Received 13 September 1965 9.3, 9.7

Coarticulation in VCV Utterances: Spectrographic Measurements

S. E. G. Öhman

Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden

In this paper, the formant transitions in vowel + stop consonant + vowel utterances spoken by Swedish, American, and Russian talkers are studied spectrographically. The data suggest a physiological model in terms of which the VCV articulations are represented by a basic diphthongal gesture with an independent stop-consonant gesture superimposed on its transitional portion. This interpretation necessitates a re-evaluation of the locus theory proposed by the workers of the Haskins Laboratories. Some conclusions about the general nature of the neural instructions behind the VCV gestures are discussed.

INTRODUCTION

SPEECH production involves two fundamental aspects: the stationary properties of phoneme realization and the dynamic rules governing the fusion of strings of phonemes into connected speech. Recent research has dealt successfully with the acoustic description of static speech sounds. The dynamics of articulation, however, have received much less attention, and reports of quantitative studies in this area are relatively rare. A selection of references is given below.[1–13]

The present work deals with some of the apparently very lawful rules that describe how voiced stops are coarticulated with vowels in vowel–consonant–vowel (VCV) context. As a preliminary illustration of the type of effects investigated, Fig. 1 shows a pair of VCV words chosen from the material described in more detail later. In both utterances, the initial vowel is /ø/ and the consonant is /g/. The final vowel is different. It is /y/ in the left and /ɑ/ in the right part of the Figure. The talker is a male Swede.

When the intervocalic consonant is inspected for changes that might be due to the variation imposed by the vowel context, the following observation presents itself immediately. The second-formant transition in the vowel preceding the stop is rising when the vowel following the stop is /y/, but it is falling when the vowel following the stop is /ɑ/. There is also some variation in the third-formant transition of the initial vowel. The articulatory motion from the initial vowel into the /g/, as mirrored by the formant transitions, is apparently modified by the vowel that is to *follow* the /g/.

Figure 2 shows a situation that is converse to that of Fig. 1. The final vowel was kept constant here and the initial vowel was varied. The utterance in the left part of the Figure is /ydø/ and that of the right part is /odø/. This time, there is a change in the formant transitions of the final vowel. The second-formant transition of the final vowel is falling when the vowel preceding

[1] R. H. Stetson, *Motor Phonetics* (North-Holland Publ. Co., Inc., Amsterdam, 1951).

[2] P. Menzerath and A. de Lacerda, *Koarlikulation, Steuerung und Lautabgrenzung* (Fred. Dümmlers Verlag, Berlin, 1933).

[3] L. Kaiser, "Some Properties of Speech Muscles and the Influence Thereof on Language," Arch. Néerl. Phon. Exptl. **10**, 121–133 (1934).

[4] M. Joos, "Acoustic Phonetics," Language **24**, 1–136, Chap. 6 (1948).

[5] B. Lindblom, "Spectrographic Study of Vowel Reduction," J. Acoust. Soc. Am. **35**, 1773–1781 (1963).

[6] A. S. House and K. N. Stevens, "Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study," J. Speech & Hearing Res. **6**, 111–128 (1963).

[7] P. F. MacNeilage, "Electromyographic and Acoustic Study of the Production of Certain Final Clusters," J. Acoust. Soc. Am. **35**, 461–463 (1963).

[8] S. E. G. Öhman and K. N. Stevens, "Cinéradiographic Studies of Speech: Procedures and Objectives," J. Acoust. Soc. Am. **35**, 1889(A) (1963).

[9] O. Fujimura, "Motion-Picture Studies of Articulatory Movements," Mass. Inst. Technol. Res. Lab. Electron. Quart. Progr. Rept. No. 62, 197–202 (15 July 1961).

[10] Y. Kato, "Analysis of Liquid Consonants," J. Acoust. Soc. Am. **36**, 1989(A) (1964).

[11] S. Inomata, "Program for Active Segmentation and Reduction of Phonetic Parameters," Paper **B7** in *Proceedings of the Speech Communication Seminar, 1962* (Speech Transmission Lab., KTH, Stockholm, 1963), Vol. 1; J. Acoust. Soc. Am. **35**, 1112(A) (1963).

[12] P. Ladefoged, "The Nature of General Phonetic Theories," paper presented at 16th Ann. Round Table Meeting on Linguistics & Language Studies, Georgtown Univ. (26 Mar. 1965).

[13] E. Fischer-Jørgensen, "Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck," Z. Phonetik **16**, No. 1–3, 19–36 (1963).

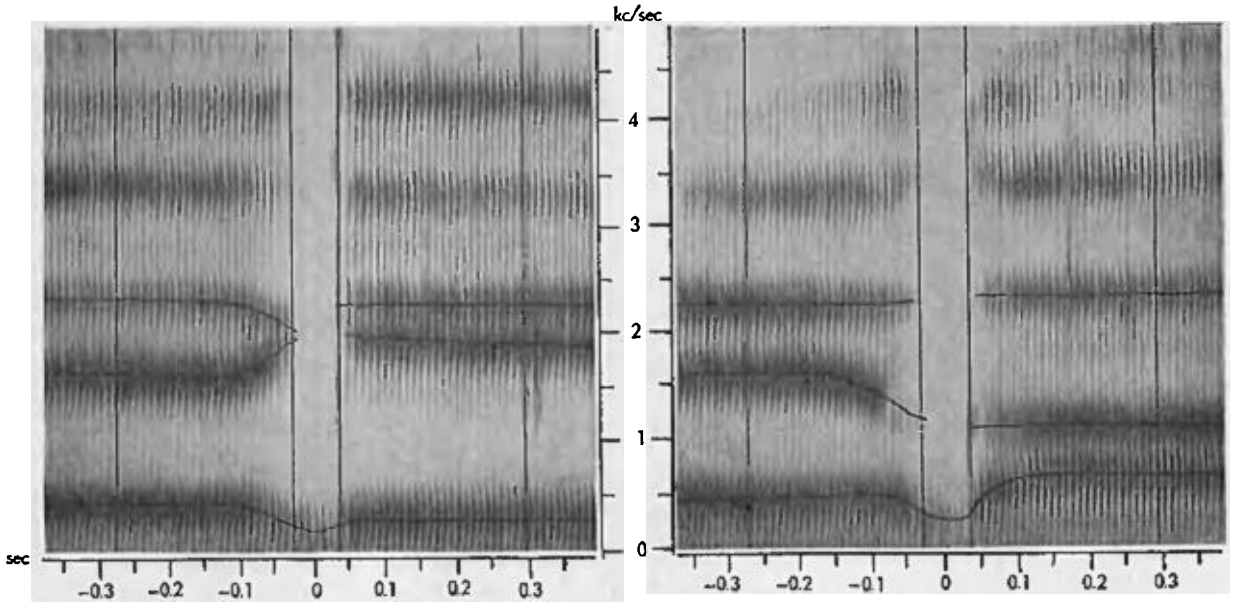

FIG. 1. Sound spectrograms of the utterances /øgy/ (*left*) and /øgɑ/ (*right*) as spoken by a male Swedish talker. The formant transitions in the initial vowel are different in the two cases, owing to influence of the final vowel. The lines superimposed on the spectrograms indicate method of measurement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

the stop is /y/, but it is slightly rising when the vowel preceding the stop is /o/. Thus, the initial vowel influences the medial stop-to-final vowel transition *across* the intervocalic consonant.

In order to obtain a more detailed picture of the interaction of the vowel and the stop-consonant gestures in VCV context, a series of studies has been made involving speakers of Swedish, English, and Russian.

I. F_2 TRANSITIONS: INVESTIGATOR'S SPEECH

A. Measurements

A list of VCV "words" was constructed using the three voiced stops /b/, /d/, /g/ and the four vowels /y/, /ø/, /ɑ/, and /u/. These are Swedish phonemes.[14] The list contained 4×3×4=48 different words.

This material was read three times by the author in an anechoic chamber. A Brüel & Kjær condenser microphone and microphone amplifier, and an Ampex 300 tape recorder were used. The system was adjusted to give an essentially flat response over the 20- to 10 000-cps frequency band.

The words were read in a monotone and both syllables were equally stressed. All vowels were phonemically long. Wide-band spectrograms were then made of all the words recorded, using a Kay Electric Company Sona-Graph with an expanded frequency scale (0–4.500 cps approximately). The frequencies of the second formant were measured in the stationary parts of the initial and final vowels and at the beginning and end of the closure of the stop consonants. Thin lines were drawn with a pencil so as to follow the center of the second-formant bars as seen in the spectrograms and vertical lines were drawn at the beginning and end of the stop closure. The second-formant frequencies were measured where the vertical lines intersected the formant lines. Frequencies measured by this method are accurate within less than 50 cps (Ref. 15).

The beginning and end of the stop-closure intervals were determined with the aid of a number of criteria derived from the acoustic theory of stop-consonant production.[16] The beginning of the closure is associated with a substantial decrease in amplitude of the second and higher formants and a downward frequency shift in the first formant. The end of the closure is usually marked by a weak-burst spike and a rapid increase in the frequency of the first formant. In some cases where it is impossible to assess the location of the second formant at the closure boundaries, an extrapolation rule was adopted that stated that the formant was assumed to move into the stop gap without change in slope from its nearest observable value in the vowel. The vertical lines of Fig. 1 exemplify the segmentation procedure.

[14] /b/ and /g/ closely resemble the English /b/ and /g/, but the Swedish /d/ is apico dental rather than apico alveolar. The vowels /y/ and /ø/ are, by and large, lip-rounded versions of the English /i/ and /e/, respectively. /ɑ/ is intermediate between American English /ɑ/ and /ɔ/. The Swedish vowels /y/ and /u/ have a narrower tongue constriction than the corresponding English vowels /i/ and /u/.

[15] B. E. F. Lindblom, "Accuracy and Limitations of Sona-Graph Measurements," in *Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki 1961* (Mouton & Co., 's-Gravenhage, 1962), pp. 208–213.

[16] G. Fant, *Acoustic Theory of Speech Production* (Mouton & Co., 's-Gravenhage, 1960), p. 169.

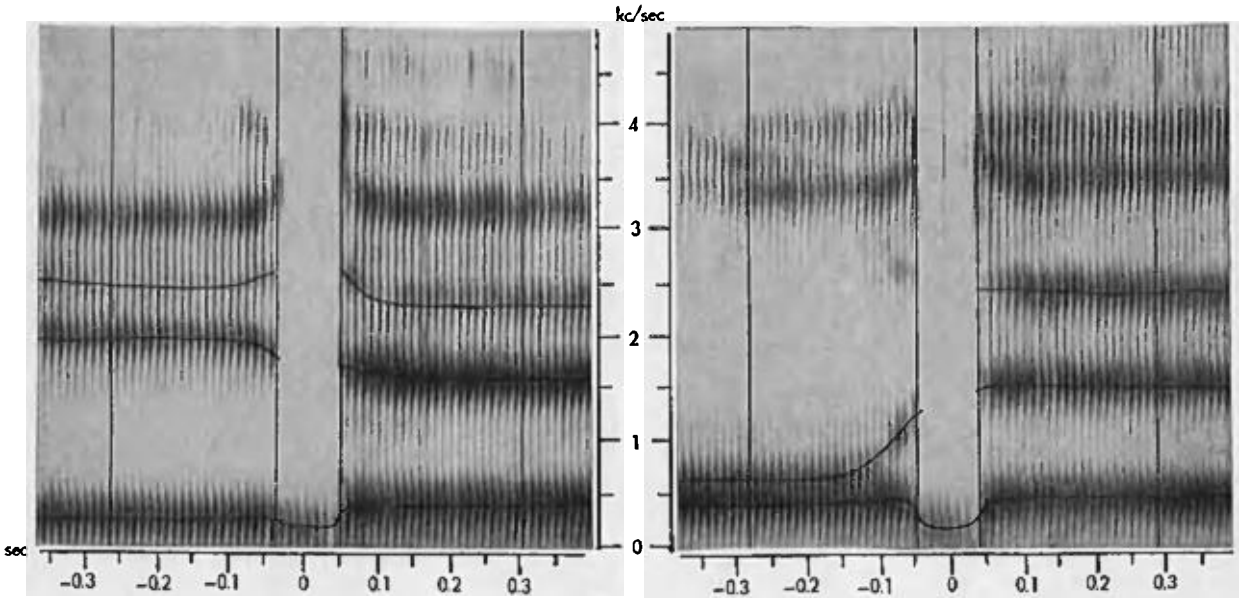

FIG. 2. Sound spectrograms of the utterances /ydø/ (*left*) and /ødø/ (*right*) as spoken by a male Swedish talker. The formant transitions in the final vowel are different in the two cases, owing to influence of the initial vowel.

B. Results

The results are summarized in Table I. Here is shown the differences between the terminal frequencies of the second-formant transitions, measured at the closure boundaries, and the frequencies of the same formant in the stationary parts of the vowels. To read the Table, note that the columns headed by VbV, VdV, and VgV have been split into two halves denoted by VC and CV. The numbers in a VC column represent the extent of a second-formant transition from an initial vowel into a stop consonant when the vowel following the stop is as indicated in the second column of the Table. The initial vowel is found in the first column of the Table. Similarly, the numbers in a CV column represent the extent of a second-formant transition into a final vowel when the vowel preceding the stop is as indicated in the second column of the Table. In this case, the first column of the Table indicates the final vowel. A number is positive if the terminal frequency of the formant is higher than its value in the stationary part of the vowel, and negative if lower. Each number in the Table is an average of three measurements. The average stationary second-formant frequencies of the vowels /y/, /ø/, /ɑ/, and /u/ for this speaker are, respectively, 2000, 1600, 1000, and 650 cps.

TABLE I. Data pertaining to investigator's second-formant transitions in VCV utterances: Each number describes the difference $F_{2t}-F_{2v}$ in cps, where F_{2t} denotes the terminal frequency of the formant transition measured in the neighborhood of the stop gap, and F_{2v} denotes the frequency of the formant as measured in the most stationary part of the vowel. The vowel in which the transition is observed is indicated in the leftmost column of the Table. The vowel on the opposite side of the stop is shown in the second column from the left. Each number is based on three measurements.

		VbV		VdV		VgV	
		VC	CV	VC	CV	VC	CV
y	y	−450	−350	−155	−130	−15	+140
	ø	−515	−370	−290	−200	−30	+55
	ɑ	−775	−370	−370	−280	−255	+115
	u	−650	−390	−200	−190	−325	+100
ø	y	−375	−65	−90	+75	0	+430
	ø	−415	−175	−200	−5	−25	+375
	ɑ	−465	−90	−340	−25	−275	+340
	u	−515	−140	−240	+40	−365	+315
ɑ	y	+100	+205	+265	+520	+175	+330
	ø	+25	+75	+240	+365	+125	+225
	ɑ	−100	−100	+165	+290	+30	+165
	u	−150	−55	+250	+315	−65	+80
u	y	+200	+255	+575	+940	+190	+115
	ø	+80	+325	+530	+840	+65	+255
	ɑ	+15	+200	+490	+370	+15	+190
	u	+5	−5	+490	+690	−50	−15

Table I shows that the second-formant transition in every VC sequence studied is variable and dependent on the formant pattern of the final vowel of the VCV utterance. In the sequence /ag/, for example, the second formant rises by 175 cps when the final vowel is /y/, and falls by 65 cps when the final vowel is /u/.

Similarly, in all the CV sequences, except /by/, there is a different second-formant transition for each preceding vowel. An example is /ba/, in which the second formant falls by 205 cps when the initial vowel of the VCV utterance is /y/ and rises by 100 cps when the initial vowel is /o/. The ranges of coarticulatory variation of the F_2 terminal frequencies in the VC and CV sequences are quite large, sometimes as large as 380 cps, the average being 210 cps.

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		ɑ		o		u	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
VbV										
F_1	290	330	400	430	670	660	400	420	340	360
	(40)	(50)	(50)	(40)	(40)	(65)	(40)	(50)	(25)	(40)
F_2	2000	1890	1650	1580	990	1040	670	730	670	680
	(50)	(140)	(75)	(90)	(75)	(100)	(50)	(50)	(40)	(40)
F_3	2350	2320	2320	2380	2670	2620		2450		2430
	(100)	(75)	(50)	(75)	(100)	(125)		(90)		(100)
VdV										
F_1	270	300	390	400	660	650	390	410	330	320
	(45)	(35)	(40)	(60)	(45)	(65)	(40)	(35)	(40)	(45)
F_2	1990	1950	1620	1580	970	1030	690	760	660	720
	(75)	(75)	(70)	(55)	(95)	(100)	(40)	(45)	(40)	(75)
F_3	2380	2350	2350	2410	2740	2700	2530	2560		2500
	(100)	(75)	(65)	(90)	(125)	(125)	(175)	(175)		(150)
VgV										
F_1	320	330	420	420	660	650	430	430	360	360
	(40)	(65)	(45)	(55)	(55)	(80)	(45)	(35)	(50)	(55)
F_2	2010	1960	1650	1650	990	1160	780	840	770	790
	(75)	(90)	(50)	(85)	(90)	(125)	(45)	(90)	(40)	(45)
F_3	2440	2340	2440	2350	2560	2480	2560	2450	2480	2460
	(75)	(100)	(115)	(75)	(150)	(165)	(115)	(145)		(115)

II. F_1, F_2, AND F_3 TRANSITIONS: MALE SWEDISH TALKER

A. Measurements

In order to avoid any bias due to the use of the investigator's own speech, it was decided to repeat the measurements using a phonetically untrained speaker who was ignorant of the purpose of the experiment. The recording equipment was the same as before, but the word material was somewhat enlarged. The VCV word ensemble consisted of all combinations of the voiced stops /b/, /d/, and /g/ with the vowels /y/, /ø/, /ɑ/, /o/, and /u/ in initial and final position. The total number of different words was thus $5\times3\times5=75$. Each of the words was written on five 2×3-in. file cards. The cards were sorted into three decks containing the VbV, VdV, and VgV words, respectively, and finally each deck of cards was shuffled. The three decks contained 125 randomly ordered words each.

The speaker read the words from the card decks in the anechoic chamber, making a pause of about 5 sec between each word pair. He was told to keep the pitch as monotone as possible, to give equal stress to the initial and final vowels, and to make the vowels long. The subject was a relatively low-pitched male speaker of a common Stockholm variety of Swedish. He experienced no difficulty in reading the words in accordance with the instruction.

Spectrograms were made in the same way as before and so were the measurements, expect that this time not only the second but also the first and third formants were recorded.

B. Results

The presentation of the results is divided into two parts: the first deals with the stationary segments of the vowels; in the second, the transitions are discussed.

1. Vowel Formants

The formant frequencies of the stationary parts of the vowels are presented in Table II. The data are displayed so as to show the small formant-pattern differences that seem to be associated with the position (initial or final) of the vowel in a VCV word as well as with the identity of the intervocalic consonant.

The physiological nature of these variations cannot be inferred with perfect certainty from the acoustic data. The fact that the formant frequencies of /y/, /ø/, and /u/ in final position seem to approach more central values than those of the same vowels in initial position indicates, however, that the final vowels are somewhat neutralized. This trend is not equally clear in the case of /o/ and /u/.

There is a small influence of the consonants on the vowel-formant frequencies in the stationary portions of the vowels. This influence is observable both in initial and final positions. Thus, the measured frequencies of the second and the third formants of any given vowel preceded by /g/ are consistently somewhat different from the values obtained when /b/ or /d/ precedes that vowel. A difference is also seen in the formant frequencies of /o/ and /u/ before /g/ as compared with the values obtained when these vowels are followed by /b/ or /d/. Figure 3 shows the F_2 distributions for /u/ and /o/

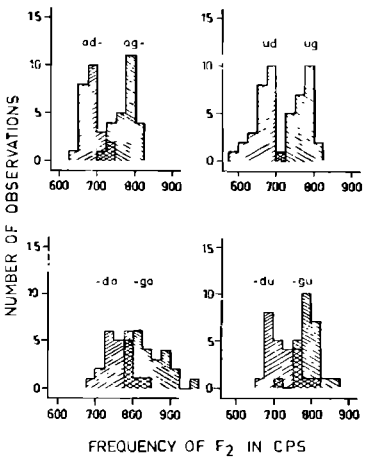

FIG. 3. Histograms representing the distribution of the second-formant frequencies of the vowels /o/ (*left*) and /u/ (*right*) in initial (*upper*) and final (*lower*) positions of VCV utterances. The two histograms in each portion of the Figure correspond to different medial-consonant contexts (/d/ and /g/). The mean second-formant frequency of any one of the vowels is lower if the medial consonant of the VCV utterance is /d/ than if it is /g/.

before and after /g/ and /d/ as measured in the VCV utterances.

The numbers given in parenthesis below the formant-frequency numbers in Table II represent dispersions. The dispersion measure used is the "interquintile range," which is the length of an interval along the frequency variable that embraces 60% of the distribution and to either side of which half (i.e., 20%) of the remaining part of the distribution is contained.[17] The formant frequency and dispersion numbers are based on 25 datum points in all cases, except those pertaining to the third formant of /o/ and /u/, which sometimes could not be measured. The formant frequency and dispersion numbers have been rounded off to the nearest 10 and 5 cps, respectively.

The dispersion data show that the variation increases with increasing formant number but is about equal for the same formant of initial and final vowels. The averages are shown in Table III.

TABLE III. Average interquintile ranges of vowel formant frequencies.

Formant	Initial vowel (cps)	Final vowel (cps)
F_1	45	50
F_2	60	80
F_3	105	110

The formant-frequency variability of each of the five vowels as averaged over all positions and contexts and without distinction of formant numbers are presented below.

y	ø	ɑ	o	u	
72.5	60.3	96.0	71.0	67.0	(cps).

Although the average interquintile range of a formant of the vowel /ɑ/ was higher than that of the remaining vowels, it cannot be concluded from this that the /ɑ/ was articulated with less precision, for a large part of the difference is probably due simply to the great sensitivity of F_2 of /ɑ/ to small fluctuations in the place of constriction of the vocal tract. The nomograms published by Fant[18] relating formant frequencies to place of constriction in four-section straight-tube models for vowels show that the F_2 gradient is maximal in the constriction region corresponding to a good /ɑ/ model. It is likely that a similar effect accounts for most of the F-pattern variability of the present speaker's /ɑ/.

In summary, it can be said that the formant-pattern variations in the stationary portions of the vowels in each subdivision of Table II are small. This conclusion is also confirmed by informal auditory analysis in which the vowels of the words recorded were easily identified with those intended by the speaker.

2. *Formant Transitions*

The formant-transition data are presented in Table IV, which is organized in the same way as Table I. A few examples will suffice to illustrate the reading of Table IV. Thus, for instance, the extent and direction of the third-formant transition of the final vowel of the word /ydɑ/—i.e., of the CV portion of /ydɑ/—is found in the CV part of the F_3 column in the VdV section. To find the row, note that the leftmost column of the Table indicates the vowel in which the formant transition is observed and that the transconsonantal vowel is indicated in the second column from the left. Since, in the present case, we are concerned with transitions in the final vowel, it is the initial vowel (/y/) that is transconsonantal and the final vowel (/ɑ/) is the "observed" vowel. The entry that we are interested in is thus found in the first (/y/) row of the /ɑ/ part of the first column. The entry in question is −20, which means that the starting frequency of F_3 in /ɑ/ of /ydɑ/ is, on the average, 20 cps lower than its frequency in the stationary part of the /ɑ/. In the same way, we find, for instance, that the F_2 transition of /ɑ/ in /ɑgø/ terminates at a frequency 400 cps higher than that of the stationary F_2 of /ɑ/, etc. The empty entries of the Table represent formant transitions that could not be measured from the spectrograms.

Table IV shows that there is substantial variation in the formant transitions of any given VC or CV sequence, depending on the identity of the vowel on the opposite side of the stop. This can be seen by selecting any vowel in the observed vowel column and comparing the five formant-frequency values found in the intersection of the row for this vowel and any column of the interior of the Table. The largest difference between any pair of these five values may be called the *range of variation* of the associated VC or CV formant transitions.

The ranges of variation of all the VC and CV transitions are given in Table V. The maximum ranges for transitions of F_1, F_2, and F_3 are here seen to be, respectively, 130, 540, and 330 cps, and the average ranges are

[17] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, N. J., 1946), p. 213.

[18] Ref. 16, p. 72 seq.

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.

Observed vowel	Transconsonantal 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	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
	ɑ	−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
	ɑ	−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
ɑ	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
	ɑ	−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
	ɑ	−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
	ɑ	−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	···	···

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_3	
	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
a	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	···	···

62, 280, and 153 cps, respectively. In other words, the terminal frequency of, e.g., the second formant in a fixed vowel+stop or a stop+vowel sequence may be expected to vary by 280 cps, on the average, owing to coarticulation with transconsonantal vowels.

The degree of statistical significance of these data can be determined by means of Fig. 4, which shows histograms representing the deviations of the terminal frequencies of the individual transitions from their associated mean values. The fitted curves show that the distributions are approximately normal with standard deviations 35, 75, and 70 cps, respectively. The difference of the means of two samples each containing n elements drawn from some one of these distributions will approach a normal distribution as n becomes large. The approximate standard deviation for a sample of n elements is $s'=s(2/n)^{\frac{1}{2}}$, where s is the standard deviation of the original distribution. Using this formula with $n=5$, we can calculate the s' corresponding to the first-, second-, and third-formant transitions. They are, respectively, 21, 48, and 45 cps. The critical differences at the 0.01 and 0.05 levels of significance are shown in Table VI.

TABLE VI. Critical ranges of variation of formant transitions at two significance levels.

	Level of significance	
Formant	0.01	0.05
F_1	57	44
F_2	118	93
F_3	115	87

In other words, if the terminal frequencies of F_2 or F_3 of a given CV or VC sequence of Table IV change by more than about 100 cps from one transconsonantal vowel condition to another, then the difference is almost certainly not due to chance. The same holds for F_1 when the difference is greater than only about 50 cps. In Table V, those ranges that exceed the critical differences have been printed in boldface. The average ranges for the terminal frequencies of F_1, F_2, and F_3, which, as noted earlier, are 62, 280, and 153 cps, respectively, are all significant. In particular, the range of the terminal frequency of F_2 is significant in every VC or CV sequence except /−gy/, where it is only 80 cps. It is interesting that the largest range observed in all of Table V pertains to the second-formant transition of the reverse of /−gy/—i.e., to /yg−/ for which it is 540 cps. A similar relation holds between F_2 of /øg−/ and /−gø/.

50% of the F_3 ranges of Table V and also 50% of the F_1 ranges are greater than the critical values. In the case of /u/, the ranges of the terminal frequencies of all three formants are significant under all conditions. In general, Table V shows that the terminal frequencies of the formant transitions in a vowel+stop or stop+vowel sequence in the speech material studied here are not unique but depend on the F-pattern of the transconsonantal vowel. This is an agreement with the results summarized in Sec. I-B.

In order to emphasize further the complexity of the stop-consonant formant transitions, we discuss certain aspects of the present data in more detail in Secs. II-C, D, and E. In Secs. III and IV, we examine transition data from an American-English and a Russian talker, respectively.

C. Nonuniqueness of Terminal Frequencies

Parts of the data of Table IV have been plotted in Figs. 5 and 6. Figure 5 shows all the second-formant transitions observed in the sequences /øC-/ (left part) and /-Cø/ (right part). The second formant of /ø/ has been represented by a thick horizontal bar and the slant lines converging upon it represent the formant transitions. No attempt has been made to indicate the correct time relations.[19] The symbols along the vertical lines

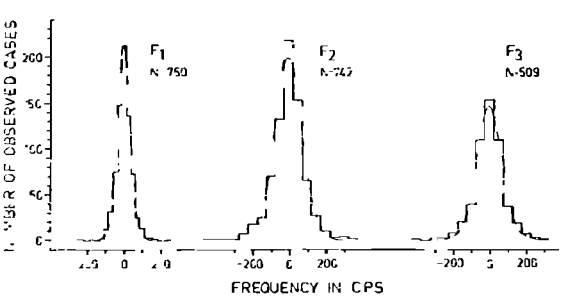

FIG. 4. Histograms representing the deviations of the terminal frequencies of individual formant transitions from their associated mean values. The superimposed curves are the best-fitting normal distributions.

[19] Measurements on the durations of the formant transitions are presented in Roy. Inst. Technol. (Stockholm) Speech Transmission Lab. Quart. Progr. Stat. Rept. 1/1965 (15 Apr. 1965).

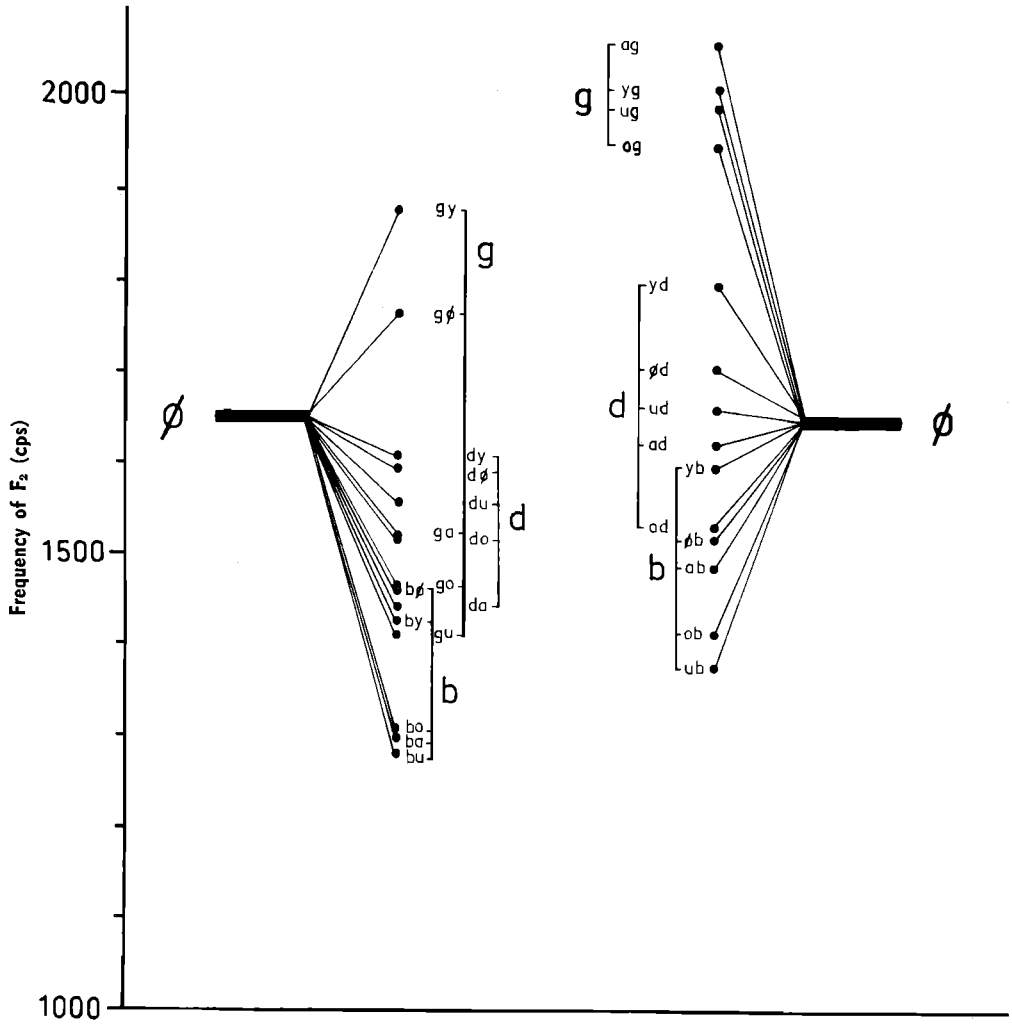

FIG. 5. Stylized second-formant transitions observed in the sequences /øC-/ (*left*) and /-Cø/ (*right*). The letters at the terminal points identify the medial consonant and transconsonantal vowel of the VCV utterance in which the transition was observed. The vertical lines cover the ranges of variation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

facing the formant transitions are used to identify the consonant and transconsonantal vowel with which each transition line is associated. The vertical lines cover the frequency ranges of the transitions corresponding to the three stops /b/, /d/, and /g/.

The vertical lines in the left part of Fig. 5 show that the /d/-transition range overlaps completely with the /g/-transition range; i.e., there is no such thing as a unique /d/-transition terminal frequency of the second formant of an /ød-/ sequence for this talker. There is even a frequency region around 1450 cps such that, if a second-formant transition from /ø/ terminates there, it may be associated with any one of the three stops /b/, /d/, and /g/. Here is, accordingly, a case in which the second-formant terminal frequency alone carries no information whatsoever about the place of articulation of the stop.

The right part of Fig. 5 shows a relatively well-defined /g/-transition range, but again the /d/ and /b/ ranges overlap.

The most striking feature of Fig. 5 is perhaps the relatively even spread of the terminal frequencies as one moves along the frequency dimension of the diagram from about 1350 to about 2100 cps. There are no fixed frequencies corresponding to the three stops. This is true not only of the case illustrated in Fig. 5, but of the whole material studied in this experiment (see Table IV).

The overlap of the terminal frequency regions of Fig. 5 can be reduced by taking account of the third formant. This is seen in Figs. 6(a) and (b), which are $F_2 - F_3$ plots of the data presented in Fig. 5. Here, the vowel /ø/ is represented by a point at $F_2 = 1650$ cps and $F_3 = 2400$ cps. The second- and third-formant transition pairs are indicated by straight lines originating from the point representing the vowel /ø/ and terminating at points representing the terminal frequencies of the two formants. The left part of the Figure shows the transitions of the /øC-/ sequences and the right part those of the /-Cø/ sequences. The labels at the terminal points identify the stop and the transconsonantal vowel. The insets show the stylized time course of the F_2 and F_3 transitions of the four quadrants about the vowel target point.

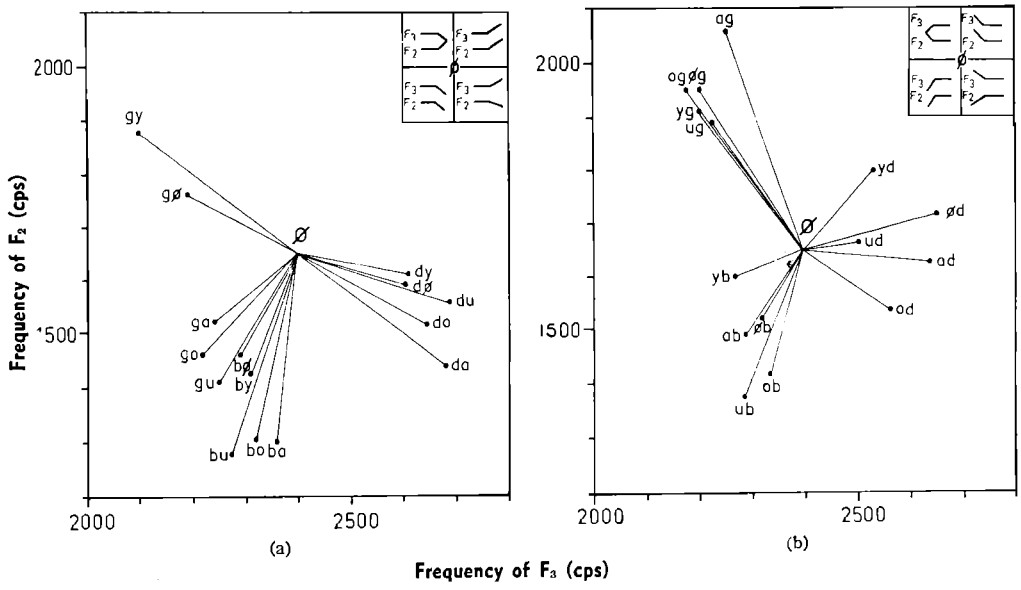

FIG. 6. (a) F_2 vs F_3 plot of the formant transitions shown in the left part of Fig. 5. The inset in the upper right corner shows the approximate course of the formant transitions represented by the points in the four quadrants about the location of /ø/ in the F_2-F_3 plane. (b) F_2 vs F_3 plot of the formant transitions shown in the right part of Fig. 5.

In Fig. 6(b), all the /b/ transitions are located in the lower-left quadrant, the /g/ transitions in the upper left quadrant, and the /d/ transitions in the upper and lower right quadrants. In this plot, unlike that of Fig. 5, there is, hence, no overlap between the /b/ and /g/ transitions. In Fig. 6(a), however, some of the /g/ transitions occupy the same quadrant as the /b/ transitions. Thus, for instance, the F_2 and F_3 terminal frequencies of the transitions in the initial vowels of /øbø/ and /øgu/ are closely similar.

D. Temporal Features of the Formant Transitions

A qualitative survey of the entire set of data for this speaker is given in Figs. 7–9, which represent spectrograms of VCV words containing the intervocalic stops /b/, /d/, and /g/, respectively. In these Figures, the words in any row have the same final vowel and the words in any column have the same initial vowel. Each of the "spectrograms" is an average of the five readings of each word obtained by tracing on transparent paper the center lines of the formants that are seen on the sound spectrograms. The five superimposed tracings were then used to get a visually determined average corresponding to each word. The accuracy of this method is not much inferior to that involving the statistical calculations and has the advantage of being much faster. The marks along the frequency scales of the graphs are 500 cps apart and the time marks are 100 msec apart.

By reading across rows or columns in Figs. 7–9, one can observe the successive shifts of locus frequencies that are due to the more or less marked influence of the transconsonantal vowel. Several other qualitative observations can also be made. A comparison of /ybo/ and /ygo/, for instance, shows that the CV parts of these two words are practically identical with respect to formant transitions. To judge from the graphs and the spectrograms, the only cues for the distinction are (1) the lower F_2 and F_3 terminal frequencies of the initial vowel of /ybo/ as compared to those of /ygo/, and (2) the different rates of change of the VC transitions of the two words, F_2 being slower and F_3 faster in /yb/.

Another interesting feature is the peculiar **S**-like shape of the F_2 transition in the final vowel of some of the VgV displays. This characteristic is particularly clear in words that end in a front vowel (/y/ or /ø/) and begin with a back vowel (/u/ or /o/)—i.e., in /ogy/, /ugy/, /ogø/, and /ugø/ (upper right corner of Fig. 9). The F_2 transition of the final vowel of /ogø/, for instance, starts with a rise that lasts for about 40 msec. The formant then falls down toward the target frequency for /ø/.

A tentative explanation of this phenomenon may be given with reference to Fant's three-parameter model for simple articulatory constrictions of the vocal tract.[20] Figure 10, which is derived from this model, shows the frequencies of F2 and F3 for a constant lip-section area (65 mm²) and a constriction of fixed minimum cross-sectional area (65 mm²) located at various places along the tract. The position x of the point of minimum cross section is measured by its distance in centimeters from the glottis. For a mediopalatal constriction ($x=13$ cm), F_2 has a maximum. In a back vowel + /g/+ front vowel utterance, the /g-/ closure may be assumed to start with a relatively posterior point of constriction corresponding approximately to the point $x=8$ cm of Fig. 10. This is suggested by the low-terminal F_2 frequency of the initial vowel. During the closure, the configuration of the final vowel is anticipated and this anticipation causes the point of constriction to move

[20] Ref. 16, p. 71.

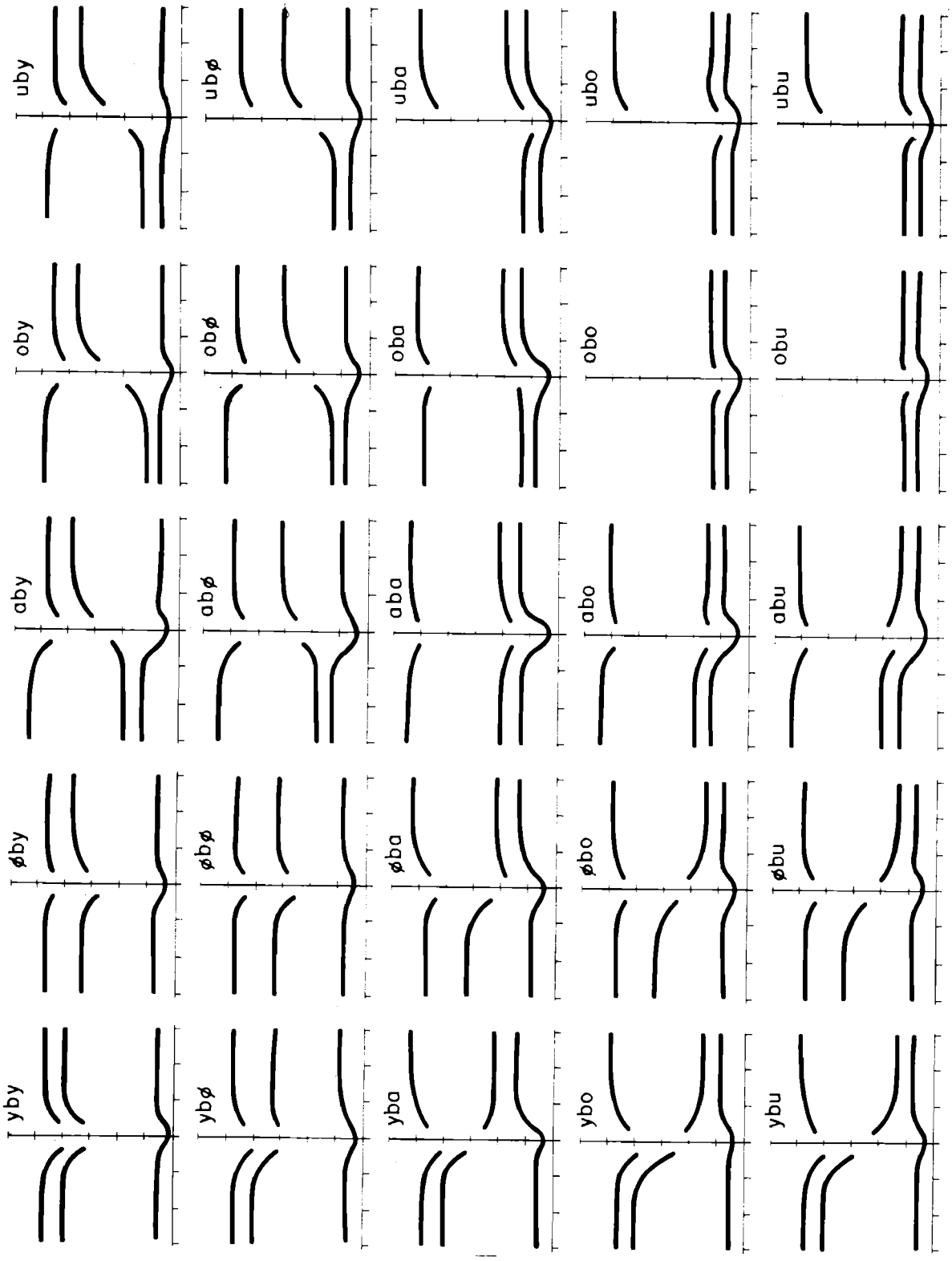

FIG. 7. Averages of pencil tracings of the formant center lines in spectrograms of VCV utterances containing the medial consonant /b/. Utterances in the same row have the same final vowel. Utterances in the same column have the same initial vowel. The time marks are 100 msec apart and the frequency marks 500 cps apart in the scales of the individual displays.

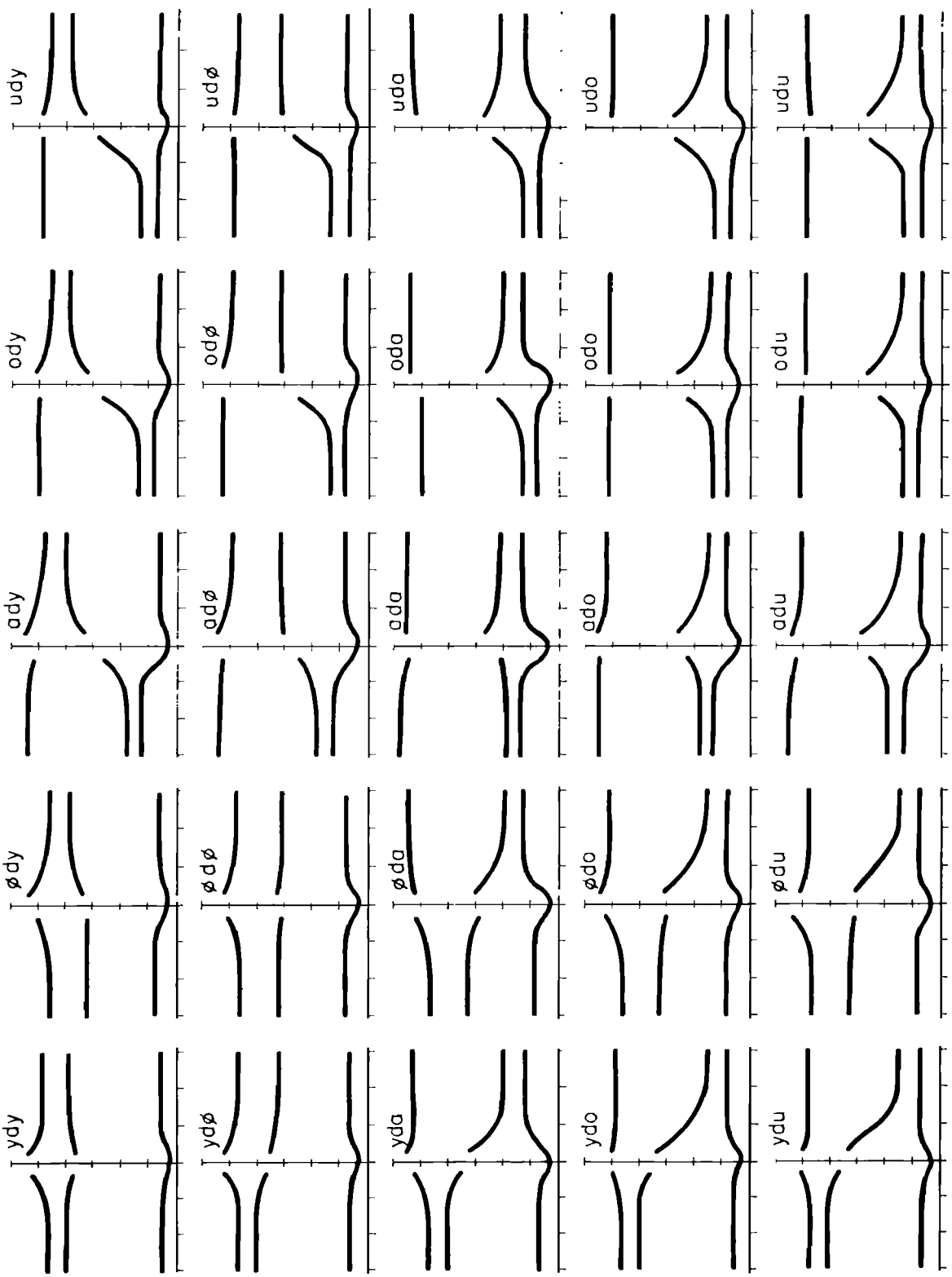

FIG. 8. Same as Fig. 7, except that the medial consonant is /d/.

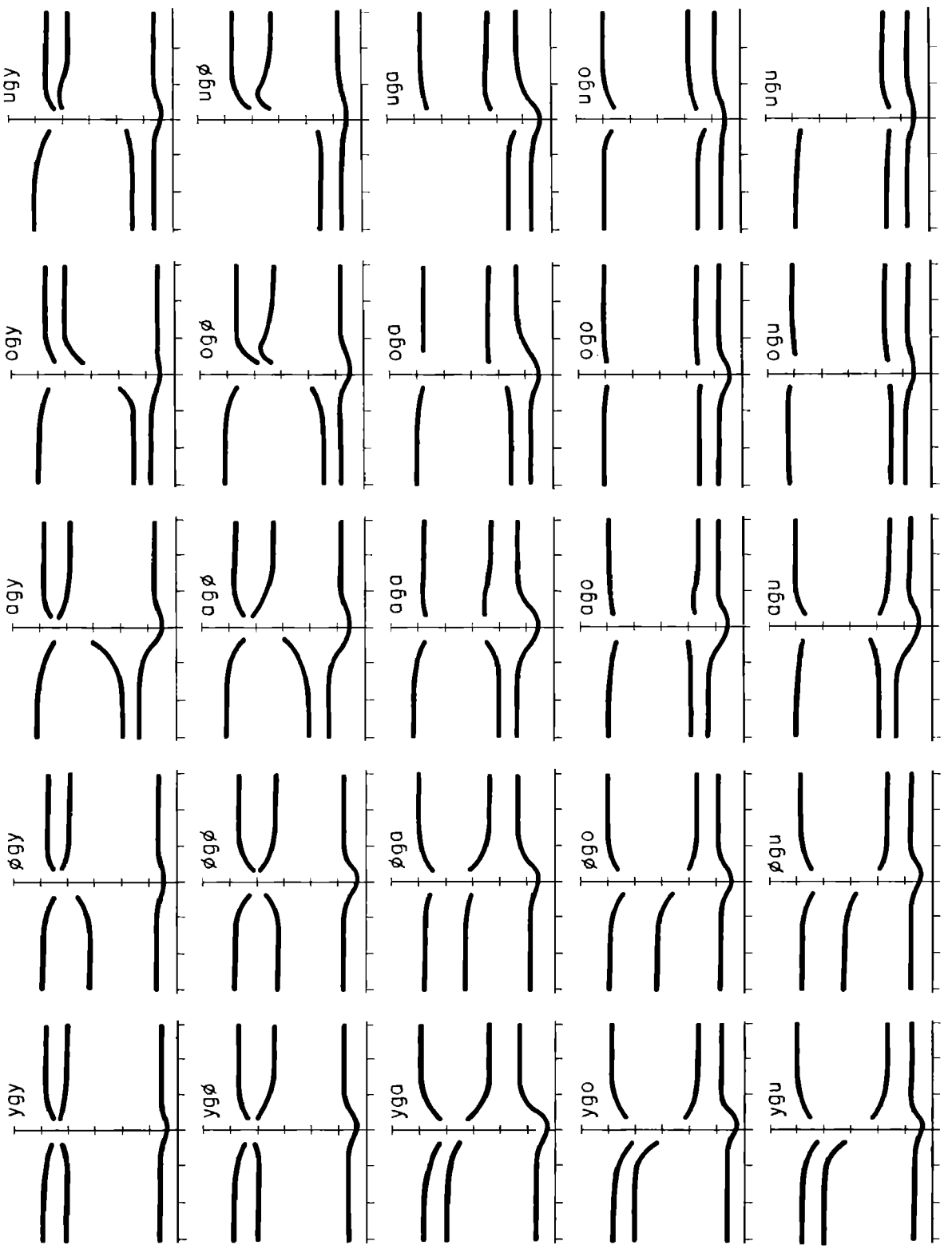

FIG. 9. Same as Fig. 7, except that the medial consonant is /g/.

forward so that at the moment of release the minimum cross-sectional area is located slightly behind the F_2 maximum of Fig. 10—i.e., at about $x=12$ cm. The F_2 terminal frequency, therefore, has a higher value at the release than at the moment of closure. The configurations of /y/ and /ø/ have their minimum cross-section points located anteriorly of the F_2 maximum. The point of constriction of the transitional configuration sequence following the relase of the /g/ must, therefore, pass through the region where F_2 has its maximum. In consequence of this, F_2 would first rise somewhat and then fall so that the typical S-shaped formant transition would result.

Figure 10 suggests that, on the above assumptions, the F_3 transition should also be S-shaped, but with a curvature opposite to that of the F_2 transition. This is not observed in the spectrograms, however. The reason for this may be that, while Fant's model postulates a catenoidal shape for the vocal-tract constriction, the actual configurations underlying the utterances of Fig. 9 were probably much more complicated. In general, the frequencies of the first and second formants can be estimated from relatively gross approximations of the vocal-tract shape. The dependence of a formant frequency on small irregularities in the shape of the vocal tract increases rapidly with the ordinal number of the formant, however.[21] X-ray studies of the utterances of Fig. 9 as spoken by a different talker suggest that the above-mentioned transitions from /g/ to /y/ or /ø/ involve a rapid decrease of the volume and length of the cavity anterior to the constriction. Model experiments by Fant[22] indicate that these changes should cause an upward shift of F_3 that could be large enough to cancel a possible downward deflection suggested by Fig. 10.

It is interesting that the transition from /g/ to a front vowel after an initial back vowel is not just a time-reversed version of the transition from a front vowel to /g/ followed by a back vowel. This is seen from a comparison of the displays in the lower left corner of Fig. 9 with those of the upper right corner. The former group does not show S-shaped F_2 transitions. Apparently, the point of closure in /ygu/, for example, is more anteriorly located than the point of release in /ugy/, and the F_2 maximum of Fig. 9 is passed during the closure.

E. Rôle of Lip-Rounding

The coarticulatory variations of the formant transitions that have been discussed above must be mainly the consequence of a corresponding variation of the shape of the tongue during the stop consonants. All the vowels of the word material were rounded and considerable changes in degree of lip-rounding would be expected to take place only in utterances containing both

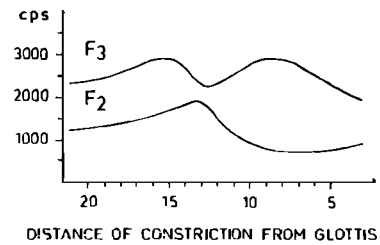

FIG. 10. Frequencies of second and third formants of a tube of length 18 cm and cross sectional area $A(x)=A_0\cosh^2(x-x_0)/h$ in the neighborhood of the point of minimum cross section x_0. The minimum cross-sectional area A_0 is assumed to be 65 mm². The laryngeal section has a fixed shape, h is a constant, and the lip-section area is also constant and equal to 65 mm². The abscissa represents x_0, the point of minimum cross-sectional area. [After C. G. M. Fant, *Acoustic Theory of Speech Production* (Mouton & Co., 's-Gravenhage 1960), p. 80].

a maximally open vowel; i.e., /ɑ/, and a maximally closed vowel such as /y/. If the difference in the course of the VC formant transitions of the two utterances /ɑgy/ and /ɑgɑ/, for instance, were due simply to a relatively more rapidly decreasing lip-opening area during the initial vowel of /ɑgy/, then the terminal-formant frequencies of /ɑg/ before /y/ would be expected to have lower values than those of /ɑg/ before /ɑ/ (Refs. 23, 24). Inspection of Fig. 1 shows that the actual situation is exactly opposite to this, however. A large number of similar cases can be found in the data. Naturally, the labial gesture will have some influence, but the lingual factor dominates. The same remark holds for the American-English material, to which we turn next.

III. AMERICAN-ENGLISH TALKER

In this study, a speaker of standard American English read a list of VCV words consisting of all combinations of the vowels /i/, /a/, /u/ and the consonants /d/ and /g/.

The average second-formant frequencies of the stationary parts of the vowels are presented in Table VII and the second-formant transition data appear in

TABLE VII. Frequencies in cps of the second formant measured in the vowels of VCV utterances spoken by an American subject. Each value is based on five measurements.

		VdV		VgV	
		Initial	Final	Initial	Final
	i	2260	2340	2290	2325
i	a	2340	2280	2240	2280
	u	2290	2260	2280	2280
	i	1180	1210	1200	1245
a	a	1140	1200	1160	1175
		1130	1180	1120	1205
	i	830	960	870	960
u	a	830	950	860	930
	u	860	950	860	960

[21] J. M. Heinz and K. N. Stevens, "On the Derivation of Area Functions and Acoustic Spectra from Cinéradiographic Films of Speech," J. Acoust. Soc. Am. 36, 1037–1038(A) (1964).
[22] Ref. 16, p. 66.

[23] V. Fromkin, "Lip Positions in American English Vowels," Language & Speech 7, 215–225 (1964).
[24] Ref. 16, p. 219.

TABLE VIII. Values in cps of $F_{2t}-F_{2v}$ for the American subject. The Table is arranged in the same way as Table I. Each number is an average of five measurements.

		VdV		VgV	
		VC	CV	VC	CV
	i	−130	−380	+10	−65
i	a	−190	−390	+100	−80
	u	−250	−370	−40	−20
	i	+360	+550	+260	+825
a	a	+275	+110	+160	+355
	u	+290	+420	+80	+165
	i	+730	+820	+130	+640
u	a	+630	+650	+90	+280
	u	+500	+650	0	+75

Table VIII. These data are based on five readings of each word.

It is apparent from Table VIII that the same general coarticulation rule holds for the American talker as that which was observed in the Swedish material. The locus shift is most prominent in the CV parts of the pairs /ida/ versus /ada/ and /iga/ versus /uga/.

IV. RUSSIAN TALKER

The palatalized/velarized distinction is an important feature of the Russian consonant system.[25] One of its acoustic manifestations is a fixation of the terminal frequencies of the formant transitions. This is illustrated by the following data.

A native speaker of Russian[26] read a list of VCV words each of which consisted of the vowel /ɛ/ followed by /b/, /d/, or /g/ or their palatalized counterparts and ending in one of the vowels /i, bI/, /ɛ/, /a/, /o/, or /u/. The words were read in random order and each word was read five times. Table IX shows the ranges of variation of the second-formant transitions as measured in the initial vowel under the different contextual conditions.

TABLE IX. Ranges of variation in cps of the second-formant transitions in the vowel /ɛ/, followed by various CV combinations as spoken by a Russian subject. Primes on the consonants represent palatalization. The dash corresponds to the five vowels /i, bI/, /ɛ/, /a/, /o/, and /u/.

ɛb′−	ɛb−	ɛd′−	ɛd−	ɛg′−	ɛg−
120	90	80	60	45	570

The variation is too small to be significant, except in the single case of the unpalatalized /g/. However, the sequence unpalatalized /g/+ front vowel is not admissible inside Russian morphemes. The measurements for /g/ may, therefore, not be representative. If this case is excluded, it appears that the type of coarticulation observed in Swedish and English does not obtain in the Russian of this talker.

An interesting acoustic feature of palatalization was observed when the formant transitions following the palatalized stops were compared with those following the corresponding unpalatalized stops in the same vowel context. The former transitions were usually convex upwards and the latter were convex downwards. With reference to Fig. 10, this is what would be expected in general if the release of the palatalized stops were associated with a forward motion of the point of maximum constriction of the tract and if the unpalatalized variants involved a backward motion.

V. FURTHER OBSERVATIONS

A large number of less systematic observations have been made on various coarticulation effects in the speech of several Swedish and American-English talkers. The results of these studies may be summarized simply by saying that they confirm the results reported in the present paper insofar as stop consonants are concerned. Sample measurements on American-English and Swedish fricatives in VCV context have not exhibited demonstrable locus shifts of the kind seen in connection with the stops, however. It would seem that the coarticulation properties of the former class of sounds are rather different from those of stops consonants.

VI. GENERAL DISCUSSION

A. Locus Theory

Before turning to the physiological interpretation of the data, it may be appropriate at this point to make a few comments about the locus theory of Delattre, Liberman, Cooper,[27] *et al.*, since the results reviewed above seem to differ from those obtained by these workers. In a now classical paper, Liberman ten years ago stated[28] (with several reservations): ". . . there are, for each consonant, characteristic frequency positions, or loci, at which the formant transitions begin, or to which they may be assumed to point. On this basis, the transitions may be regarded simply as movements of the formants from their respective loci to the frequency levels appropriate for the next phone, wherever those levels might be. The spectrographic patterns . . . , which produce /d/ before /i/, /a/, and /o/, show how this assumption suggests itself for the case of the second-formant transitions. We observe that all of these transitions seem to be pointing to a locus in the vicinity of 1800 cps." The Haskins workers were able to extrapolate fixed-locus frequencies also for American English /b/ and /g/. In the case of /g/, two loci had to

[25] M. Halle, *Sound Pattern of Russian* (Mouton & Co., 's-Gravenhage, 1959).

[26] Male native of Leningrad.

[27] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," J. Acoust. Soc. Am. 27, 769–773 (1955).

[28] A. M. Liberman, "Some Results of Research on Speech Perception," J. Acoust. Soc. Am. 29, 117–123 (1957), quotation from p. 118.

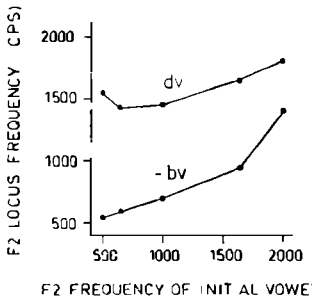

FIG. 11. Frequencies of second-formant transition loci of /b/ + vowel and /d/ + vowel sequences for different transconsonantal vowel conditions.

be assumed: a low-frequency locus for contexts involving back vowels and a high-frequency locus for front vowels.

In Fig. 11, a plot is presented that summarizes the CV second-formant locus frequencies of the data contained in Figs. 7 and 8. The /b/+ vowel and /d/+ vowel second-formant transitions were traced on transparent paper for each initial vowel of the VCV words separately. The locus frequency was then estimated in each of the five cases following the rules of Ref. 27. I.e., a common CV second-formant terminal frequency was extrapolated from the five utterances /yCy/, /yCø/, /yCɑ/, /yCo/, and /yCu/. A second one was obtained from /øCy/, /øCø/, /øCɑ/, /øCo/, and /øCy/, and so on for each initial vowel (C=b or d). In Fig. 11, these CV loci are plotted along the vertical axis as a function of the mean stationary F2 frequency of the initial vowel of the VCV words (horizontal axis).

It is apparent from this graph that the CV /b/ and /d/ loci are *not* independent of the vowel that precedes the CV sequence. Thus, the second formant in a transition from /b/ to a following vowel must originate from a point at 500 cps when the vowel /u/ precedes the stop. When /y/ precedes, the same point must be located at 1300 cps. An analogous statement holds for /d/. However, although the loci are not constant, the estimation of a locus frequency from VCV utterances having the same initial vowel (and consonant) was easy to perform in all cases presented in Fig. 11. In the case of /g/, more than one locus must be assumed even when the initial vowel is fixed (see for instance the second column from the left in Fig. 9).

The locus values presented by Delattre, Liberman, and Cooper were based on spectrographic studies of American-English CV utterances and were tested by means of synthesis experiments. These locus values are apparently close to those that are obtained if the consonants are studied in symmetric VCV contexts. The resulting loci will then be intermediate between the largest and smallest values given in Fig. 11. Thus, the single F_2 locus frequency obtainable for each of the consonants /b/ and /d/ by the methods of Ref. 27 will probably be an average of the several values that arise when all VCV contexts are considered.

It is possible, on the basis of the data presented in Figs. 7–9, to synthesize two VCV utterances that have the same VC and CV first-, second-, and third-formant *terminal* frequencies but that nevertheless differ with respect to the auditory impression of the consonant that they give to a Swedish listener. At least one of the vowels must be chosen differently in the two cases. The consonant of one of the stimuli may be perceived as a /b/ and that of the other stimulus as a /g/. Evidently, in such cases, the terminal frequencies themselves are less important for the perception of the stops than the manner in which these frequencies are approached and left by the formants during the transitional segments. It would thus seem that the perception of the intervocalic stop must be based on an auditory analysis of the entire VCV pattern rather than on any constant formant-frequency cue.

B. Physiological Interpretation

The data that have been presented in this paper suggest that in Swedish and English the variability of the formant transitions in VC sequences is controlled by the postconsonantal vowel (Fig. 1). 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 as 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.[29]

To understand something of the physiology behind these facts, we must consider the articulatory activity of the tongue. In Swedish and English, the invariant features of /d/ and /g/ are, respectively, the apico-alveolar[2] and dorsopalatal contacts. The precise shape of the vocal tract during the stop closure is quite irrelevant, phonemically, in these languages insofar as the identities of the stops are concerned. This shape may, in fact, vary a great deal without affecting the abovementioned characteristic closure features. Indeed the progressive assimilation of the stop to the final vowel manifests itself in adjustments of the very shape of the unconstricted parts of the vocal tract. This is demonstrated by the x-ray tracings of Fig. 12 (Ref. 30).

The second column from the left shows, from top to bottom, tracings of the vowels /y/, /ɑ/, and /u/. The column to the left of it shows /d/ closures in the environments /y–y/, /u–ɑ/, and /u–u/, and the column to the right shows /g/ closures in the same environments. It is

[29] The variability of the formant transitions in fixed-CV sequences apparently arises from the circumstance that the final vowel is approached in different ways depending on which preconsonantal vowel configuration the motion starts from.

[30] The tracings of Fig. 12 were obtained from x ray motion pictures of the author's speech. The cooperation of Prof. Negelius, Wenner-Gren Res. Lab., Stockholm, is acknowledged.

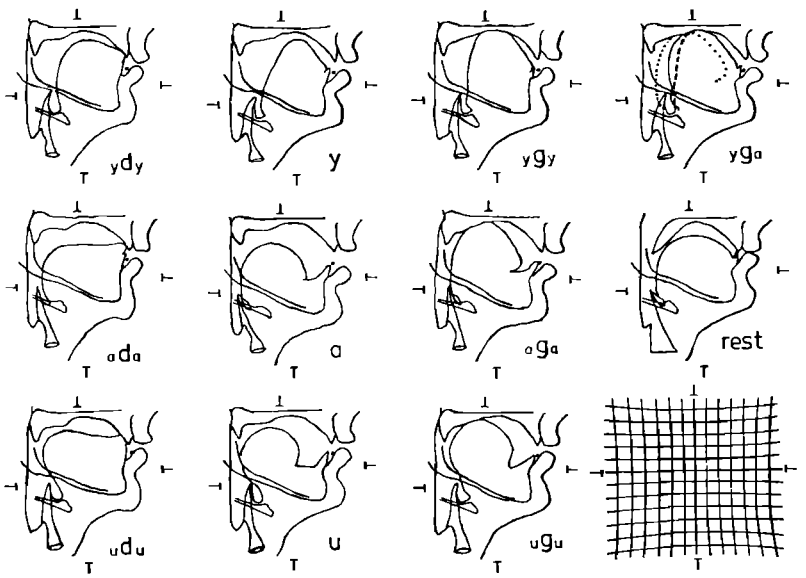

FIG. 12. Contour tracings from x-ray motion pictures of VCV utterances. The edges of the hyoid bone, the mandible, and the epiglottis are indicated. The grid in the lower right corner shows the spatial distortion introduced by the x-ray system. The real distances between adjacent lines in the grid are 1 cm.

apparent from these tracings that the unconstricted parts of the vocal tract assimilate the shape of the initial and final vowels. Note also, the vowel-dependent variability of the place of constriction in the /g/ column.

The rightmost part of the Figure (top) shows a tracing of the /g/ closure in /ygu/ sampled a few milliseconds after the beginning of the closure (solid lines). The superimposed dotted and dashed tongue contours represent the /g/ closure of /aga/ and /ygy/, respectively. It is seen here that the vocal-tract shape delimited by the solid line is intermediate between that of the dotted and dashed lines. This illustrates the fact that the cavities may change their shape even during the stop closure. In other words, the tongue is able to make a distorted vowel gesture, *while* it is executing the stop consonant.

We may summarize these observations in terms of the following more-hypothesized statement. The production of the vowels and of the apical and dorsal consonants involve activity in three (probably partly overlapping) sets of muscles. These sets have separate neural representations in the motor networks of the speaker's brain. Articulatory commands may be transmitted over the three neural control channels independently of each other. However, the dynamic response of the tongue to a compound instruction is a complex summation (neural, muscular, and probably mechanical also) of the responses to each of the components of the instruction.

In other words, we may probably regard the tongue as three separate articulatory systems that share some muscles. The three systems may be controlled in a purely independent manner. Different languages may impose phonetic restrictions overruling this freedom of control, however. Thus, in Swedish and English, the stop consonants seem to coarticulate relatively freely with the vowels. This implies, in terms of the present model, that these stops are produced almost exclusively by commands over the apical and dorsal channels, leaving certain subsets of the muscles responding to vowel instructions free to anticipate a following vowel. On the other hand, there are languages, such as Russian, in which the instructions for the stop consonants are made up of an apical or dorsal command as in English or Swedish but with the additional feature that the vowel channel must simultaneously receive exactly one of two fixed commands [i] or [bI], where [i] produces palatalization and [bI] velarization.[31]

To say that the apical and dorsal constriction systems of the tongue may be controlled independently of the vowel activity is, thus, analogous to the statement that nasalization and voicing are independent parameters in speech. This is a universal feature of the basic phonetic competence of all normal speakers. The fact that some languages do not make a distinction between voiced and voiceless nasals, whereas others do,[32] must be interpreted as a difference in the linguistic programming of the available physiological apparatus.

The model discussed thus far may perhaps also shed some light on the dynamic behavior of the lips. Accordingly, we should probably distinguish two physiologically independent types of labial activity in speech: namely (Type 1), the closing motions that take place in the vertical dimension and width are observable, e.g., in the English consonants /p b m f v/; and (Type 2) the rounding-spreading dimension of motion, which is used for the phonetic feature of rounding in vowels.

There are rounded as well as unrounded labial consonants in English (compare the /p/ of "Oh Pooh!"

[31] S. E. G. Öhman, "Note on Palatalization in Russian," Mass. Inst. Technol. Res. Lab. Electron. Quart. Progr. Rept. No. 73, 161–171 (15 April 1964).

[32] R. Jakobson, G. Fant, and M. Halle, *Preliminaries to Speech Analysis* (MIT Press, Cambridge, Mass., 1963), 4th printing, p. 40.

with that of "Yippee!"). This suggests that the vowel component of the total labial system, however nonphonemic in English, in principle could be used to add phonemic distinctions to *labial* consonants that are otherwise produced by the consonantal (Type 1) activity mentioned above.[33] The behavior of the lips would thus seem to be analogous, in this respect, to that of the tongue.[34]

C. Timing of the Instruction Components

The model suggested here postulates only three motor instructions corresponding to a Swedish or English vowel + stop + vowel utterance: namely, one for the initial vowel, one for the consonant, and one for the final vowel, the latter being applied before the response to the consonant instruction has died out so that the observed coarticulation effect may result.[35]

However, the data presented in this paper support only the assertion that *some* components of the vowel gesture are active already at the onset of the consonant gesture in certain VCV words, but they are inconclusive as to the question whether or not further vowel instruction components are added on during or at the offset of the consonant gesture. To judge from x-ray motion pictures of /ydu/ for instance, the /d/ could not be properly articulated if the back part of the tongue assumed the /u/ position too early for, then, the tip of the tongue would apparently not be able to reach the alveolar ridge. This is also observable in /udu/, where the back part and indeed the whole mass of the tongue is seen to move forward during the /d/ gesture, seemingly for the purpose of making the alveolar contact possible. Thus, in /ydu/, the motion in the direction of the final vowel during the /d/ gesture cannot go too far but has to await the offset of the consonant for its completion. The consonant gesture somehow overrules the vowel gesture if the latter is antagonistic to the former. It is therefore not clear whether the vowel gesture may be said to be the result of *one* temporally invariant instruction or whether the relative timing of the components of the instruction is variable from context to context. Further physiological measurements are needed to clarify this problem.

The complexity of the relative timing relations is also illustrated by the data of Fig. 3, which show that the formant frequencies during the stationary parts of the *initial* vowels of the Swedish word material were different for different medial stop consonants. Figure 13 shows two superposed x-ray tracings of the tongue shape

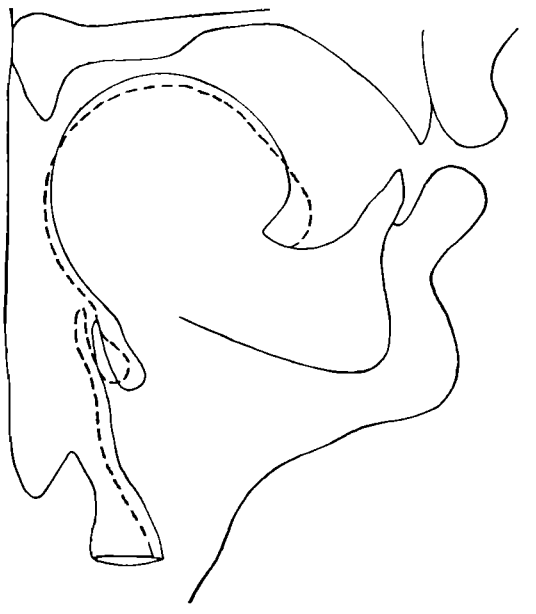

FIG. 13. Contour tracings from x-ray motion pictures of the vowel /u/ before /g/ (solid line) and before /d/ (dashed line) in VCV utterances.

during /u/ as recorded several hundred milliseconds before the closures of /g/ and /d/, respectively. It is seen that /u/ before /g/ has a narrower dorsopalatal region than /u/ before /d/. This physiological difference correlates well with the acoustic difference seen in Fig. 3.

Apparently, the articulatory system "prepares" for the medial consonant during all of the initial vowel. The effect resembles the nasalization of vowels before nasal consonants, which in many languages may effect the entire vowel, from the moment it starts, even when the vowel is prolonged.

Evidently, the fact that a consonant may color a prolonged preceding vowel speaks against the notion that the neural instructions that control the stream of articulatory gestures of an utterance may be regarded as a simple sequence of temporally well-defined impulses, each of which corresponds to a single unit of a linguistic transcription of that utterance. A more-appropriate model would perhaps be one that operates with a hierarchy of instruction levels corresponding, e.g., to the linguistic units of phoneme, syllable, word, and phrase. The coarticulation phenomena studied in the present paper could then be termed *phonemic*, since they concern interactions between phoneme gestures. The preparatory effect of consonants upon the vowel of a preceding syllable mentioned in the last paragraph might be the consequence of some higher-level coarticulation rule.

VII. SUMMARY

For the speakers of Swedish and English who were discussed in the present paper, the terminal frequencies

[33] According to the authors of Ref. 32, p. 50, this is the case in Circassian.

[34] The retroflex vowels of English exemplify the use of consonantal (apical) gestures for the addition of an extra vowel distinction.

[35] A model of this kind was presented by the author in "Numerical Model for Coarticulation, Using a Computer-Simulated Vocal Tract," J. Acoust. Soc. Am. 36, 1038(A) (1964). A more complete summary is in preparation.

of the formants in VCV utterances depend not only on the consonants but on the entire vowel context. The stop-consonant loci are, therefore, not unique. The reason for this fact seems to be that the production of the consonant involves concomitant articulatory adjustments partially anticipating the configuration of the succeeding vowel. Neural commands for the consonant and the following vowel must, hence, be active simultaneously. However, certain components of the final vowel instruction are probably inhibited during the consonant. Furthermore, the medial consonant configuration may be slightly anticipated during the initial vowel. The latter effect is observable even at the beginning of prolonged initial vowels.

The freedom of coarticulation of the Russian stops is small. Apparently, these stops must always be coarticulated with one of two fixed vowels, the palatal [i] or the velar [bI]. Similarly, Swedish and American-English fricatives do not seem to enjoy the same coarticulation properties as the stops.

Probably, these observations may be accounted for in terms of an articulatory speech synthesis model in which (1) apical constrictions, (2) dorsal constrictions, and (3) the position of the body of the tongue are represented by a small number of independent control parameters. The operation of this model would involve a hierarchy of instructions corresponding to the phoneme, syllable, word, and phrase levels of a linguistic representation.

ACKNOWLEDGMENTS

I am deeply indebted to Dr. C. G. M. Fant, Royal Institute of Technology (KTH), Prof. K. N. Stevens, Massachusetts Institute of Technology, and B. E. F. Lindblom, Royal Institute of Technology (KTH), for the invaluable support, encouragement, and intellectual stimulus that I have received from them during my work with the present paper. Stimulating discussions with Professor M. Halle, Massachusetts Institute of Technology, Professor P. Ladefoged, University of California, Los Angeles, and Professor J. M. Heinz, Massachusetts Institute of Technology, have given me useful instruction of a kind that I could not have obtained otherwise. I also want to express my sincere thanks to all members of the Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden, the Speech Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, and the Haskins Laboratories, New York, among whom I have had the opportunity to work. I am particularly grateful to S. Felicetti for her expert aid in the preparation of the manuscript.

This work has been supported by the U. S. Air Force, by the U. S. Army, by the National Institutes of Health, U. S. Department of Health, Education, and Welfare, by the Swedish Technical Research Council (Statens Tekniska Forskingsråd), and by the Sweden–America Foundation (Sverige–Amerika Stiftelsen).