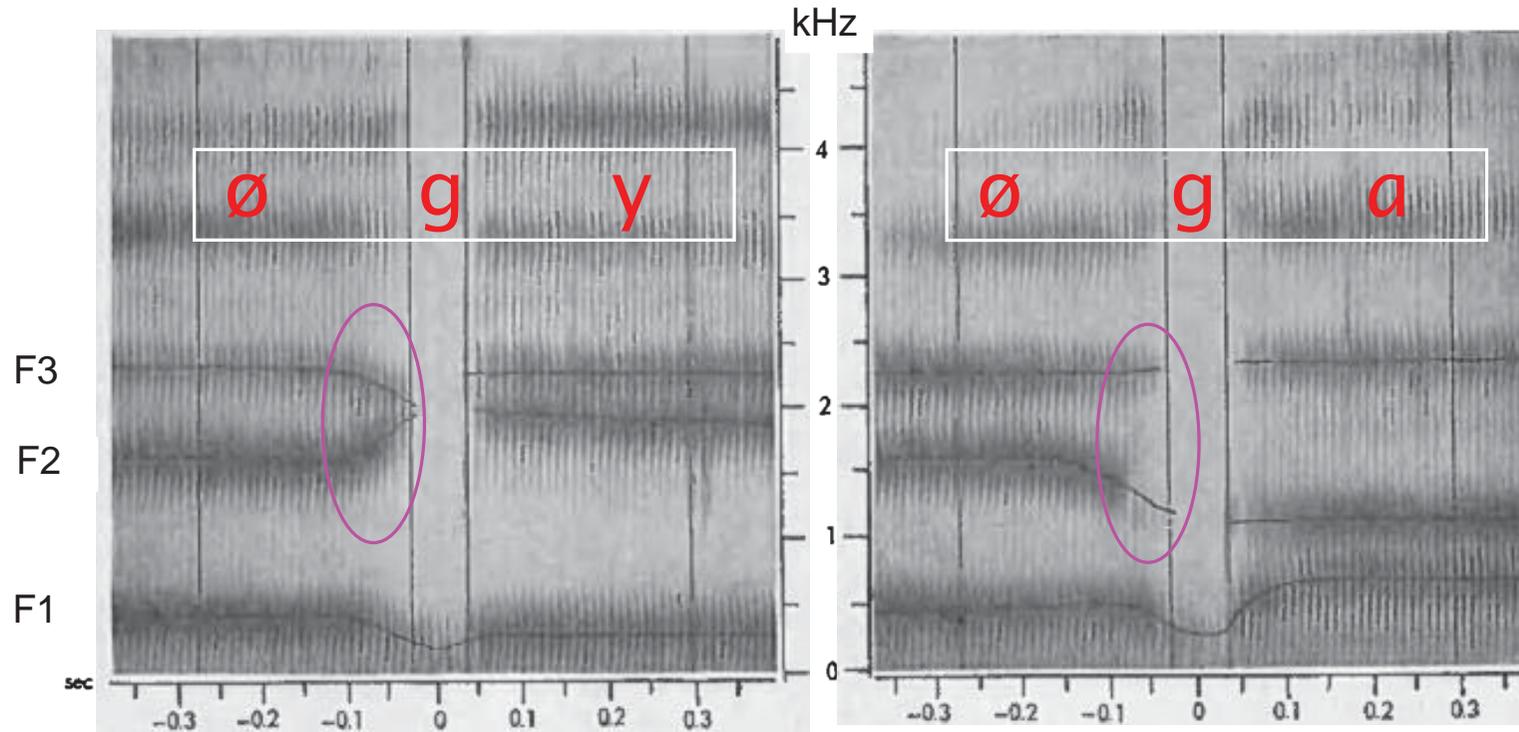


FIG. 1. Sound spectrograms of the utterances /øgy/ (left) and /øga/ (right) as spoken by a male Swedish talker. The formant transitions in the initial vowel are different in the two cases, owing to influence of the final vowel. The lines superimposed on the spectrograms indicate method of measurement discussed in the text.

unterschiedliche V2 (/y/ vs. /a/)

====> unterschiedliche Formantlage am Ende von V1 (/ø/)

Öhmans Material

Vokale: alle Kombinationen von /y, ø, a, o, u/ für V1 und V2

Konsonanten: /b, d, g/

Außer /a/ nur gerundete Vokale

====> Koartikulationseffekte sind dann auf unterschiedliche
Zungenpositionen zurückzuführen

Übung 1: Skizzieren von VC-Transitionen

Benutzen Sie die Information in Table II und Table IV, um für **einen** Vokal, alle VC-Transitionen für F2 zu zeichnen.

Die linke Hälfte von Fig. 5 aus Öhman, 1966, zeigt ein Beispiel für $V1 = / \emptyset /$

Formantwerte in der Vokalmitte. Initial = V1, Final = V2.

Diese Werte entsprechen den Startwerten der VC-Transitionen, bzw. den Endwerten der CV-Transitionen

TABLE II. Vowel-formant frequencies in cps of male Swedish speaker. Values are given corresponding to the initial and the final positions of the VCV utterances and corresponding to the different intervocalic stop-consonant contexts. The numbers in parentheses denote interquintile ranges. Each formant-frequency value is an average of 25 measurements.

	y		ø		α		o		u	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
VbV										
F ₁	290 (40)	330 (50)	400 (50)	430 (40)	670 (40)	660 (65)	400 (40)	420 (50)	340 (25)	360 (40)
F ₂	2000 (50)	1890 (140)	1650 (75)	1580 (90)	990 (75)	1040 (100)	670 (50)	730 (50)	670 (40)	680 (40)
F ₃	2350 (100)	2320 (75)	2320 (50)	2380 (75)	2670 (100)	2620 (125)		2450 (90)		2430 (100)
VdV										
F ₁	270 (45)	300 (35)	390 (40)	400 (60)	660 (45)	650 (65)	390 (40)	410 (35)	330 (40)	320 (45)
F ₂	1990 (75)	1950 (75)	1620 (70)	1580 (55)	970 (95)	1030 (100)	690 (40)	760 (45)	660 (40)	720 (75)
F ₃	2380 (100)	2350 (75)	2350 (65)	2410 (90)	2740 (125)	2700 (125)	2530 (175)	2560 (175)		2500 (150)
VgV										
F ₁	320 (40)	330 (65)	420 (45)	420 (55)	660 (55)	650 (80)	430 (45)	430 (35)	360 (50)	360 (55)
F ₂	2010 (75)	1960 (90)	1650 (50)	1650 (85)	990 (90)	1160 (125)	780 (45)	840 (90)	770 (40)	790 (45)
F ₃	2440 (75)	2340 (100)	2440 (115)	2350 (75)	2560 (150)	2480 (165)	2560 (115)	2450 (145)	2480	2460 (115)

TABLE IV. Data pertaining to formant transitions in the VCV utterances of a male Swedish speaker. The Table is arranged in essentially the same manner as Table I. Further explanations are given in the text in Sec. II-B2.

Ob- served vowel	Trans- conso- nantal vowel	VbV								VdV								VgV							
		F ₁		F ₂		F ₃		F ₁		F ₂		F ₃		F ₁		F ₂		F ₃							
		VC	CV	VC	CV	VC	CV	VC	CV	VC	CV														
y	y	-35	-90	-350	-265	-330	-235	-50	-115	-85	-85	245	255	-60	-75	130	145	-195	-120						
	ø	-20	-20	-445	-235	-285	-85	-80	-95	-170	-185	150	265	-80	-125	95	185	-220	-15						
	a	-40	-45	-500	-325	-295	-95	-65	-95	-260	-290	250	320	-65	-70	-185	110	-300	-35						
	o	-20	-20	-575	-320	-260	-115	-65	-110	-140	-290	180	220	-10	-55	-410	190	-310	-125						
	u	-10	-90	-475	-390	-320	-130	-70	-65	-155	-235	165	135	-40	-55	-375	185	-245	-25						
ø	y	-95	-130	-275	-50	-90	-155	-115	-170	-40	150	210	135	-130	-215	230	360	-300	-185						
	ø	-95	-125	-190	-130	-110	-80	-145	-180	-55	65	205	255	-130	-175	115	300	-205	-195						
	a	-85	-150	-355	-160	-40	-110	-110	-135	-210	-25	280	240	-115	-150	-130	410	-155	-140						
	o	-105	-140	-345	-235	-80	-65	-135	-135	-135	-115	245	165	-95	-130	-185	300	-180	-215						
	u	-115	-150	-370	-270	-125	-115	-155	-180	-95	15	290	105	-90	-140	-240	340	-150	-175						
a	y	-300	-325	150	35	-290	-255	-335	-355	365	440	-95	-20	-320	-420	510	430	-315	-355						
	ø	-220	-270	95	-15	-265	-280	-385	-390	310	480	-95	-50	-335	-370	400	405	-425	-275						
	a	-280	-330	-170	-55	-140	-210	-320	-350	125	225	-65	-50	-270	-325	225	110	-275	-145						
	o	-260	-330	-155	-80	-140	-145	-385	-270	175	170	-20	-75	-275	-355	50	55	-200	-25						
	u	-350	-300	-150	-185	-190	-225	-365	-375	280	220	-125	-65	-310	-300	110	20	-130	-80						
o	y	-85	-105	290	255	...	-300	-80	-215	590	870	50	100	-100	-175	195	260						
	ø	-120	-140	255	80	-250	-150	-85	-160	580	710	...	50	-105	-125	190	165	...	-190						
	a	-110	-105	-25	60	-50	-55	-50	-200	330	500	...	120	-115	-140	110	110						
	o	-140	-95	-5	0	-10	...	-75	-95	365	475	...	40	-120	-135	35	10						
	u	-100	-130	-10	-20	...	-125	-110	-175	425	520	...	30	-105	-120	15	-25						
u	y	-60	-50	375	225	-130	-315	-65	-125	665	900	...	-50	-65	-135	105	280	...	-265						
	ø	-60	-110	175	275	...	-135	-70	-110	665	800	-55	-125	60	180	...	-235						
	a	-90	-80	-135	170	...	-90	-60	-45	345	700	-75	-50	-60	135						
	o	-80	-65	-50	5	-25	-35	430	650	-100	-95	-65	40						
	u	-85	-90	-120	-25	...	-155	-60	-95	570	620	...	-50	-70	-95	-120	-50						

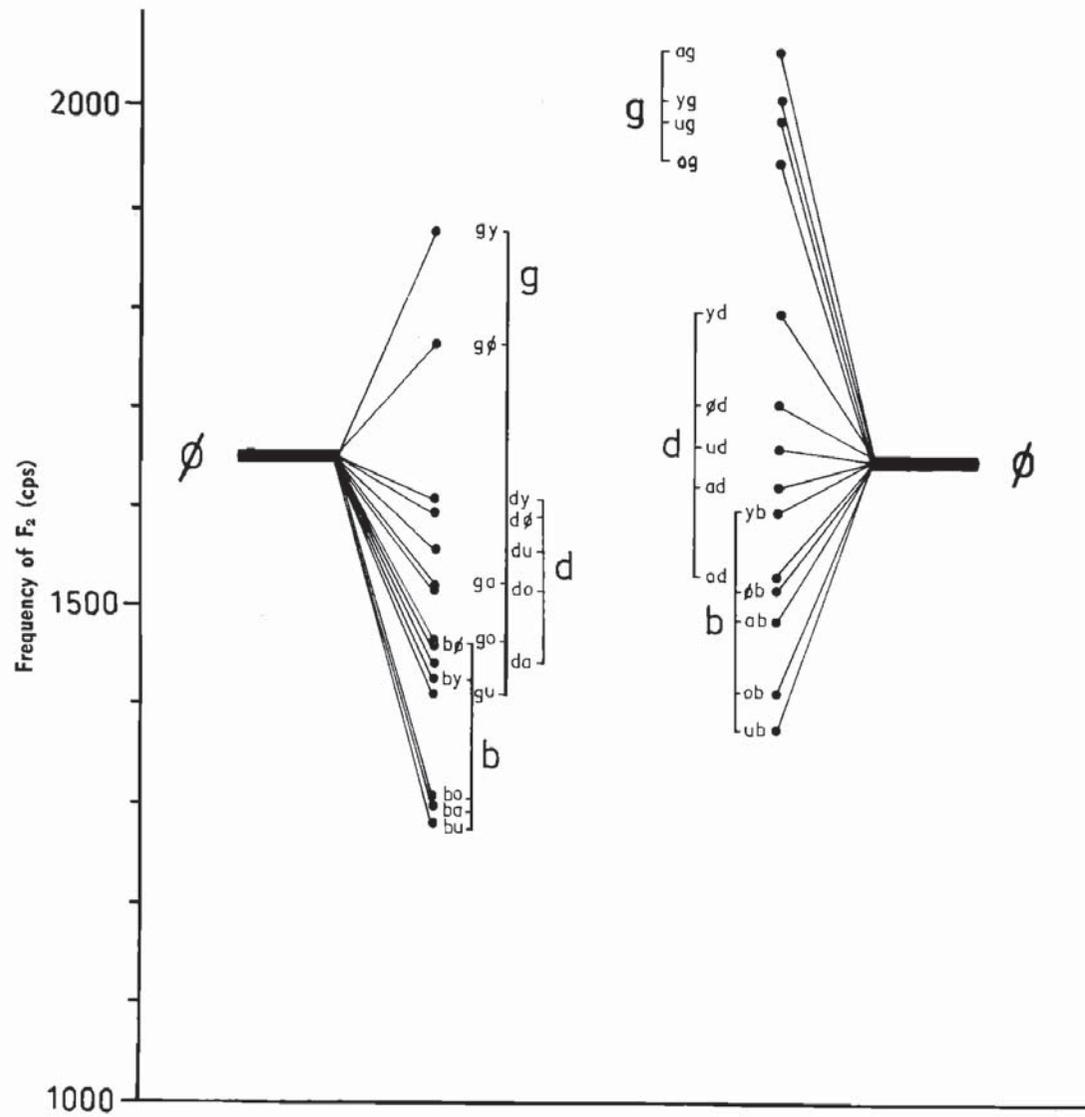


FIG. 5. Stylized second-formant transitions observed in the sequences /øC-/ (left) and /-Cø/ (right). The letters at the terminal points identify the medial consonant and transconsonantal vowel of the VCV utterance in which the transition was observed. The vertical lines cover the ranges of variation of the formant transitions associated with the medial consonants /b/, /d/, and /g/.

Übung 2: Hängt die Stärke der Koartikulation von folgenden Faktoren ab? Benutzen Sie hierzu die *ranges of variation* in Table V

(1) Richtung

anticipatory (Variabilität in VC-Transitionen) vs.
carryover (Variabilität in CV-Transitionen)

(2) Vokal

Sind bestimmte Vokale stärker betroffen als andere?

(3) Konsonant

Breitet sich der Einfluss von V1 auf V2 (und umgekehrt)
gleich stark über die 3 Konsonanten /b, d, g/ aus?

TABLE V. Ranges of variation in cps of the formant transitions presented in Table IV. Values that are too large to be due to errors of measurement or chance are printed in boldface.

	VbV						VdV						VgV					
	F ₁		F ₂		F ₃		F ₁		F ₂		F ₃		F ₁		F ₂		F ₄	
	VC	CV	VC	CV	VC	CV	VC	CV	VC	CV	VC	CV	VC	CV	VC	CV	VC	CV
y	30	70	225	155	70	150	30	50	175	205	100	185	70	50	540	80	115	110
ø	30	25	180	220	85	90	45	45	170	265	85	150	40	85	470	110	150	75
u	130	60	305	220	150	135	65	120	240	310	105	140	65	120	460	410	295	330
o	55	45	315	275	240	245	60	120	260	395	...	90	20	55	180	285
u	30	60	510	300	...	225	45	90	320	280	45	85	225	330

Weitere Aspekte

“Coarticulatory resistance”

vgl. Übung 2, Fragen (2) und (3)

Sehr umfangreiche Literatur hierzu. Übersicht in Farnetani & Recasens (1999)

Sprachspezifisch: sehr wenig V-V-Koartikulation im Russischen.

Wegen Palatalisierung der Konsonanten?

Probleme mit *locus theory*:

Lässt sich die Artikulationsstelle des Konsonanten durch die Zielfrequenz der Transitionen kennzeichnen?

Typisches Beispiel:

bei /di, da, du/ beginnt die CV-Transition bei ca. 1800Hz

Aber diese Locusfrequenz verschiebt sich bei unterschiedlichen Vokalen vor dem Konsonanten.

(näheres z.B Harrington/Cassidy, S. 87-91)

Diskussion

“Since traces of the final vowel are observable already in the transition from the initial vowel to the consonant, it must be concluded that a motion toward the final vowel starts not much later than, or perhaps even simultaneously with, the onset of the stop-consonant gesture. A VCV utterance of the kind studied here can, accordingly, *not* be regarded a linear sequence of three successive gestures. We have clear evidence that the stop-consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture” (p.165)

====> Zunge als Mehrfachartikulator

2. Öhman (1967)

“Numerical Model of Coarticulation”

Was bedeutet *Modell*?

2. Öhman (1967)

“Numerical Model of Coarticulation”

Ähnliches Material wie Öhman (1966).

VCV-Sequenzen

V1 und V2: alle Kombinationen von /y, a, u/.

C = /d, g/

Jetzt aber Analyse der Vokaltraktform anhand von cineradiographischen Aufnahmen.

“The tongue may be regarded as three independently controllable mechanical systems the apical articulator, the dorsal articulator, and the tongue-body articulator” (p.310)

(Consonants: apical, dorsal. Vowels: tongue body)

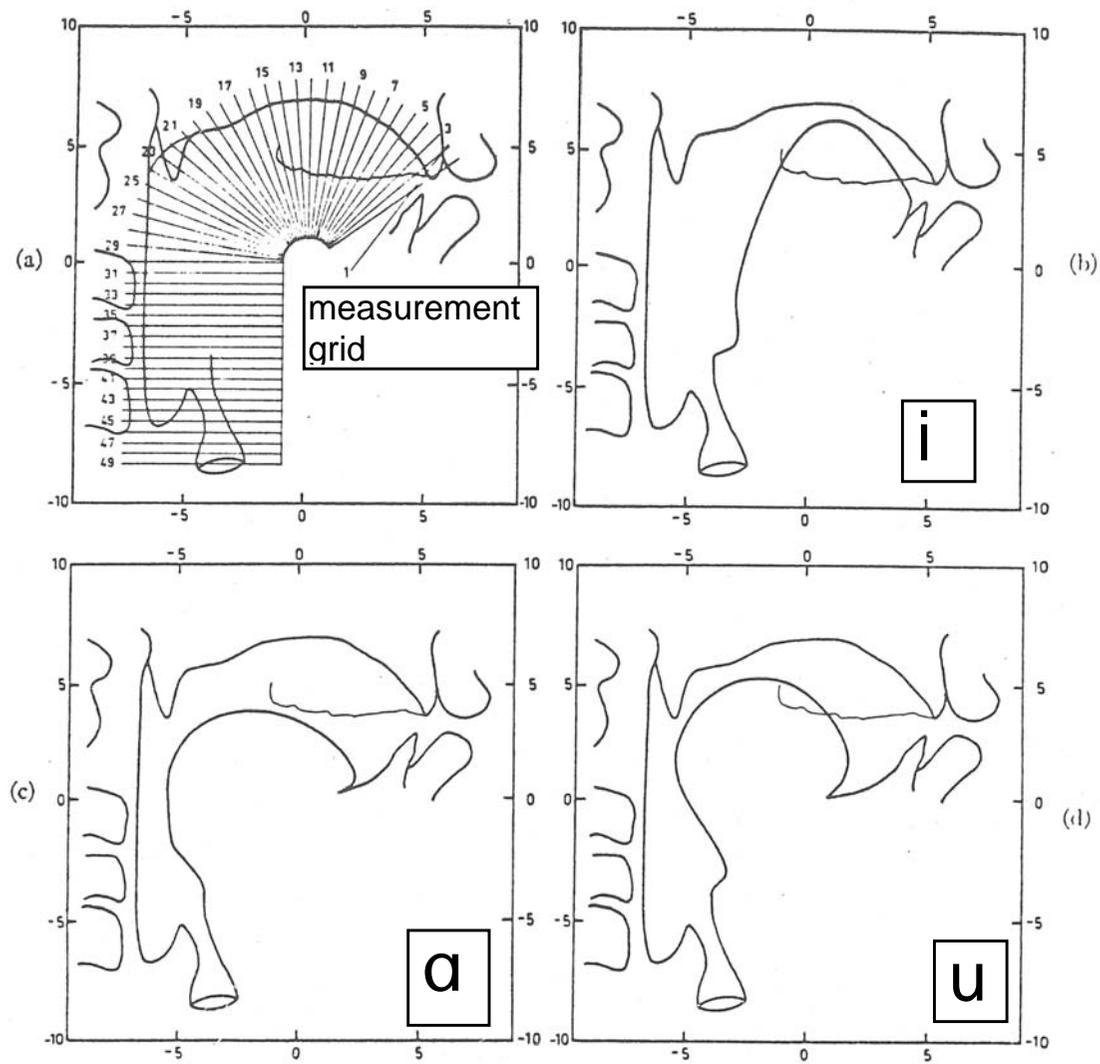
“We also observed that the production of vowel+stop consonant+vowel utterances ... seemed to involve two simultaneous gestures, viz. a diphthongal gesture of the tongue-body articulator and a superimposed constrictory gesture of the apical or dorsal articulators” (p.310)

Entwicklung des Modells

Messungen:

Vokaltraktform definiert als die Entfernung zwischen Zunge und Vokaltraktbegrenzung auf den Linien des folgenden Rasters.

====> 49 Messwerte pro Röntgenbild



Vokaltraktform (*shape function*): $s(x)$, $x = 1 \dots 49$

Das Modell als Synthesystem betrachtet:

Wie kann man die Vokaltraktformen für eine VCV-Sequenz generieren (zuerst für den Fall $V_1=V_2$)?

Wird folgende Rechenvorschrift reichen?

$$s(x) = (1-k) \cdot s_v(x) + k \cdot s_c(x) \quad (1)$$

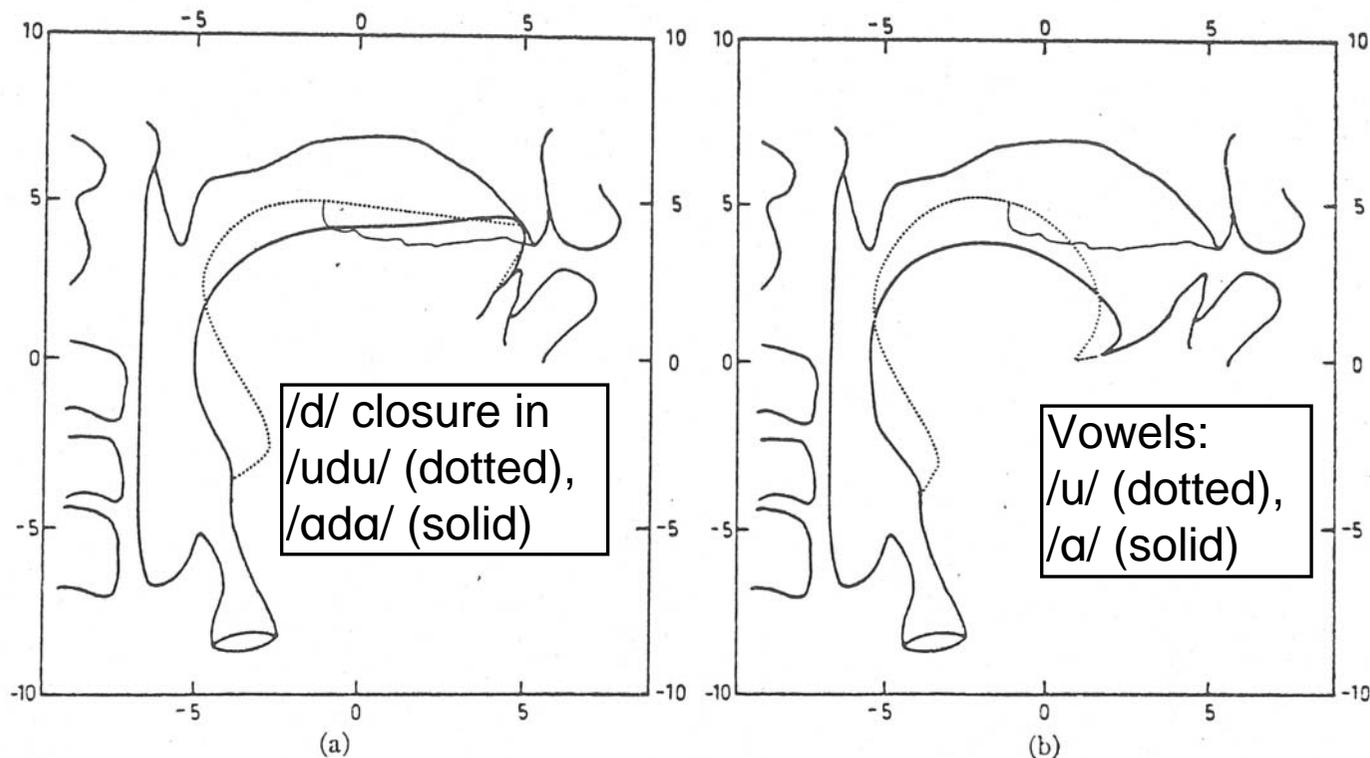
s_v, s_c : Grundform des Vokaltrakts für Vokal bzw. Konsonant

k : variiert als Funktion der Zeit von 0 nach 1 nach 0 für V - C - V

Leider nicht - Koartikulation wird nicht erfasst.

Bei $k=1$ wird s vom Vokal nicht beeinflusst.

k muss aber beim Konsonant 1 sein, sonst ist der Verschluss unvollständig.



Auf dem Weg zu einem realistischeren Modell

Neue Annahme:

Jeder Konsonant besteht aus zwei zeitinvarianten Komponenten

1. $c(x)$ “ideal target shape”
2. $w_c(x)$ “coarticulation function”; bestimmt durch ein Gewicht zwischen 0 und 1, wie stark $c(x)$ vom Vokal beeinflusst werden kann

neues Synthesemodell für symmetrische VCV-Sequenz

$$s(x) = v(x) + k \cdot [c(x) - v(x)] \cdot w_c(x) \quad (2)$$

$v(x)$: neue Schreibweise für Grundkonfiguration des Vokals

An welchen Stellen im Vokaltrakt dürfte $w_c(x)$ hohe Werte haben (ca. 1)?

Im Bereich der Artikulationsstelle.

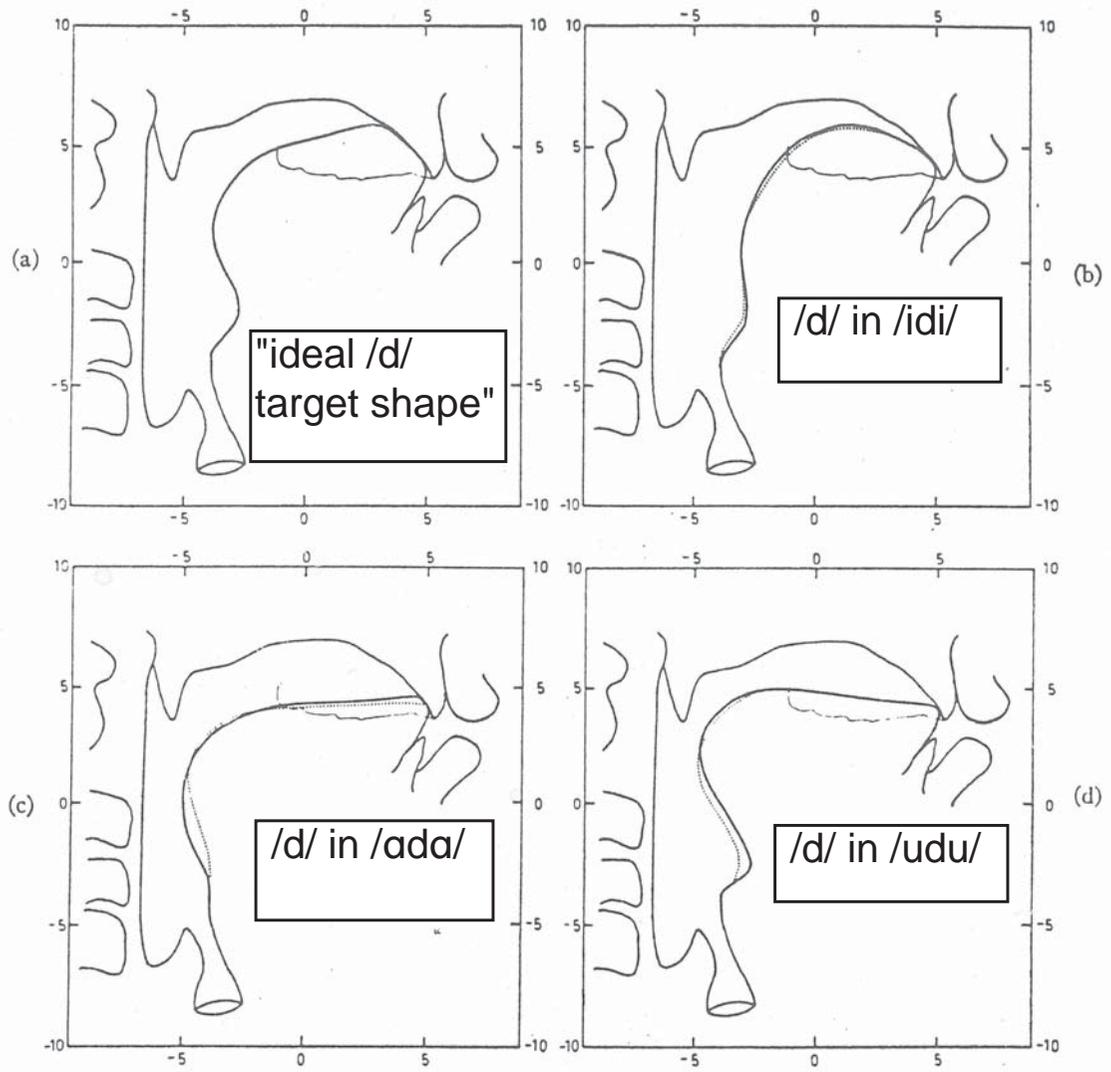
Um Verschluss zu garantieren, wenn $k = 1$

Aber wie bestimmen wir $c(x)$ und $w_c(x)$ (sie sind nicht direkt messbar)?

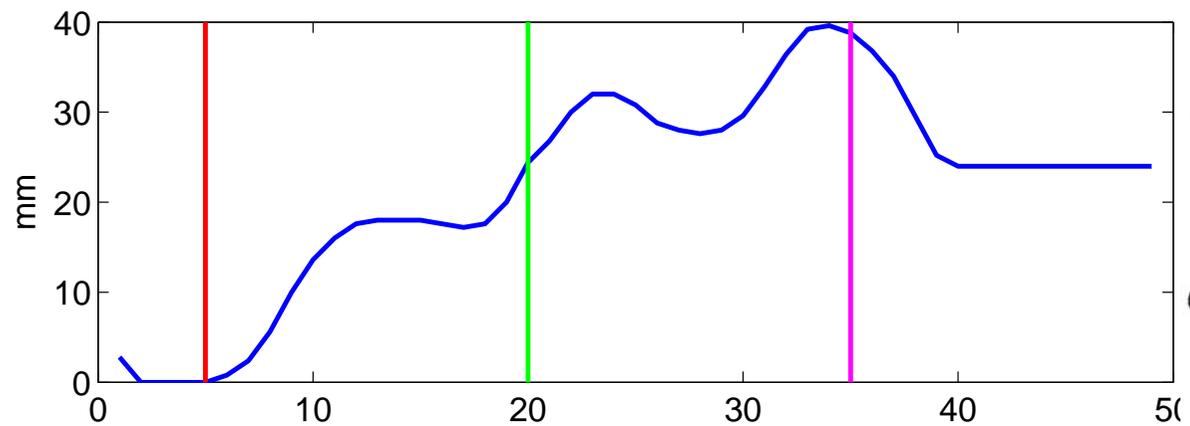
Durch Lösung von Gleichungen der Form

$${}_u d_u(x) = u(x) + [d(x) - u(x)] \cdot w_d(x)$$

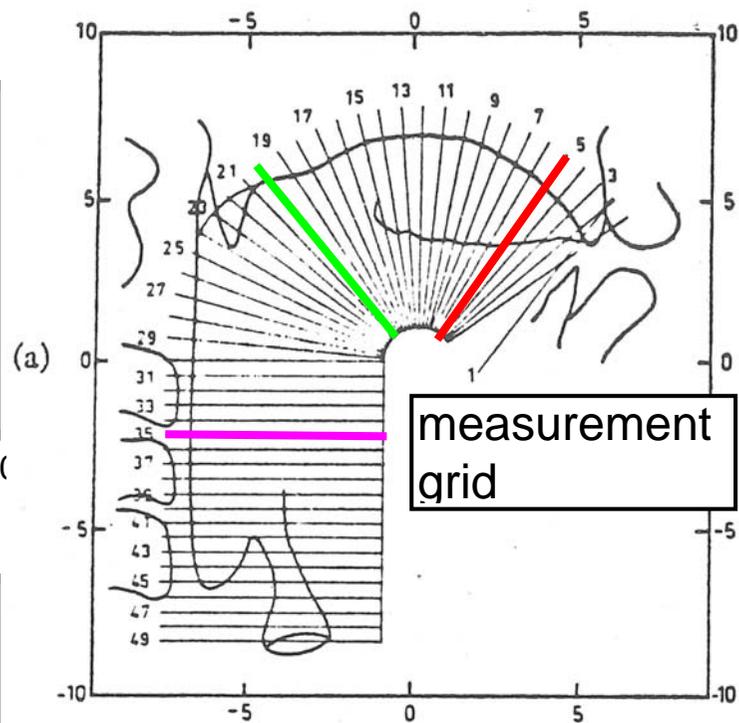
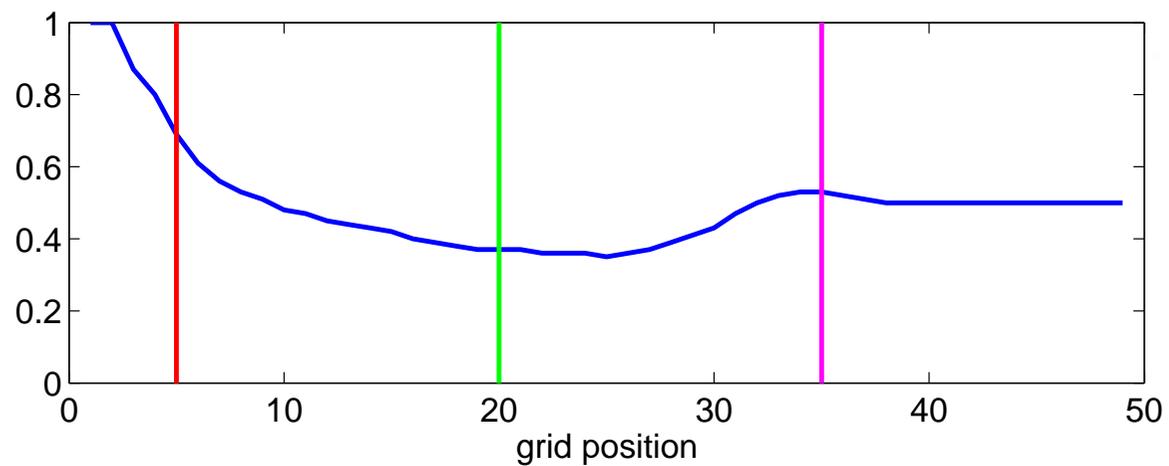
$${}_i d_i(x) = i(x) + [d(x) - i(x)] \cdot w_d(x)$$



Vocal-tract distance function for /d/



Coarticulation function for /d/



Nicht-symmetrische VCV-Sequenzen

In Gl. 2 muss v auch eine Funktion der Zeit sein: $v(x;t)$

Annahme: Alle vokalischen Konfigurationen lassen sich als gewichtete Summe der Eckvokale i , a , u darstellen.

Für das folgende Beispiel nur Fälle, wo $V1$ und $V2$ Eckvokale sind.

Dann kann man einen Vokalparameter q definieren (analog zum Konsonantenparameter k), der die "Mischung" der zwei Eckvokale angibt

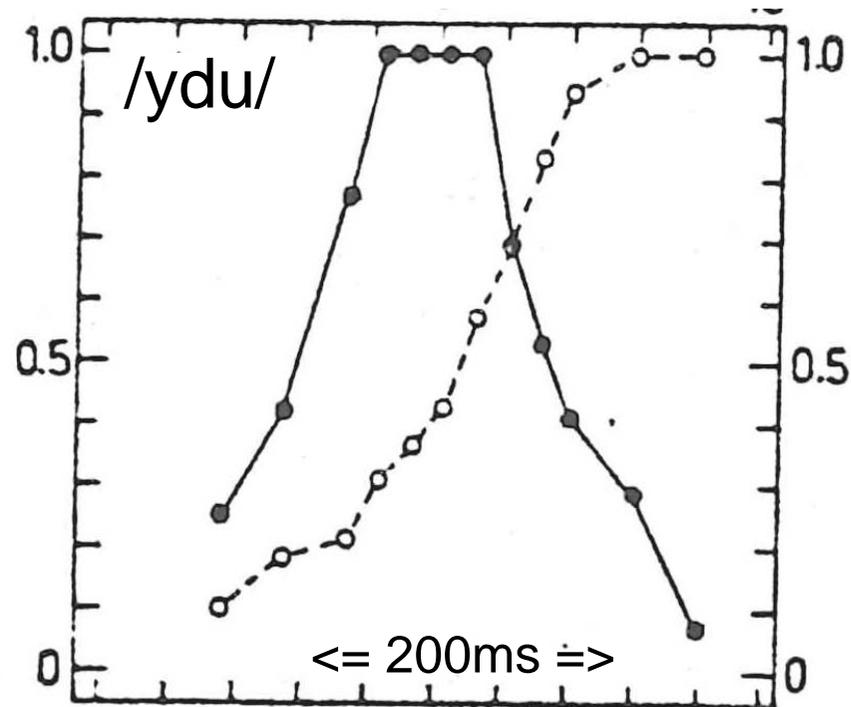
$$\text{z.B. } v(x) = q \cdot a(x) + (1-q) \cdot i(x)$$

Für eine gemessene VCV-Sequenz kann man dann für jeden Frame im Röntgenfilm den Wert von k und q bestimmen, der anhand von Gl. 2 die beste Rekonstruktion/Vorhersage der gemessenen $s(x)$ liefert.

====> k und q als Zeitfunktion

Bei Modellierung immer wichtig: Ein Maß für die Rekonstruktionsgenauigkeit.

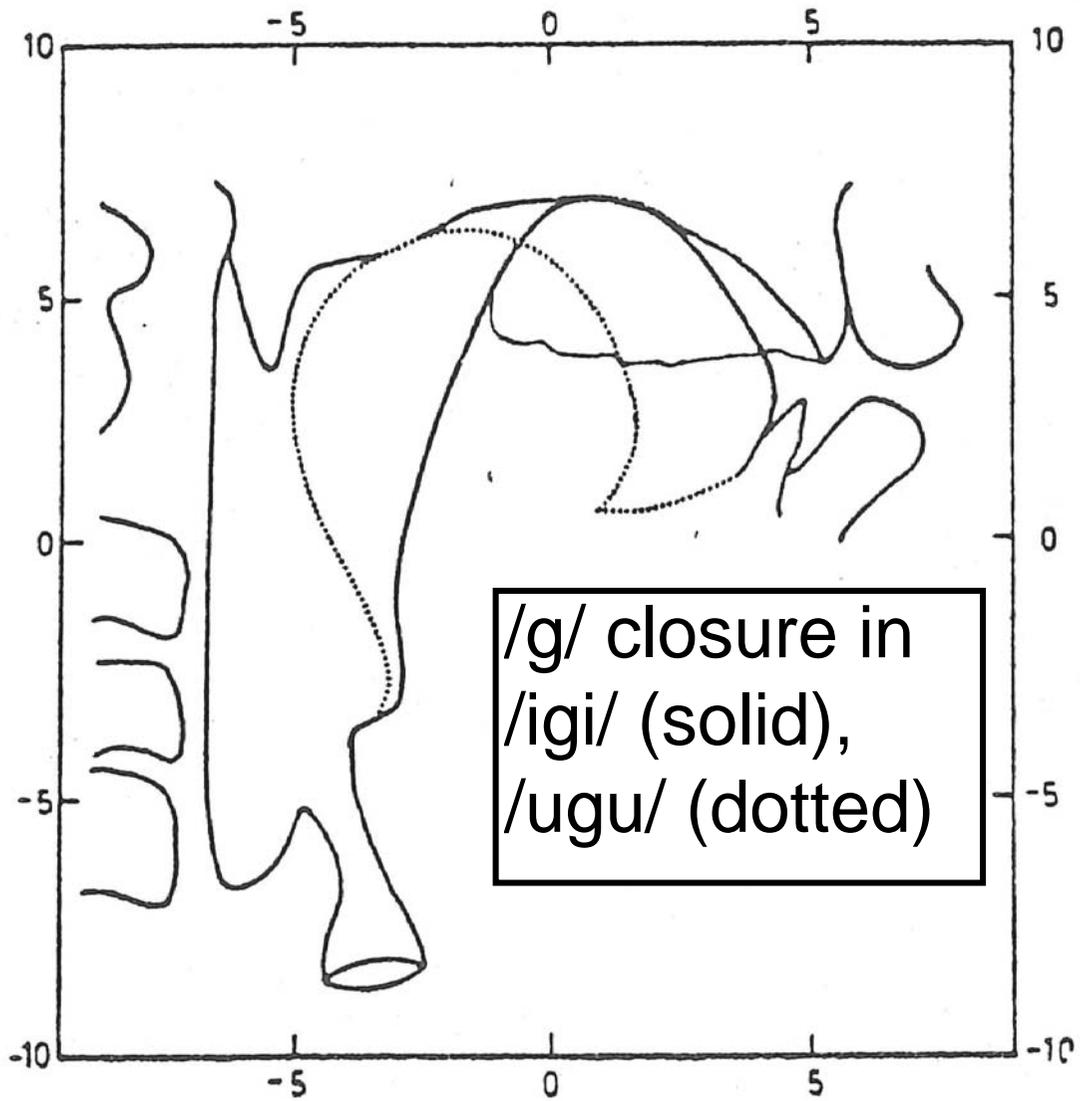
Time course of consonant parameter (solid)
and vowel parameter (dashed)



Wie ist Koartikulation hier sichtbar?

q erreicht den Extremwert für V2 erst nach der Verschlusslösung, d.h. zu diesem Zeitpunkt Einfluss von V1 noch vorhanden.

Welche Koartikulationsphänomene lassen sich mit diesem Modell nicht erfassen?



CC-Koartikulation auch nicht ohne weiteres möglich.

ad-hoc Lösung möglich.

Aber Neuformulierung des Modells mit **Artikulatoren** (nicht 'shapes') als Ausgangspunkt ev. sinnvoller (Apex, Dorsum, tongue body)

Ausblick:

Die Grundidee einer kontinuierlichen vokalischen Bewegung mit überlagerten konsonantischen Verengungen ist aber sehr einflußreich geblieben.

Beispiel: Synthesystem von B. Story

<http://sal.shs.arizona.edu/~bstory/>

Literatur

Öhman, S. E. G. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. 39, 151–168.

Öhman, S. E. G. (1967). "Numerical model of coarticulation," J. Acoust. Soc. Am. 41, 310–320.

Farnetani, E. & Recasens, D. (1999). "Coarticulation models in recent speech production theories". In W. Hardcastle & N. Hewlett (eds.) "Coarticulation. Theory, data and techniques". CUP, pp. 31-65.

Story, B. (2005) A parametric model of the vocal tract area function for vowel and consonant simulation. J. Acoust. Soc. Am. 117 (5), 3231-3254