ASSIMILATION AS A CONTINUOUS COARTICULATORY PROCESS: FIRST ARTICULOGRAPHIC RESULTS

B. Pompino-Marschall *

Institut für Phonetik und Sprachliche Kommunikation der Universität München, Germany.

ABSTRACT

Alveolar to velar assimilation in stop place of articulation was studied in German utterances by spectrographic analysis and by recording tongue movements with an electromagnetic articulograph. Spectral and positional differences between the stop consonants and between different speaking rates were analyzed. Increased speaking rate for the alveolar stops clearly resulted in positional changes of the tongue towards a position appropriate for velar stops but still significantly different from the latter. This seems to be in accordance with the view that assimilation is a continuous coarticulatory process.

1. INTRODUCTION

In 1988 Nolan [2] reported an EPG study on English alveolar to velar assimilation of stop place of articulation in utterances of the form "... bed girls ..." (vs all velar "... beg girls ..."). Assimilation resulted in EPG patterns with reduced as well as with a total loss of alveolar contact. This finding seemed to be in accordance with the view that assimilation is rather a continuous coarticulatory process than a process of featural change (cf. also [1]).

Since despite the lack of alveolar contact even those 'totally' assimilated items could be discriminated from the all velar utterances better than chance with the following pilot experiments we wanted to study the differences in tongue *movement* in alveolar to velar assimilation.

2. PROCEDURE

Two male German speakers read ten utterances of the type "Wir wollen zu Bett gehen" (vs all velar "Wir wollen zu Beck gehen") ten times at two different speaking rates (normal/fast) in randomized order. Besides the audio signal tongue movements were recorded with the help of an electromagnetic articulograph (AG100, Carstens Medizinelektronik; cf. [3]) via three coils placed on the midsaggital line of the tongue: (1) as far back as possible (back coil), (2) ca. 0.5 cm behind the the tip of the tongue (front coil), and (3) midway between the others (mid coil).

Besides spectrographic analysis of formant frequencies in the middle of the preceeding vowel and at implosion of the alveolar/velar stop, and of acoustic segment duration, the position of the three coils (as their x/y-coordinates on the midsaggital plane) were determined at the following points in time: (1) in the middle of the preceeding vowel, (2) at the beginning of stop closure, (3) at the first stop release (if present[1]), and (4) at the second stop release.

3. RESULTS

In contrast to the English study the auditory analysis of our data revealed that there is only a weak tendency for assimilation in this German material constructed in parallel. One of the speakers almost never produced perceivable assimilations. For the numerical analysis we therefore chose the item with the most occurences of assimilation, i.e. "Er wird es bald kriegen" (vs all-velar "Er wird das Balg kriegen").The mean coil position at the four points in time for both subjects and both places of articulation are seen in Figure 1. The differences in coil placement for the two subjects are clearly seen besides the differences in stop place of articulation.

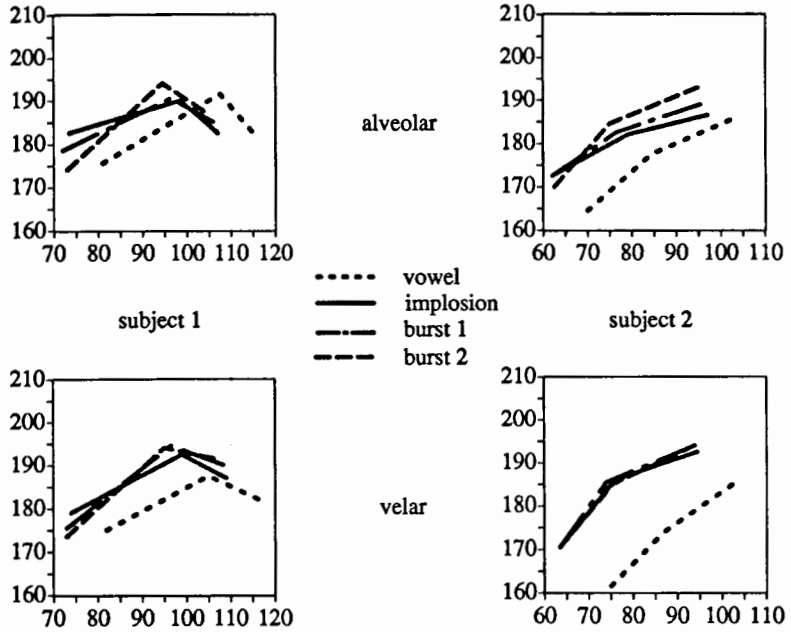


Figure 1: Mean 'tongue contours' at four different points in time for alveolar/velar stop production for both subjects (x/y coil positions in mm).

For the statistical analysis the positional data as well as the measured frequencies for the second and third formants at implosion were subjected to separate two-factorial analyses of variance (with the factors speech rate and place of articulation). There was no influence of speech rate on the formant frequencies at implosion for either speaker. The only significant effects were a higher F2 ($p < .001$) and a lower F3 ($p < .001$) for velars. The positional data showed significant effects only for one speaker (S1). His mean coil positions and standard deviations are shown in inTable I and II. The differences in tongue contour are also shown in Figure 2.

The analyses of variance showed significant interactions between the factors speech rate and place of articulation only for the y-position of the mid and back coil at the first stop release: while for alveolar/velar utterances at fast rate of speech coil 2 is on the average 4.7 mm higher ($p < .001$) than in the case of normal rate of speech, no such rate effects are seen for the all velar utterances. A parallel effect is seen for the y-position of the back coil at the first stop release: the value for the fast alveolar utterance is on the average 3.9 mm higher than for the utterance at normal rate. At the same time the differences between fast alveolar/velar and all velar utterances always remain significant ($p < .01; .001$).[2]

4. DISCUSSION

These results (cf. Figure 2) can be summarized as follows: Whereas tongue position

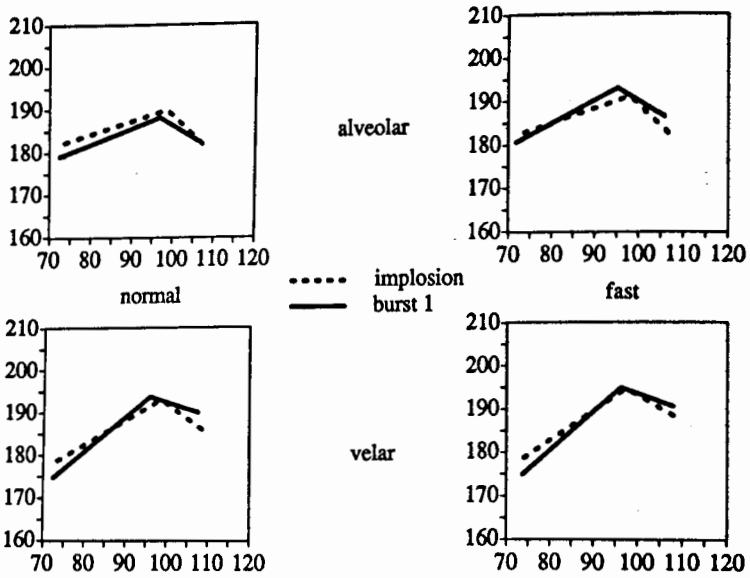


Figure 2: Mean 'tongue contours' at implosion and first burst for alveolar/velar stop production at normal vs fast rate of speech for subject 1 (x/y coil positions in mm).

Table I:
Mean coil postion at implosion (a) and at stop release (b) in cm
(1st line: x, 3rd line: y, 2nd/4th line: standard deviations) at slow rate for S1
(a: alveolar, v: velar)

(a)

coil

1		2		3	
a	v	a	v	a	v
7.36	7.34	9.84	9.87	10.74	10.90
.13	.13	.10	.15	.08	.15
18.23	17.87	18.99	19.30	18.17	18.57
.21	.26	.15	.10	.10	.16

(b)

coil

1		2		3	
a	v	a	v	a	v
7.24	7.24	9.67	9.60	10.61	10.76
.13	.10	.08	.15	.07	.11
17.91	17.47	18.81	19.37	18.26	19.00
.16	.22	.13	.10	.13	.12

Table II:
Mean coil postion at implosion (a) and at stop release (b) in cm
(1st line: x, 3rd line: y, 2nd/4th line: standard deviations) at fast rate for S1
(a: alveolar, v: velar)

(a)

coil

1		2		3	
a	v	a	v	a	v
7.36	7.40	9.81	9.76	10.71	10.84
.05	.16	.11	.13	.12	.11
18.29	17.87	19.10	19.46	18.23	18.83
.13	.24	.12	.10	.11	.11

(b)

coil

1		2		3	
a	v	a	v	a	v
7.17	7.37	9.53	9.64	10.59	10.81
.08	.16	.05	.13	.07	.16
18.06	17.49	19.29	19.49	18.64	19.06
.19	.26	.12	.07	.05	.11

at implosion and the first stop release does not change dramatically with speaking ratefor the all velar utterances, the back of the tongue clearly adopts a higher position in the fast alveolar/velar utterances. But on the other hand this higher position does not reach the configuration of the all velar utterances. This clearly seems to be in accordance with the view that assimilation rather is due to a continuous coarticulatory process than a process of featural change.

5. FURTHER EXPERIMENTS

In a second pilot experiment these effects were studied in more detail with another male German subject. Here, additionally, we wanted to study the influence of context on alveolar/velar assimilation: Besides preceding /_ald/g/ as in the experiments above ("bald/Balg") simple /_ad/g/–endings ("Tat/Tag") were used with following accented vs unaccented /ge/ ("geben" vs "gestanden" or "gehalten"; accented syllables bold).

6. NOTES

* The experiments reported here were conducted in the course of an experimental workshop. I am indepted to my students S. Burger, P. Janker, L. Kuffer, C. Mooshammer, D.Stein and A. Zimmer for help in carrying out the experiments and part of the analyses.

1 This was almost always the case. For the statistical analyses only items showing two stop releases were used.

2 Besides these interactions there is a further for the x–position of the mid coil at the first stop release, only showing a marginal ($p < .05$) fronting effect (.5 mm) for fast alveolars (all other simple effects being not significant). Another interaction – not of relevance here – is seen for the y–position of the back coil at implosion: Here the simple effect of speaking rate for alveolars is not significant, the one for velars showing on average a 2.5 mm higher position for the fast rate ($p < .001$).

The other main effects – not of relevance here – are: mid coil, y–position at implosion, 1.4 mm higher at fast rate ($p < .01$) and 3.3 mm higher vor velars ($p < .001$); front coil, y–position at the first stop release, 5.1 mm higher for alveolars ($p < .001$); back coil, x–position at the first burst, 1.9 mm more back for velars ($p < .001$).

6. REFERENCES

[1] ENGSTRAND, O. & KRULL, D. (1988), "Discontinuous variation in speech", *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm PERILUS, VIII*, 48–53.

[2] NOLAN, F. (1988) "The descriptive role of segments: Evidence from assimilation" *2nd Conference on Laboratory Phonology*. Edinburgh.

[3] SCHÖNLE, P. MÜLLER, C. & WENIG, P. (1989) "Echtzeitanalyse von orofacialen Bewegungen mit Hilfe der elektromagnetischen Artikulographie" *Biomedizinische Technik, 34*, 126–130.

AUTOMATIC PROCEDURE FOR LARYNGOGRAPHIC (Lx) ANALYSIS OF PHONATION CONTRASTS

J. H. Esling**/*, B. C. Dickson*/** and J. R. Woolsey*

**University of Victoria, Canada
*Speech Technology Research, Ltd., Victoria, Canada

ABSTRACT

This report describes an automated, microcomputer-based procedure for comparing laryngographic (Lx) waveforms. The program enables researchers to analyze the cycle-by-cycle changes in vocal fold vibratory characteristics that may signal linguistic contrasts or differences in long-term voice quality. The first differential of the original Lx signal, captured digitally, is marked to indicate beginning, ending, upper and lower limits. A set of ratios is then obtained relating increasing voltage to decreasing voltage for each period of the signal. Considerations such as the inherent variability of the Lx signal, techniques of Lx recording, and applications of the algorithm are discussed.

1. LARYNGOGRAPHIC ANALYSIS

Electrical impedance laryngography has been used in phonetic research to quantify differences between contrasting types of phonation [4]. Such contrasts appear linguistically in languages like Korean at the syllable level in conjunction with phonologically distinct manners of consonantal articulation [1] [6] [7] [8] [9]. Phonatory contrasts also appear as long-term postures in voice quality with largely indexical significance [3, 10]. One problem in the analysis of the larynx waveform (Lx) has been the highly variable data that it yields. The signal is obtained by means of superficial throat electrodes which measure decreasing impedance as the vocal fold mass comes together, and increasing impedance as these structures separate [5].

Different models of laryngograph and differing recording procedures result in Lx signals with varying phase characteristics. This makes it difficult to analyze characteristics of individual waveform periods to distinguish, for example, a breathy voice from a harsh voice. Another problem is the DC float that characterizes many Lx signals and which makes establishing a baseline for reliable measurement of individual period characteristics particularly difficult. Aspects of obtaining an initial, workable Lx signal are dealt with in section 2. A solution to the baseline problem is presented in section 3. The method of segmenting the waveform to obtain a ratio is presented in section 4.

2. RECORDING Lx

Recordings of Lx signals made on a standard AM tape recorder tend to be distorted by phase shift. For this reason, it is valuable to develop procedures for direct digitization of the Lx signal, using an adequate (16-bit) data acquisition system. However, if this is not possible, the signal may be recorded using a system that does not introduce phase distortion, such as a Sony PCM digital audio processor and recorder, as has been used for these experiments.

Attempts at controlling DC float in the Lx signal include assuring that proper coupling with the input preamplifier or similar analog conditioner is maintained. However, low-frequency oscillations can be expected as a result of laryngeal movement around the axis of the electrodes. This is more apparent during continuous speech than during examples taken from sustained vowels, owing to the natural raising and lowering of the larynx in the less controlled situations.

The polarity of the Lx signal must also be

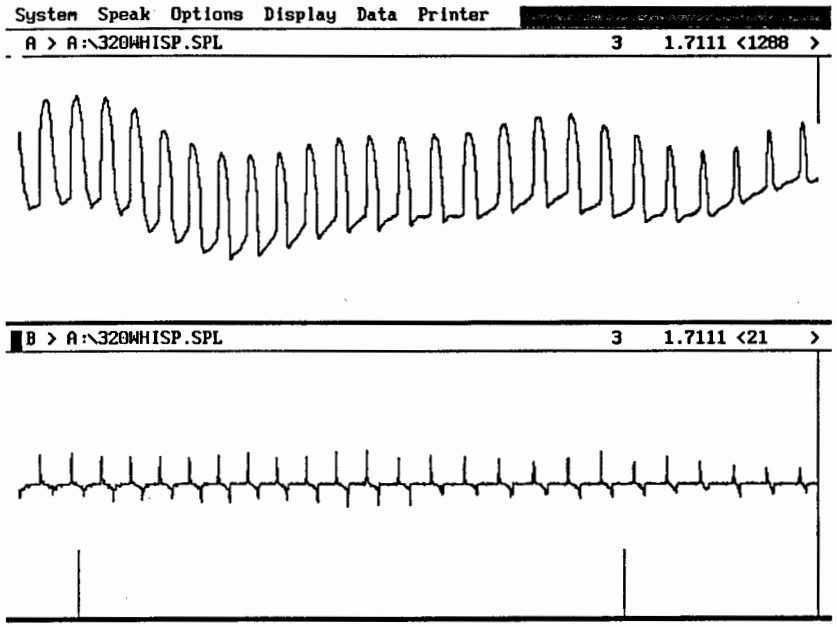


FIGURE 1. Original Lx waveform (top) and difference waveform (bottom) of a sample of whispery voice. Data marked (cursors), ready for ratio analysis.

considered. When the signal is taken directly from the laryngograph, the high impedance component of the signal is converted to the maximum positive voltage in the waveform, while the low impedance component results in a negative voltage. As the low impedance component is a result of maximum current flow across the electrodes, this will occur when the glottis is in its maximally closed phase of the voicing cycle. Because it is more representative of laryngeal behaviour to display the closed phase on the positive side of the waveform, we prefer to invert the polarity of a signal that has been taken directly from the laryngograph. However, if the signal is passed through a preamplifier at any stage, this will result in the polarity being reversed.

3. DATA ACQUISITION

Data acquisition is carried out using the CSL digital signal processing system [2], operating on an IBM-AT workalike. Data acquisition is performed at a rate of between 10K and 40K samples/second and the resulting sampled data files are passed to the EDIT320 software package. In that package, the waveform is displayed graphically and manipulated to enhance the laryngeal characteristics of interest (see FIG. 1).

The first differential of the waveform emphasizes the change in voltage over time, thus providing a representation of the Lx signal that closely models significant changes in current (as the impedance changes from, e.g., high impedance during the open phase of the laryngeal cycle to low impedance during the closed phase) as in equation (1).

$$(1) \qquad d_i = y_i - y_{i-1}$$

where $i = \{1, 2, 3 \ldots n\}$ sampled data points

A side effect of taking the first differential is that low frequency oscillations attributed to larynx movement, as well as DC float, are eliminated.

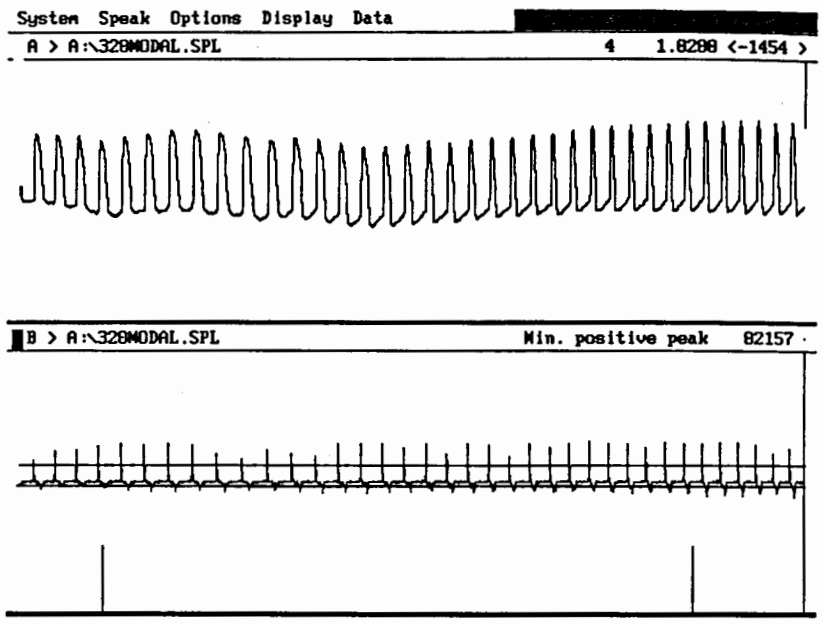


FIGURE 2. Lx waveforms of a sample of modal voice.
Cursors marking minimum positive value and minimum negative value.

4. PROCESSING ALGORITHM

To derive the ratios for each Lx period of a range of voiced speech, the original Lx signal is loaded using EDIT320 and flipped if necessary (the sign is changed on each amplitude value) depending on recording conditions, and the first differential is computed. A minimum positive threshold and a minimum negative threshold are then selected, using horizontal cursors as illustrated in FIG. 2, to eliminate the effect of arbitrary zero crossings. For each period in the marked range, the greatest negative excursion (i_1) that is less than the negative threshold and the greatest positive excursion of the period (i_2) that is greater than the positive threshold are identified. A third value (i_3) is defined as the subsequent greatest negative excursion below threshold, beginning the following period. A ratio is then calculated for each Lx period as shown in equation (2).

$$(2) \qquad (i_2 - i_1)/(i_3 - i_2)$$

where i = the selected sampled data point

For each succeeding period, the previous i_3 becomes the new i_1, until the end of the range is reached. The resulting ratios are stored in a file; as shown in FIG. 3, computed for a portion of the differenced signal for harsh voice.

5. APPLICATIONS

Applications of this analysis algorithm focus on the identification of phonatory differences at the segmental, CV, or long-term level. The hypothesis that a distinctive 'breathy' phonatory quality is associated prosodically with the lenis (vs. aspirated or fortis) consonant series in Korean as a principal cue in identifying meaning in CV sequences, for example, can now be tested. Lx rise-time to fall-time ratios of sets of controlled phonetic models can also be compared with specific language data or with examples of phonation in pathological speech.

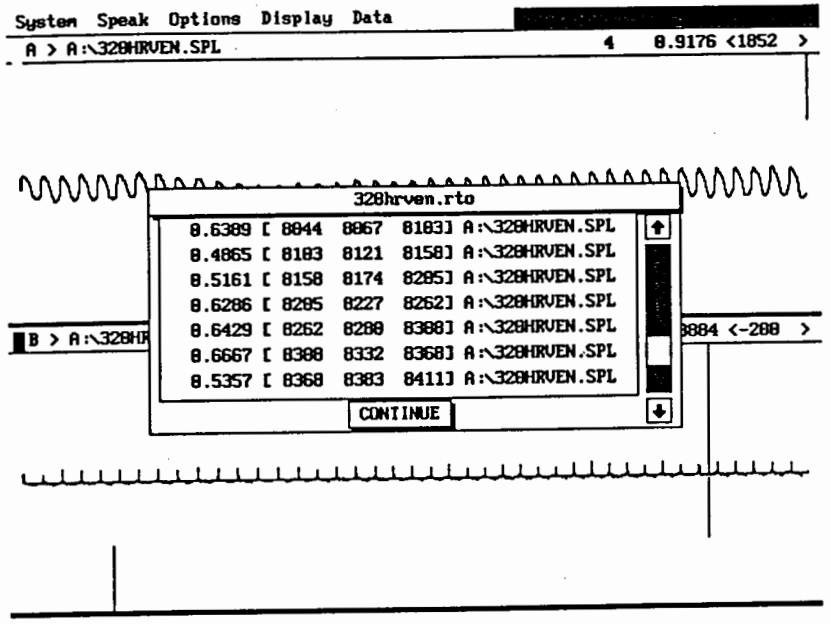


FIGURE 3. EDIT320 display showing part of the result file (ratio; start, peak and end point times for each period) for harsh (ventricular) voice.

Initial examination of phonation types using this procedure illustrates that harshness and creakiness, which have low Lx ratios, differ from modal voice, and from whisperiness and breathiness, which increase progressively in Lx ratio range, as predicted in prior research [3].

6. REFERENCES

[1] ABBERTON, E. (1972), "Some laryngographic data for Korean stops", *JIPA, 2,* 67-78.
[2] DICKSON, B. C., et al. (1990), *CSL User's manual"*, Pine Brook, NJ: Kay Elemetrics Corporation.
[3] ESLING, J. H. (1984), "Laryngographic study of phonation type and laryngeal configuration", *JIPA, 14,* 56-73.
[4] FOURCIN, A. J. (1974), "Laryngographic examination of vocal fold vibration", *Ventilatory and phonatory control systems* (B. Wyke, ed.), 315-333. London: Oxford University Press.
[5] FOURCIN. A. J., & ABBERTON, E. (1976), "The laryngograph and the voiscope in speech therapy", *XVIth Int. Congr. Logop. Phoniatr.,* 116-122. Basel: Karger.
[6] FUJIMURA, O. (1977), "Control of the larynx in speech", *Phonetica, 34,* 280-288.
[7] IVERSON, G. K. (1983), "Korean *s*", *Journal of Phonetics, 11,* 191-200.
[8] KIM, C-W. (1965), "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)", *Word, 21,* 339-359.
[9] LADEFOGED, P. (1973), "The features of the larynx", *Journal of Phonetics, 1,* 73-83.
[10] LAVER, J. (1980), *"The phonetic description of voice quality"*, Cambridge: Cambridge University Press.

ELECTROPALATOGRAPHY OF CONVERSATIONAL SPEECH

L. Shockey

University of Reading, U.K.

ABSTRACT

Electropalatography was used to sample natural conversational English. A tabulation was made of cases where alveolar obstruents could occur and of how these underlying consonants were realised. The results reflect large scale reduction of alveolars in conversational speech, some of which (e.g. reduced lateral contact) seem to be common to all members of the set and some of which are more particular to the class of speech sounds involved (laterals, nasals, stops, fricatives).

1. INTRODUCTION

Until recently, little or no research using electropalatography has focused on tongue-palate contact during relaxed, unselfconscious speech such as that which we use in everyday discourse. The reason for this is presumably the unease which besets phoneticians when they think about doing research on non-laboratory speech: in collecting free conversation, one cannot control for any of the variables known to influence articulation, among them segmental environment, stress, place in utterance, and word class. In addition, one never knows how many tokens of a given type will appear on any particular occasion, thereby making it hard to apply standard statistical measures to the results. Yet, surely if our goal as linguists is to model speech as it is used by ordinary people in daily life, it is vital to develop techniques for collecting and analysing data about this type of speech. Electropalatography provides an indirect but dynamic picture of articulator movement and as such is an invaluable adjunct to auditory and acoustic analysis of natural speech.

2. EXPERIMENTAL METHOD

In this study, acoustic and EPG data were collected from two subjects involved in conversation. The subjects were both longterm EPG users, having been on the team which developed the system currently in use at Reading University. They reported feeling very comfortable wearing the palate and experiencing no interference with articulation. Each of the subjects was seated comfortably in a small room and asked to talk to another member of the research team whose speech was not being monitored. The experimenter was in an adjoining room, listening to the conversation. After an initial period during which the conversants seemed to have become involved in discussion and to be producing unselfconscious output, the experimenter collected three-second samples of acoustic and EPG data. The acoustic signal was sampled at 10KHz and the EPG output at 100Hz. One minute of speech was collected from subject WJ, a West Midlands speaker with considerable Standard Southern overlay and 1.5 minutes collected from subject FG, a Standard Southern British speaker.

An impressionistic phonetic transcription of the collected corpus was done as well as a phonemic transcription. A tabulation was then made of cases of /t,d,s,z,n,l/ (the alveolar consonants involving contact in English), and each phonemic form related to both its phonetic transcription and the span of 10-millisecond EPG patterns which corresponded to it. The phonemic category provided a list of places where it would in theory be possible to find a maximally-articulated alveolar consonant; the phonetic realisations were divided into three categories: complete closure, incomplete closure, and deletion. These are very crude divisions. Complete closure was defined as the case in which every column of the palatogram indicated contact in at least one of the first four rows. Many kinds of complete closure were noted. For example, several degrees of lateral contact could be seen for everything except [l]: some showed a great deal of lateral contact, presumably indicating a high tongue position. Less side contact was visible in others, suggesting a laxer closure. The tokens with weak lateral contact were very common: this may prove to be a predictable feature of English conversational speech.

Complete closure per se cannot be said to apply to fricatives at all, since they require an incomplete closure in their production. For the same reason, the notion of incompleteness is not well-specified for fricatives: some with a very wide central channel were found, but as they were heard to produce friction, they could not be judged as incomplete.

Deletion in this case was defined as "showing no palatal contact": clearly inadequate, since a gesture of considerable proportions can be made without actually making contact with the palate.

While these categories will, therefore, have to be amended in a more detailed report, they allow us to shed some light on the behaviour of the elements investigated and so have been preserved here.

3. RESULTS

Not all underlying alveolars were fully realised, and in a pattern which was relatively similar from speaker to speaker. Table 1 shows summary data averaged over all consonants for each speaker and for both speakers combined.

TABLE 1

	WJ	%	FG	%	Both	%
all alveolars						
complete	113	71	182	69	295	69
incomplete	13	8	39	15	52	12
deleted	31	19	24	9	55	13
glottalled	3	2	20	8	23	5
total	160		265			

Characteristic realisation patterns emerged for each manner of articulation:

1. /n/ -- Reduction of /n/ can be attributed to two main factors, a) a Vn sequence is often reduced to a nasalised vowel before another alveolar consonant, and b) [n] often shows incomplete closure intervocalically.

In addition, [n] shows, in common with most of the other consonants investigated, a tendency to be articulated with a central groove before a fricative. It is a well-established tenet of phonetics that the production of the near-closure for a fricative involves finer motor control than the (theoretically) complete closures found for stops and nasals. Electropalatograms show that preparation for the groove configuration begins in preceding alveolar consonants and can sometimes be detected in vowels preceding such clusters.

2. /l/ -- In these subjects, there were two distinct realisations of /l/. One involved contact with the palate and was found syllable-initially, at the trailing end of a cluster, and intervocalically. The other involved no contact and was found at the leading end of a cluster and finally. The light or "semivocalised" closure which was noted by Hardcastle and Barry [1] in some environments was not found to be characteristic of these subjects: subject FG showed four anomalous cases, but these were a very small proportion of the total.

3. /s/ and /z/ -- These sounds tend to be preserved in some form, but (as mentioned above) often get a very wide channel in these data, implying (in agreement with the lateral contact discussed above) less raising of the tongue toward the palate than is found in citation forms.

4. /d/ -- A fully closed [d] is normally found after another non-nasal alveolar, especially word-finally when the next word begins with a vowel. The closure tends to weaken intervocalically, even if the [d] is word-final. (The resulting segment does not sound like a fricative or look like one on an acoustic display. This is presumably because there is little or no airflow through the constriction). [d] is especially prone to deletion in the environment n__C.

5 /t/ -- Fully-articulated tokens tend to be found syllable-initially, especially word-initially and especially in stressed syllables. After the alveolar nasal or fricative and intervocalically, [t] can be either fully closed or incomplete. No closure is normally found in the environment C__#C.

For both speakers, /t/ was usually realised as a glottal stop in the environment V__#C and in absolute final position.

4. DISCUSSION

Let us return briefly to the notion of a normal or target articulation. While it is clearly desirable for all speakers to be able to produce a maximally-differentiated set of alveolars in citation-form words in a laboratory situation, it seems obvious from the above that less fully realised tokens are very much a part of conversational speech and are in themselves normal. The implication for those using EPG didactically is obvious: it would be excessively demanding and in some sense even incorrect to expect maximally differentiated tokens of most alveolar consonants (in some environments) in unselfconscious speech. Variation in production which comes about not only through coarticulation with surrounding segments but also through position in the linguistic unit (syllable, word) and position with respect to stress must be taken into account. There might also be a generally lower longterm jaw/tongue setting in conversational speech, which leads to less side contact and bigger fricative grooves, and may be one of the reasons for the observed incomplete closures. (See [2] for further discussion of this question).

The latter point must be reiterated with respect to general phonetic theory: these data provide further evidence for the assertion that the physical properties of the vocal tract alone cannot account for the patterns of reduction we find in conversational speech. An /nt/# sequence behaves very differently from an /n/#/t/ sequence with respect to reduction: it is the higher-level linguistic construct which determines the possibility of phonetic variation, though the construction of the vocal tract is one of several factors which determine the nature of the variation.

BIBLIOGRAPHY

[1] Hardcastle, W.J. and Barry, W., (1989), "Articulatory and Perceptual Factors in /l/ Vocalisation in English," Journal of the International Phonetic Association 15, 3-17.

[2] Hewlett, N. and Shockey, L., "On Types of Coarticulation," in D.R. Ladd and G.J. Docherty (eds), Papers in Laboratory Phonology II, Cambridge University Press, to appear.

PHONOLOGICAL DISRUPTION IN WORD PRODUCTION

Gregory V. Jones

University of Warwick, Coventry, England.

ABSTRACT

In naturally occurring speech, people occasionally find that they have a word "on the tip of the tongue". In this state, they may produce other words related in either sound or meaning to the targets. Are these other words instrumental in causing the TOT states, or are they merely by-products of the TOT states?

1. INTRODUCTION

In spontaneous utterances, most people occasionally experience difficulty in producing an intended word. In this state, a person may be confident that the word he or she wishes to generate is within his or her mental lexicon. The word nevertheless remains temporarily unavailable, seemingly "on the tip of the tongue" While people are in this tip-of-the-tongue (TOT) state, they often do not remain mute but instead produce words other than the target-word at which they are aiming. Such words have been termed "interlopers" [5,6]. An early example was reported by the writer George Lewes, partner of the novelist Mary Ann Evans (George Eliot), as follows.

I was one day relating a visit to the Epileptic Hospital, and intending to name the friend, Dr. Bastian, who accompanied me, I said, "Dr. Brinton;" then immediately corrected this with, "Dr. Bridges," - this also was rejected, and "Dr. Bastian" was pronounced. I was under no confusion whatever as to the persons, but having imperfectly adjusted the group of muscles necessary for the articulation of the one name, the one element which was common to that group and to the others, namely B, served to recall all three [7, p. 128].

Lewes's observation was discussed widely, for example in France by Ribot [11, p. 19] and by Binet [1, pp. 113-114]. However, greater generality was clearly to be obtained by the collection of a corpus of such observations. Early corpora were assembled by Woodworth [14] and Wenzl [12,13]. More recent corpora have been described by Reason and his colleagues [9,10], Cohen and Faulkner [4] and Burke, MacKay, Worthley, and Wade [3]. In all of these studies, a considerable number of TOT states were found to be characterised by the occurrence of interlopers that were related to their respective targets in either their sound or their meaning. The nature of the empirical stochastic contingency between relatedness in sound and relatedness in meaning of an interloper to its target is still, however, unclear. For this reason I have in a recent unpublished study collected a small corpus of naturally occurring TOT states.

2. TOT CORPUS

TOT experiences were collected from undergraduates at the University of Warwick over a period of several weeks. In this sample, the number of interlopers generated by the participants themselves (as opposed to those generated by bystanders) was 100. The interlopers were classified as being related both phonologically and semantically (PS), phonologically

alone (Ps), semantically alone (pS), or neither phonologically nor semantically (ps). The observed incidences were PS = 29, Ps = 3, pS = 67, and ps = 1.

A striking and unexpected aspect of the preceding results was that almost all (96%) of the interlopers were semantically related to their targets. At first sight, this result appears to conflict with the previous observation of many interlopers categorised as phonologically related to their targets [10, p. 124]. However, closer examination of the examples provided by Reason and Mycielska indicates that in each case their "mostle phonological pathways" display semantic relatedness also (e.g., *target* = pomander, *interloper* = pot-pourri).

3. INTERLOPER ORIGINS

What are the origins of the interlopers that commonly occur in TOT states? Two logical possibilities may be distinguished. The interlopers may arise either before or after the disruption in target word generation. In particular, the interlopers may either be instrumental in causing the disruption or be merely a consequence of the disruption. To use medical terminology, the interloper could be considered either as a pathogen (i.e., cause of disruption) or as a sequela (i.e., consequence of disruption).

In the case of words related in meaning, the Sequela hypothesis seems *a priori* plausible. Words produced in normal utterances are presumably selected largely on the basis of their meaning. Thus after a target word becomes unavailable, it might be expected that a person's attempts at word generation will yield other words which are related in meaning to the target. In contrast, the Pathogen hypothesis (that the interlopers themselves cause the disruption) seems implausible. It is obvious that other words related in meaning to intended target words are routinely generated in many normal meaningful utterances (e.g., consider the target word "water" in the sentence "The swimming pool water was chlorinated"). Since we generally have no difficulty in speaking such sentences, we may infer that target generation is not likely to be prevented by the activation of words related in meaning that act in a pathogenic manner.

In the case of interloper words related in sound to the target, it is in contrast difficult to establish their role by *a priori* reasoning. On the Sequela hypothesis, such interlopers might arise if it is the case that, for some unrelated reason, only a partial phonological specification of the target word becomes activated. Subsequently, other words sharing this partial specification might be generated as sequelae. Most would be expected to be also related in meaning to the target, since semantic factors would presumably remain important in guiding word production. On the Pathogen hypothesis, it might be possible for phonological interlopers to act as blockers. Perhaps a word which is similar in sound to the target word receives activation by chance shortly before generation of the target is completed, and acts as a phonological decoy receiving in sum more activation than the target itself. Again, this is clearly more likely to occur if the interloper and target are related in meaning as well as in sound.

4 SOME INTERLOPER EXPERIMENTS

How can one distinguish between the Pathogen and Sequela hypotheses for the origins of phonological interlopers in TOT states? Two recent studies [5,6; see also 8] developed further an experimental method of investigating the TOT state introduced by Brown and McNeil [2]. Brown and McNeill showed that reading people definitions of moderately rare words induces TOT states on the order of 10% of occasions.

In the new studies, people were again presented with definitions of moderately rare words, such as

"Something out of keeping with the times in which it exists". But now the definition was followed immediately by an interloper word also presented by the experimenter. Equal numbers of the four types of interlopers distinguished earlier (PS, Ps, pS, and ps) were used - that is, the interloper was either related both phonologically and semantically to the target, phonologically alone, semantically alone, or neither phonologically nor semantically, respectively. For the present example definition, the interloper was "abnormality". This was of the PS type since it was related in both sound (initial phoneme and number of syllables) and meaning to the target "anachronism".

It was found that interlopers which were related in sound to their targets were more likely to lead to TOT states, irrespective of whether they were related in meaning. This result is consistent with the Pathogen hypothesis since that hypothesis asserts that phonological similarity between interloper and target is instrumental in engendering TOT states, in contrast to the Sequela Hypothesis's assertion that the interloper is merely a by-product of naturally occurring TOT states. Nevertheless, considerable empirical work remains to be done to examine further the effects of artificially supplied interlopers, and in particular more extensive work with a wider range of experimental materials is needed.

5 REFERENCES

[1] BINET, A. (1886). *La psychologie du raisonnement: Recherches expérimentales par l'hypnotisme.* Paris: Germer Baillière.

[2] BROWN, R., & McNEILL, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325-337.

[3] BURKE, D. M., MacKAY, D. G., WORTHLEY, J. S., & Wade, E. (in press). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language.*

[4] COHEN, G., & FAULKNER, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology, 4,* 187-197.

[5] JONES, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition, 17,* 69-76.

[6] JONES, G. V., & LANGFORD, S. (1987). Phonological blocking in the tip of the tongue state. *Cognition, 26,* 115-122.

[7] LEWES, G. H. (1879). *Problems of life and mind* (3rd series, cont.). London: Trübner.

[8] MAYLOR, E. A. (1990). Age, blocking and the tip of the tongue state. *British Journal of Psychology, 81,* 123-134.

[9] REASON, J., & LUCAS, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J. E. Harris & P. E. Morris (Eds.), *Everyday memory, actions and absent-mindedness* (pp. 53-70). Orlando, FL: Academic Press.

[10] REASON, J., & MYCIELSKA, K. (1982). *Absent-minded? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice-Hall.

[11] RIBOT, T. (1881). *Les maladies de la mémoire.* Paris: Germer Baillière.

[12] WENZL, A. (1932). Empirische und theoretische Beiträge zur Erinnerungsarbeit bei erschwerter Wortfindung. *Archiv für die gesamte Psychologie, 85,* 181-218.

[13] WENZL, A. (1936). Empirische und theoretische Beiträge zur Erinnerungsarbeit bei erschwerter Wortfindung. *Archiv für die gesamte Psychologie, 97,* 294-318.

[14] WOODWORTH, R. S. (1929). *Psychology* (2nd rev. ed.). New York: Holt.

ARTICULATORY-ACOUSTIC CORRELATIONS IN THE PRODUCTION OF FRICATIVES

N. Nguyen-Trong[1], P. Hoole[2], & A. Marchal[1]

1. CNRS, URA 261, Univ. Provence, Aix-en-Provence, France
2. Institut für Phonetik, München Univ., FRG

ABSTRACT

This work is aimed at exploring the relationships between a set of articulatory parameters, and the acoustic output, in the production of French fricatives /s, S/. More precisely, we attempt to find out whether the dimensions of maximal contrast among the fricative spectra, are correlated with movements of lingual transducers monitored by means of an electromagnetic (EMA) system. Our results show that the EMA measurements can be considered to be very reliable. It appears that the spectra can be regenerated with a good accuracy from these measurements, with the help of a statistical method the advantages of which are pointed out. In conclusion, implications of this work in the domain of articulatory modelling are discussed.

1. INTRODUCTION

In a recent work [3], Hoole et al. have shown that the EPG tongue-contact patterns in the production of the fricatives /s/ and /S/ in English, were strongly correlated with a set of acoustic parameters extracted from the corresponding spectra by means of a factorial analysis. It has appeared that this relationship was close enough to allow a prediction of the acoustic data from the EPG data, with the help of a multiple linear regression. The results supported the conclusion that an empirical investigation of this kind, was suitable for providing information on the articulatory-acoustic correlations, which could be fruitfully incorporated into a model of fricative production [2,5]. The present experiment was based on the same methodological principles, and was aimed at investigating in a more extensive way two specific points. First, the question could be raised to know whether it is possible to relate the articulatory parameters with the spectra themselves by regenerating these spectra from the acoustic factors. Second, it seemed important to compare the results of the multiple linear regression, with those of a method giving the possibility to detect non linear articulatory-acoustic relationships similar to the ones which are described in Stevens' quantal theory [6].

2. MATERIAL

The experiment has been carried out at the Institute of Phonetics of Munich University. It consisted of an audio recording synchronized with a parallel EMA tracking. The electromagnetic system used is a commercially available device (Articulograph AG 100, Carstens, Göttingen, FRG) which has been recently assessed in [4,7], and which allows monitoring of articulatory movements with the help of five electromagnetic transducers (coils). In the present experiment, three coils were attached to the mid-line of the tongue, one was attached to the lower incisors, and one reference coil to the upper incisors. The tongue rearmost coil was placed as far back as possible, the frontmost coil about 1 cm back from the tip, and the third coil in between these two. The output of the EMA system was digitized (sampling rate 250 Hz) and transferred online to a PC AT-386 computer where software compensation for the effects of possible tilt of the receiver coils was applied. The digital signal, which represented the displacements of each of the five coils in the x-y plane, was finally stored on a hard disk. The audio signal was recorded by means

of a B&K microphone on a DAT recorder, digitized on a LSI 11/73 computer (sampling rate 16000 Hz, LP filtered at 7500 Hz), and aligned with the EMA signal thanks to a set of synchronization pulses recorded on the second track of the DAT tape. The estimated accuracy of the alignment was +/-3 ms. The speech material consisted of the following combinations: /aS#/, /aSa/, /iSa/, /#sa/, /#si/, /as#/, /asa/, /asi/, /isa/, /isi/ embedded in 9 sentences which have been pronounced from 8 to 9 times by two male native speakers of French (AM, NN). In this paper, results will be presented for speaker AM.

3. ANALYTICAL PROCEDURES

3.1. EMA measurements

To minimize the variations in the articulatory signal which could have been generated by any head movements, the coordinates of the coil affixed to the upper incisors has been subtracted from those of the other coils. Moreover, for each repetition, the whole cloud of data has been rotated around the origin in the x-y plane, so as to achieve a vertical orientation of the first principal axis of the jaw movement. Figure 1 displays the positions, averaged over all the repetitions, of the tongue-back, tongue-mid, tongue-tip and jaw coils, at the mid-point of the fricative (which has a coordinate on the time axis determined with the help of the acoustic signal) for each item.

3.1. Acoustic analysis

The acoustic analysis consisted in calculating an FFT spectrum within a 32 ms Hamming window centered at the fricative mid-point. This spectrum was next reduced to 21 components by averaging the spectral energy over 1 Bark intervals from 0 to 8 kHz. Moreover, the information below 8 Barks has been ignored, in such a way that the acoustic data were finally made up in the present experiment of a set of 13-dimensional vectors. For reasons that are given below, we have chosen to proceed to a new data reduction by means of a principal-components analysis, which proved that 4 linear combinations of the 13 original parameters could account for more than 90% of the variance among the spectra. It has appeared that the 2 first factors were sufficient to differentiate the fricatives /s/ and /S/ from each other. Factor 1 is interpretable as a dimension of average energy; factor 2 can be considered as underlying an opposition between the spectra which show a local maximum within the 12-15 Barks range, and those in which the energy is relatively higher above 15 Barks.

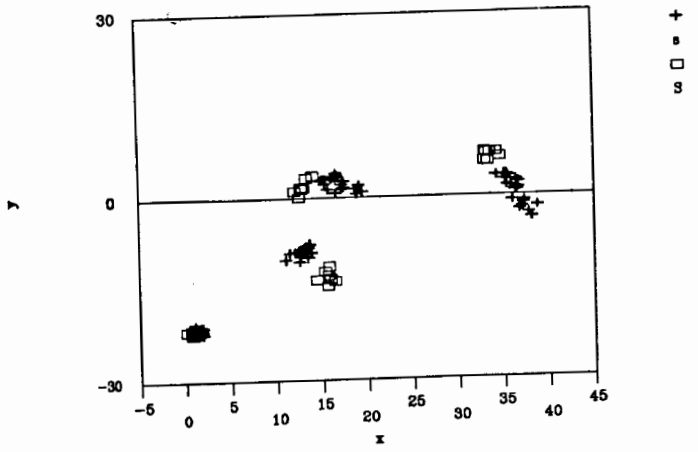


FIGURE 1

4. RESULTS

4.1. empirical regression of the acoustic factors from the EMA data

One approach to exploring EMA-acoustic relationships consists of analyzing separately the way in which the data are distributed in the EMA space, and in the acoustic space, to examine whether the sources of variation can be considered as being the same in the two cases (pronounced consonant, immediate context, carrier sentence, etc.). But it is also possible to check the existence of such relationships, by attempting to predict the acoustic parameters from the articulatory ones. At the present time, the information obtained with EMA on the midsagittal section of the vocal tract, doesn't allow calculation of the area function required by a standard acoustic model to resynthesize the output signal. In this experiment, the predictions have been based on the so-called statistical regression, which has been performed in two different ways, since we have compared the results of a classical, multiple linear regression, with those of an empirical, non linear variant [1]. In the second case, the predicted value of a given acoustic parameter **y** for a given articulatory input (a tongue «profile» composed of 3 points) was simply defined as the **y** mean value

TABLE 1: correlation coefficients between measured and predicted values for the first 4 acoustic factors (results given for two different regression techniques).

	REG.1	REG.2
fac.1	0.765	0.882
fac.2	0.851	0.925
fac.3	0.722	0.877
fac.4	0.384	0.536

associated with the input k-nearest neighbours in the articulatory space (k being determined by the user). It can be easily shown that the regression achieved by means of such a local approximation, is suitable for modelling non linear relationships, between any number of independent variables and the to-be-predicted one.

The calculations have involved the articulatory and acoustic data relative to all the (V)C(V) sequences recorded by speaker AM. The prediction quality has then been assessed on a test set which was composed of the tongue «profiles» averaged over all the repetitions for each of the two consonants, in each possible context. In table 1 are given the *r* 's corresponding to the correlation between the measured and the predicted values for factors 1, 2, 3 and 4. It appears clearly that the *r*'s are high whatever regression technique is used, and reflect a close relationship between the EMA parameters and the dimensions of maximal variance among the spectra, while the empirical regression (referred to as REG.2 in table 1) produces results which are systematically superior to those of the multiple linear regression (REG.1). The number of neighbours, on the basis of which a value has been given to each tongue profile for each of the 4 factors in REG.2, was fixed to 5.

4.2. Regeneration of the original spectra from the acoustic factors

The fricative spectra have been finally regenerated from the output parameters of the empirical regression, through the usual operation based in the present case on the eigenvectors of the covariance matrix relative to the 14 original acoustic variables [8]. Figure 2 proves that the accuracy of this regeneration is quite good (average *r* between the measured and predicted components for each of the spectra in the test set >0.9). It is especially encouraging to note that the shape shown by a spectrum is restored in a very satisfying way.

6. CONCLUSION

Our results could be explained in part by the fact that speaker AM has pronounced /s/ and /S/ in a rather stable way across repetitions and across contexts. Consequently, the variability among the spectra was likely to be accounted for by a relatively small

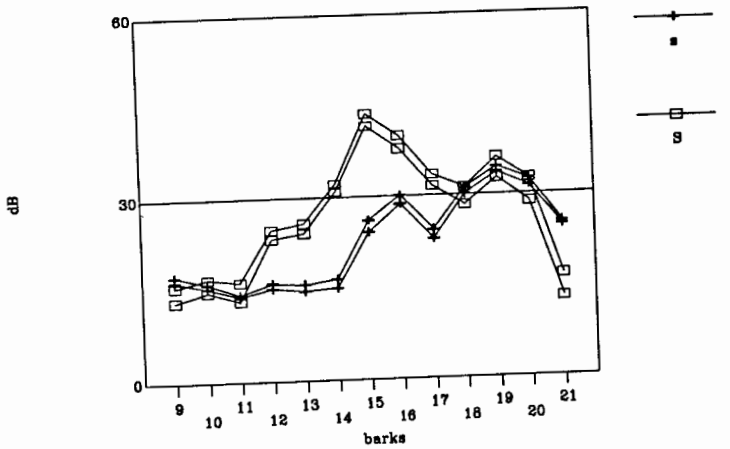


FIGURE 2: average spectra for /s/ and /S/ displayed together with the output of the empirical regression from the corresponding EMA profiles.

number of factors which in return allowed to regenerate these spectra without any major distortion. It remains that the factors themselves (which can be considered as dimensions of maximal contrast among the acoustic data) have proven to be accurately predictable from the EMA measurements. Therefore, it can be said that under the conditions adopted in this experiment, the articulatory parameters extracted by EMA are closely correlated with the acoustic output.

The empirical regression is very interesting in that it is gives the possibility to make predictions which have an accuracy calculated for each point in the articulatory space (while this calculation wouldn't have much sense in the case of the linear regression, in which the criterion to be optimized concerns an average accuracy). From our point of vue, this issue should give rise to a more systematic investigation. An experiment carried out according to the same methodological principles as those of the present work, on a more extensive material, would probably allow to determine whether a given articulatory neighborhood is stable (with respect to the acoustic output), or unstable. It seems to us that this kind of empirical exploration of the articulatory space, could constitute an quite interesting way to verify hypotheses on the mechanisms of speech production, such as the ones which are supported in the quantal theory [6].

7. ACKNOWLEDGMENTS

This work was supported by ESPRIT II/ BRA n°3279 ACCOR.

9. REFERENCES

[1] CAZES, P. (1976). Rev. Stat. Appl., 24, 5-57.
[2] FANT, G. (1960). Acoustic Theory of Speech Production (Mouton, The Hague).
[3] HOOLE, P., ZIEGLER, W., HARTMANN, E., and HARDCASTLE, W.J. (1989). Clin. Ling. & Phon., 3, 59-69.
[4] SCHÖNLE, P. MÜLLER, C., WENIG, P., HÖHNE, J., SCHRADER, J., and CONRAD, B. (1987). Brain & Lang., 31, 26-35.
[5] SHADLE, C.H. (1985). «The acoustics of fricative consonants», Ph. D. Diss. (MIT).
[6] STEVENS, K.N. (1989). J. Phon., 17, 3-45.
[7] TULLER, B., SHAO, S., and KELSO, J.A.S., (1990). J. Acous. Soc. Am., 88(2), 674-679.
[8] ZAHORIAN, S.A., and ROTHENBERG, M. (1981). J. Acous. Soc. Am., 69, 832-845.

INITIALISATION, ARRET ET VARIATION DE FREQUENCE FONDAMENTALE DE LA VIBRATION LARYNGEE : ETUDE ELECTROMYOGRAPHIQUE

B. Roubeau, G. Dassau, J. Lacau et C. Chevrie-Muller

INSERM, Laboratoire de Recherche sur le Langage
Paris, France.

ABSTRACT
In order to analyze the mechanisms of fundamental frequency controls, the electromyogram (EMG) of four laryngeal muscles (Cricothyroïd and three strap muscles) was recorded, as well as the acoustic signal, in two subjects (one male, one female) producing ascending and descending tones (glissandi). The relationship between EMG activity patterns and frequency variations were described ; in addition, the specific patterns related to the glissando's beginning and terminal part were analyzed. According to the different vocal events taken into consideration the EMG patterns of the four muscles were compared.

Les études électromyographiques des muscles laryngés corrélées avec la fréquence vibratoire sont nombreuses (2, 3, 6, 1). Celles consacrées aux "mouvements" de fréquence sont déjà plus rares de même que celles qui considèrent parallèlement l'activité de muscles intrinsèques du larynx et de muscles sous-hyoïdiens (5).

1. PROTOCOLE EXPERIMENTAL

Le protocole vocal comporte des variations de fréquences au cours de la réalisation sur la voyelle "O", d'une part, de glissandos ascendant puis descendant, et d'autre part descendant puis ascendant sans contrainte de hauteur limite ni d'intensité. La seule contrainte imposée aux sujets est le maintien de la tête dans une position aussi fixe que possible. Au cours de nos expériences, l'activité de 4 muscles a été enregistrée :
- le cricothyroïdien (CT), le thyrohyoïdien (TH), le sternothyroïdien (ST) et le Sterno-cleïdo-hyoïdien (SCH).

La technique EMG employée est celle décrite par Hirose (5) utilisant des électrodes bipolaires implantées à l'aide d'aiguilles intramusculaires. Des tests non vocaux tels que la déglutition, l'ouverture de la bouche et l'inclinaison de la tête, permettent de vérifier l'emplacement des électrodes. Ces tests sont répétés plusieurs fois au cours de l'enregistrement, afin de vérifier le maintien en place des électrodes.
Les résultats provenant de deux sujets (une femme et un homme) parmi les 6 qui ont participé à l'expérience ont été retenus pour la stabilité des tracés fournis par les 4 muscles explorés et seront présentés ici.

2. RESULTATS ET DISCUSSION

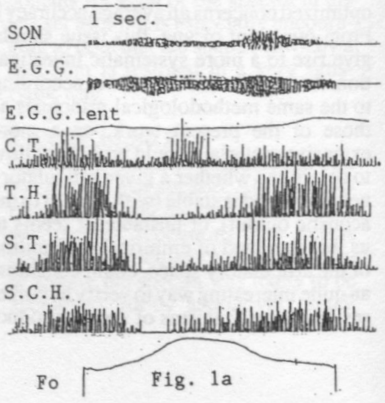


Fig. 1a

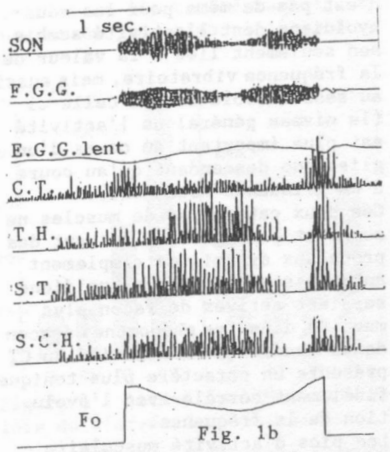


Fig. 1b

Figure 1 : Tracés EMG redressés des 4 muscles explorés ainsi que le signal acoustique, le signal électroglottographique (EGG) et la courbe mélodique a) Glissando ascendant et descendant, b) Glissando descendant puis ascendant. Sujet masculin CT, TH, ST et SCH. La courbe mélodique (en bas) représente l'évolution de la fréquence caractéristique d'un glissando.

L'activité EMG semble corrélée à plusieurs évènements que nous analyserons séparément : la variation de fréquence, l'initialisation et l'arrêt de la vibration.
Les mesures d'activité ont été effectuées sur les signaux redressés puis intégrés et lissés.

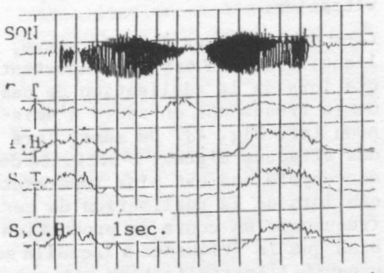
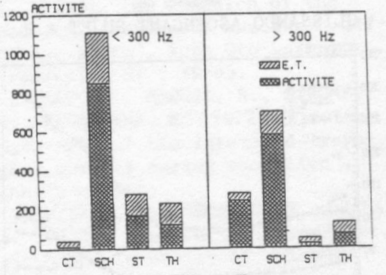
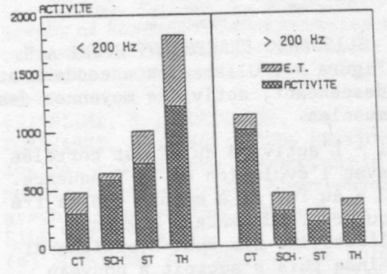


Figure 2 : Glissando ascendant puis descendant, sujet masculin, signal intégré.

On note une activité importante des sous-hyoïdiens lors de la réalisation des fréquences les plus basses au début et à la fin de la production, de même qu'une activité importante du CT précédant la réalisation des fréquences les plus élevées. Les pics d'activité musculaire précédant les points d'inféxion de la courbe mélodique ont été relevés et moyennés. Pour le sujet masculin (DA), les activités EMG sont moyennées pour des valeurs de F_o inférieures et supérieures à 200 Hz, et inférieures et supérieures à 300 Hz pour le sujet féminin (AH).



SUJET A.H.

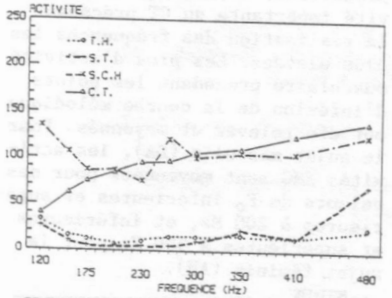


SUJET D.A.

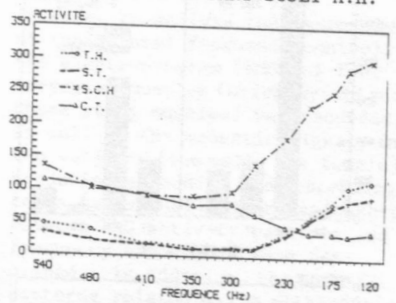
Figure 3 : Activités musculaires moyennes aux points d'inflexion de la courbe mélodique.

Ces résultats montrent l'importance de l'activité du CT lors de la réalisation des fréquences élevées et des sous-hyoïdiens lors de la réalisation des fréquences basses.
Les courbes d'activité des muscles en fonction de l'évolution dynamique de la fréquence ont été réalisées à partir de moyennes effectuées sur 34 productions pour le sujet féminin AH et 28 pour le sujet masculin DA.
Ces valeurs sont relevées après élimination des phénomènes d'ini-

tialisation et d'arrêt de l'émission.



GLISSANDO ASCENDANT SUJET A.H.



GLISSANDO DESCENDANT SUJET A.H.

Figure 4 : Glissandos ascendant et descendant, activités moyennes des muscles.

1 - L'activité du CT est corrélée avec l'évolution de la fréquence.
2 - Au fur et à mesure que la fréquence fondamentale augmente, l'activité des sous-hyoïdiens diminue puis s'accroît à nouveau lors de la réalisation des fréquences les plus élevées.
3 - De même, les glissandos descendants mettent en évidence au cours de la réalisation des fréquences élevées une activité non négligeable des sous-hyoïdiens. Celle-ci s'accroît considérablement lors de la réalisation des fréquences les plus basses. Il faut évidemment considérer le fait que l'anticipation de l'activité pour la réalisation d'une fréquence donnée est ici difficile à évaluer.
L'activité du CT apparaît corrélée de manière relativement stable avec la fréquence fondamentale (glissandos ascendant et descendant sont assez bien symétriques). Il n'en n'est pas de même pour les sous-hyoïdiens dont l'activité semble non seulement liée à la valeur de la fréquence vibratoire, mais aussi au sens d'évolution de celle-ci (le niveau général de l'activité est plus important au cours d'un glissando descendant qu'au cours d'un glissando ascendant).
Ces deux catégories de muscles ne semblent pas agir ici suivant des processus dynamiques simplement antagonistes. Les sous-hyoïdiens seraient activés de façon plus massive dans les phénomènes descendants tandis que l'activité du CT présente un caractère plus tonique fidèlement corrélé avec l'évolution de la fréquence.
Les pics d'activité musculaire précédant l'attaque ont été moyennés et regroupés en fonction de la hauteur du son au moment de celle-ci (cf. Fig. 1).

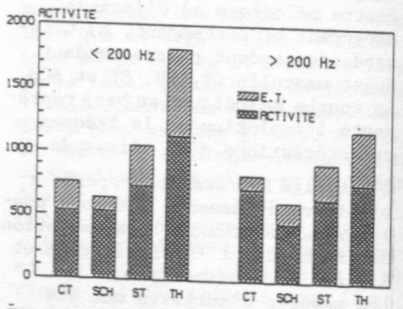


Figure 5 : Activité EMG moyennée lors de l'initialisation de la vibration.

Lors des attaques dans le grave, l'activité du CT est plus importante que lors de la réalisation de ces mêmes fréquences en cours d'émission (cf. Fig. 4). De même, lors des attaques aigües, l'activité des sous-hyoïdiens est plus importante que lors de la réalisation de ces fréquences en cours de production. Bien que la fréquence à laquelle se produit l'initialisation ait une influence sur l'amplitude du tracé EMG (CT plus actif dans les attaques supérieures à 200 Hz et TH plus actif pour des attaques inférieures à 200 Hz), le phénomène d'initialisation lui-même provoque une globalisation de l'activité musculaire suivant un processus plutôt phasique.

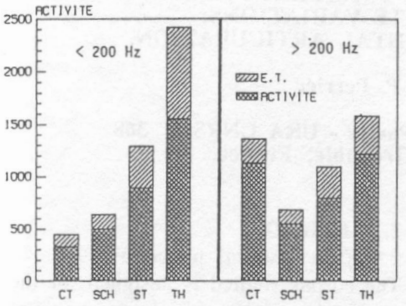


Figure 6 : Activités EMG moyennées lors de l'arrêt de l'émission.

On note ici une différence de comportement des deux sujets. Chez AH, l'activité lors de l'arrêt de l'émission demeure liée principalement à la fréquence vibratoire telle qu'on l'observe en cours d'émission. Par contre, chez DA, les arrêts dans l'aigü indiquent une forte activité des sous-hyoïdiens supérieure à celle que l'on observe en cours de production (cf. Fig. 4). Ce phénomène délicat à interpréter à partir de deux sujets peut être considéré comme une régulation de la fréquence ou des mouvements liés à son évolution grâce à un système d'"antagonisme" agissant sur les mobilisations générales du larynx.
Ces différentes observations sur l'initialisation et l'arrêt de l'émission mettent en évidence le caractère phasique de l'activité musculaire tandis que la régulation de la fréquence en cours d'émission est liée à une activité tonique, cette distinction est particulièrement nette au niveau du CT. Les différences entre les deux sujets semblent liées soit aux durées de réalisations des glissandos qui sont considérablement pour courtes chez le sujet masculin soit à l'utilisation de différents mécanismes vibratoires ou registres.
Dans tous les cas observés, les sous-hyoïdiens semblent fonctionner en synergie. Il ne faut pas négliger dans cette interprétation, la possibilité d'un "parasitage" des signaux EMG entre eux du fait de la proximité des muscles.
Cette étude confirme l'importance de la considération simultanée des muscles intrinsèques et extrinsèques du larynx lors de l'étude des variations mélodiques et de la distinction des évènements mettant en jeu des activités musculaires de caractère tonique ou phasique.

3. BIBLIOGRAPHIE

(1) FAABORG-ANDERSEN, K, SONNINEN, A. (1960) "The function of the extrinsic laryngeal muscles at different pitch", Acta Oto-Laryngol., Stockholm, 51 : 89-93.
(2) GAY, T., HIROSE, H., STROME, M., SAWASHIMA, M. (1972)."Electromyography of the intrinsic laryngeal muscles during phonation", Annals of ORL.
(3) HIRANO, M., VENNARD, W., OHALA, J. (1970), "Regulation of register, pitch and intensity of voice", Folia Phoniat., 22 : 1-20.
(4) HIROSE, H. (1971). "Electromyography of the articulatory muscles : current instrumentation and techniques", Haskins Labs. RS-25/26, 73-86.
(5) NIIMI, S., HORIGUCHI, S., KOBAYASHI, N. (1988). "The Physiological role of the sternothyroid muscle in phonation and electromyographic observation", Ann. Bull. RILP, 22 : 155-172.
(6) SONNINEN, A. (1956). "The role of the external laryngeal muscles in length-adjustments of the vocal cords in singing", Acta Oto-Laryngol., Suppl. 130.

PHASE MODIFICATIONS IN TONGUE MOVEMENTS ACROSS SPEECH RATE VARIATIONS: INFLUENCE OF CONSONANTAL ARTICULATION.

C. Delattre & P. Perrier

Institut de la Communication Parlée - URA CNRS n° 368
Université Stendhal,Grenoble, France.

ABSTRACT

A corpus of CV_1CV_2 sequences is analysed in order to emphasize the concept of synergy among gestures in speech production. The vertical movements of tongue dorsum were measured for V_1=[a] and V_2=[u], and for two consonants [d] and [g] which production recruit respectively this articulator at two quite different levels. Results are interpreted in terms of changes of (1) the consonant-vocalic and (2) the peak velocity phasing, for raising and lowering gestures.

1. INTRODUCTION

Speech production implies a spatial and temporal coordination of different articulatory gestures. The concept of synergy, introduced by Haken [3] in the field of motor control, could be appropriate to characterize and predict some articulatory patterns of speech production. In their work, Kelso et al. [5] explained the jump from the production [ipip] to the production [pipi] as the speech rate strongly increases, within this theoretical framework. They introduced the concept of intergestural phasing, and identified this jump as an obvious phasing restructuration between lip and glottal gestures.

In the present study, we propose to analyse this synergy phenomenon in a quite different paradigm. The behaviour of the consonant-vowel phasing in CV_1CV_2 sequences is observed for consonants involving different articulators. This paper presents preliminary results for two consonants [d] and [g].

2. METHOD

2.1 Experimental procedure

The corpus studied is designed for the observation of the vertical movements of tongue dorsum. It consists in the repetitions of isolated productions [au], [dadu], [gagu], for which we assume that tongue dorsum movement is pertinent to describe the [a] to [u] articulation. The consonants [d] and [g] are chosen, because of their two extreme behaviours towards this articulator: [d] production does not recruit the tongue, whereas the [g] articulation recruit mainly the tongue dorsum. Thus, in a [gagu] sequence, the vocalic and consonantal gestures are produced with the same articulator.

The task consisted in 10 repetitions of each sequence, produced at normal and fast speaking rates by a French male speaker. Tongue dorsum displacement was monitored at an 1 kHz rate with a computerized ultrasound transducer system (see [4]). Simultaneously the acoustic speech signal was recorded at the same rate.

2.2 Data collection

As data are available for one articulator only, namely the tongue dorsum, the choice of temporal events defining the vocalic and consonantal phases was not obvious: the events related to consonantal gestures were detected on the speech wave. In all cases the reference is the underlying vowel-to-vowel movement, divided into two components, the raising gesture ([a]-[u]) and the lowering gesture ([u]-[a]). For the first component, the consonantal movement towards the occlusion is in the same raising direction as the vocalic movement; the release movement is in the opposite direction. In the second component ([u]-[a] transition), the relations are reversed. It is thus interesting to detect these two consonantal events (closion and release) inside the vocalic phase. For each vocalic transition ([a]-[u] and [u]-[a]) the boundaries of the vocalic phase correspond to the points of zero velocity of the tongue dorsum raising gesture (See Fig. 1). We define so for each movement (raising and lowering) one vocalic phase and two consonantal phases, the "occlusion phase" and the "release phase".

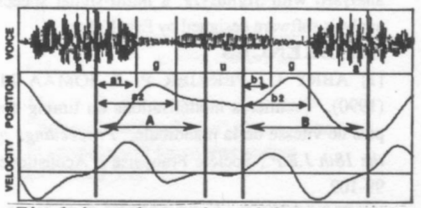


Fig.1 shows ultrasound recorded movement and velocity profile of tongue dorsum and the corresponding acoustic speech signal, during the production of /gagu/, at normal speaking rate. Duration A determines the raising gesture for the transition [a]-[u], in which a1 ("the occlusion phase") and b2 ("the release phase") are plotted in percentage. Duration B determines the lowering gesture for the transition [u]-[a], in which the same phases ("the occluion phase"=b1) and ("the release phase"=b2) are plotted.

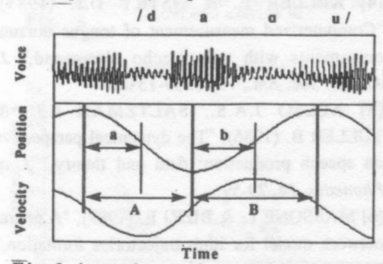


Fig.2 shows ultrasound recorded movement and velocity profile of tongue dorsum and the corresponding acoustic speech signal, during the production of /dadu/, at fast speaking rate. Duration A and B were measured assessing that points of zero velocity determine the initiation and termination of the gesture. The acceleration phases, duration a and b are plotted in percentage.

The second point of our study focuses on the kinematic changes of the raising and lowering gestures depending on the consonant and on the speech rate. In this aim and according to different studies that advocate the importance of velocity profiles in motor control analysis (see [1], [2] and [6]), the occurence of the velocity peak between two successive points of zero velocity is measured (see Fig. 2).

3. PRELIMINARY RESULTS AND DISCUSSIONS

3.1. Consonantal-vocalic phasing

The data obtained for the raising gesture are presented in Fig. 3 and 4. The examination of the time proportion of the release phase for [dadu] (Fig.3), reflects obviously no significant modification with changes in speech rate (mean values: 74.9 vs 74.2); on the contrary, for [gagu], we observe an obvious increase of this time proportion for the fast rate. Such a behaviour can easily be explained by a time constraint on the consonantal hold: this durational value must be sufficient for a good perception of the consonant; hence the vocalic durations are more affected by the change in speech rate than the consonantal one. This phenomenon becomes indeed more obvious, when vowel and consonant are produced with the same articulator. At the same time the proportion of the occlusion phase decreases in fast rate for both consonants (Fig. 4). Moreover, whereas the differences between [d] and [g] are not significant (α>0.10) at normal rate, they become highly significant at fast rate (α<0.01), which means that the articulation of the consonant induces different behaviours when speech rate increases. As the raising vocalic gesture ([a]-[u]) and the consonantal raising occlusion gesture occur simultaneously, the delay between the onsets of these two gestures tends to decrease significantly, especially when the same articulator is recruited. These results attest, in the case of a monoarticulator production, a tendency towards synchronization of the

two raising gestures, when the speech rate constraints are strong. Hence, it supports the idea of a synergetic production.

The same kind of observations can be made for the release phase in the lowering vocalic transition. For both consonants, the proportion of the release phase increases (Fig. 5), as the proportion of the occlusion phase decreases (Fig. 6). But these behaviours are hardly significantly different for the occlusion phase ($\alpha > 0.05$) and highly significantly different for the release phase ($\alpha < 0.01$). The delay between the onsets of the lowering vocalic gesture and the consonantal one remains important, due to the constraints on the consonantal hold duration; but a tendency towards synchronization of these two gestures could well furnish a reliable explanation for the more important reduction of the consonantal hold, in the case of a monoarticulator production. This phenomenon supports the hypothesis of synergy among consonantal and vocalic gestures.

3.2. Kinematic changes

This investigation is essentially based on the duration of the lowering and raising gestures in which the occurence of peak velocity is observed (see Fig. 2). In both gestures, these acceleration phases are plotted in Fig.7 & 8. At normal rate, and for both gestures, our results show an important dispersion of the data for [au] and [dadu]; the constraints seem obviously stronger for the [gagu] production. For fast speech rate, the data converge towards the same value in all cases: the velocity profile tends to become symmetric as in optimized movements minimizing the jerk (see [7]). This variation is however less important in [gagu]. An increase in speech rate seems thus to produce an optimization of the coordination between vocalic and consonantal gestures. This optimization is already perceptible for [gagu] at normal rate.

3.3. First conclusions

These two kinds of data seem to attest the existence of synergy between consonantal and vocalic gestures when the same articulator is recruited: (1) the consonant-vowel phasing is specific for this kind of production; (2) in this last case, the kinematic properties reflect a tendency towards optimization. These results and conclusions are preliminary. A further study will be made with other consonants as [R], [b], [k], to confirm the assumptions resulting from the observation of [d] and [g].

ACKNOLEDGEMENTS

We are very grateful to Eric Keller, Université du Québec à Montréal, who allowed us to use his ultrasound system and for his very efficient assistance in data collection. The data were analysed with *Signalyze*, a multi-signal speech analysis software designed by Eric Keller.

REFERENCES

[1] ABRY C., PERRIER P. & JOMAA M. (1990), "Premières modélisations du timing des pics de vitesse de la mandibule," *Proceedings of the 18th J.E.P.*, Société Française d'Acoustique, 99-102.

[2] BULLOCK D. & GROSSBERG S. (1988), "Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation," *Psychological Review, 91*, 1, 49-90.

[3] HAKEN H. (1977), "*Synergetics: an introduction. Nonequilibrium phases transitions and self-organization in physics, chemistry, and biology,*" Heidelberg, Springer-Verlag.

[4] KELLER E. & OSTRY D.J. (1983), "Computerized measurement of tongue dorsum movements with pulsed-echo ultrasound," *J. Acoust. Soc. Am., 73*, 1309-1315.

[5] KELSO J.A.S., SALTZMAN E.L. & TULLER B. (1986), "The dynamical perspective on speech production: data and theory," *J. of Phonetics, 14*, 29-59.

[6] MASSONE L. & BIZZI E. (1989), "A neural network model for limb trajectories formation," *Biol. Cybern.*, 1-9.

[7] NELSON W.L. (1983), "Physical principles for economies of skilled movements," *Biol. Cybern.*,46,135-1470.

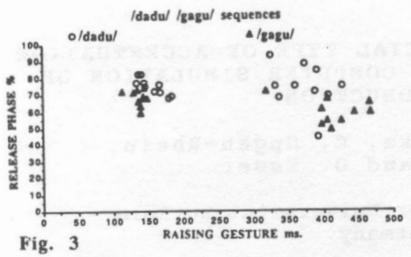


Fig. 3

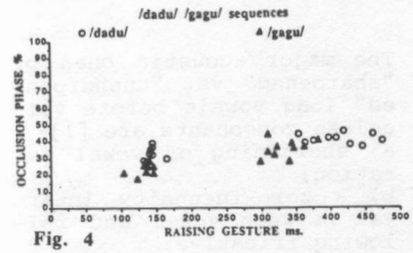
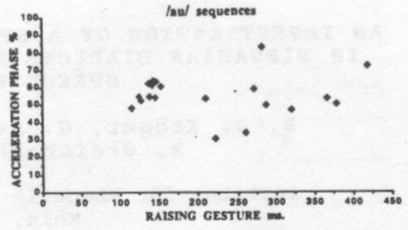


Fig. 4

Fig. 3 & 4
"Release" and "occlusion" phases plotted in percentage of the raising gesture (see text § 3.1)



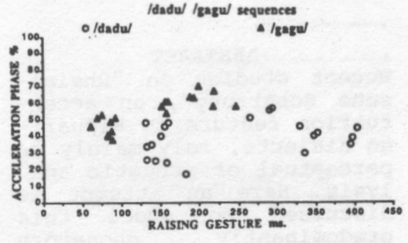


Fig. 7
"Acceleration phase" plotted in percentage of the raising gesture (see text § 3.2)

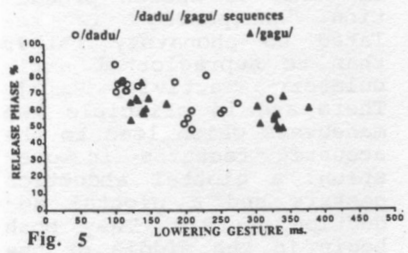


Fig. 5

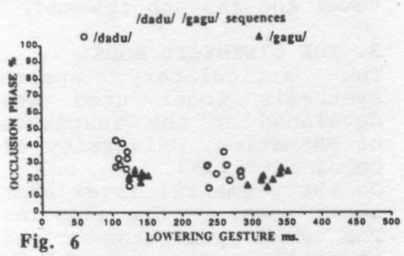
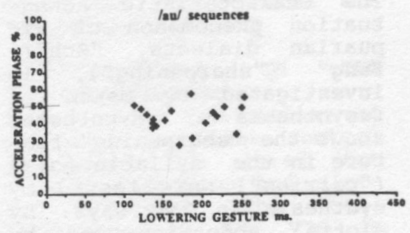


Fig. 6

Fig. 5 & 6
"Release" and "occlusion" phases plotted in percentage of the lowering gesture (see text § 3.1)



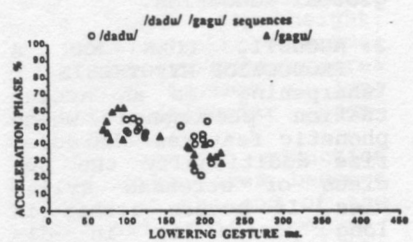


Fig. 8
"Acceleration phase" plotted in percentage of the lowering gesture (see text § 3.2)

AN INVESTIGATION OF A SPECIAL TYPE OF ACCENTUATION IN RIPUARIAN DIALECTS BY COMPUTER SIMULATION OF SPEECH PRODUCTION

B. J. Kröger, G. Heike, C. Opgen-Rhein, R. Greisbach and O. Esser

Institut für Phonetik der Universität zu Köln, Köln, Germany.

ABSTRACT

Recent studies on "Rheinische Schärfung", an accentuation feature in Ripuarian dialects, rely mainly on perceptual or acoustic analysis. Here an attempt is discussed to model this predominantly phonatory phenomenon by coupling a two-mass-model of the glottis to the Cologne articulatory speech synthesis system.

1. INTRODUCTION

The characteristic accentuation phenomenon of Ripuarian dialects, "Schärfung" ("sharpening"), is investigated by means of resynthesis. Hypotheses about the "sharpening" feature in the syllable [ɔ:s] ("carrion") were tested by synthesis in two ways: by glottal abduction and by glottal adduction.

2. ACOUSTIC CUES AND A PRODUCTION HYPOTHESIS

"sharpening" is an accentuation phenomenon whose phonetic features characterize additionally the nucleus of stressed syllables. It occurs either in long vowels or in diphthongs or in short vowels followed by a sonorant ([hy:] "height", [zɛi] "sieve", [ɔ:s] "carrion", [vo:t] "rage", [al] "all", [hyŋ]"dogs", [tant]"aunt")

The major acoustic cues of "sharpened" vs. "unsharpened" long vowels before voiceless consonants are [1]:

a) shortening of vowel duration;

b) a zero-intensity interval between vowel and following fricative;

c) a marked intensity decrease in the vowel segment;

d) a marked decrease in fundamental frequency in the vowel segment.

In terms of speech production "sharpening" is related to phonatory rather than to supraglottal articulatory activity [2]. There are in principle two maneuvers which lead to the acoustic features in question: a glottal abduction gesture and a glottal adduction gesture. They both begin in the middle of the vowel and last to its end.

3. THE SYNTHESIS MODEL

The articulatory speech synthesis model used was developed at the Institute of Phonetics, University of Cologne [3,4,5].

On the segmental level phonetic segments are put in. The model generates a set of articulatory control parameters (e.g. tongue height, tongue position, jaw opening) and three phonatory control parameters (lung pressure, cord tension, abduction area) which are continuous in time.

The vocal tract is modelled with the Kelly-Lochbaum reflection-type line model; it simulates the wave propagation in the vocal tract by scattering partial waves at impedance discontinuities.

The self-oscillating glottis model is of major importance in this context. It comprises a static (non-oscillating) part or bypass and an oscillating part, the two-mass model [6]. While the two masses of the latter differentiate between the motion in the lower and the upper glottal region, the combination of the static plus the oscillating part differentiates between the motion in the posterior and anterior part. The bypass models the posterior portion, the less flexible part of the vocal folds. The control parameter abduction area represents the vocal fold position and changes the opening area of the bypass and the phonation neutral area of the two-mass model. Negative abduction area produces a glottal state in which the vocal folds are compressed medially.

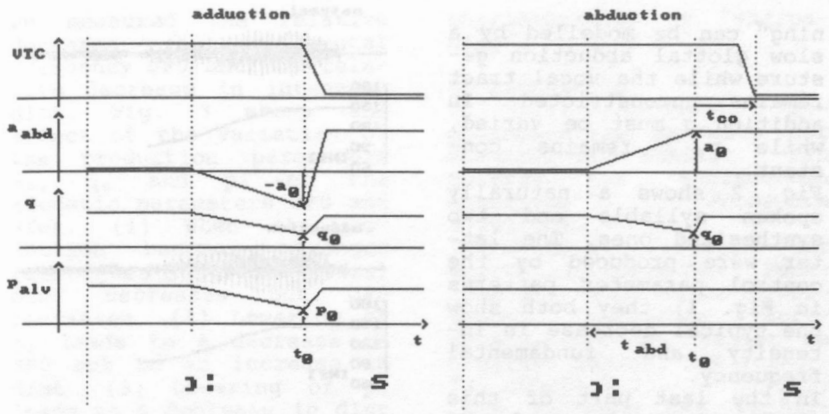


Figure 1: The control and production parameters for the "sharpening" for glottal abduction and glottal adduction.

4. "SHARPENING" IN THE PRODUCTION MODEL

Fig. 1 shows the pattern of the most important control parameters for [ɔ:s] for glottal abduction and glottal adduction. The vocal tract constriction VTC is given by the control parameter tongue tip height. The other control parameters are: abduction area a_{abd}, cord tension q, lung pressure (alveolar pressure) p_{alv}.

In the case of adduction the abduction area becomes negative at the end of the vowel, and medial compression is produced. At the same time, cord tension and lung pressure decrease, resulting in the decrease of F0 and signal intensity.

In the other case "sharpe-

ning" can be modelled by a slow glottal abduction gesture while the vocal tract remains unconstricted. In addition q must be varied, while p_{alv} remains constant.
Fig. 2 shows a naturally spoken syllable and two synthesized ones. The latter were produced by the control parameter patterns in Fig. 1; they both show the typical decrease in intensity and fundamental frequency.
In the last part of this study we investigated the influence of the most important production parameters on the acoustic cues of "sharpening".
In the case of glottal adduction these are the abduction area a_0, the cord tension q_0 and the lung pressure p_0 at instant t_0 (see Fig. 1). The time t_0 is defined as the instant at which the adduction/abduction gesture ends.

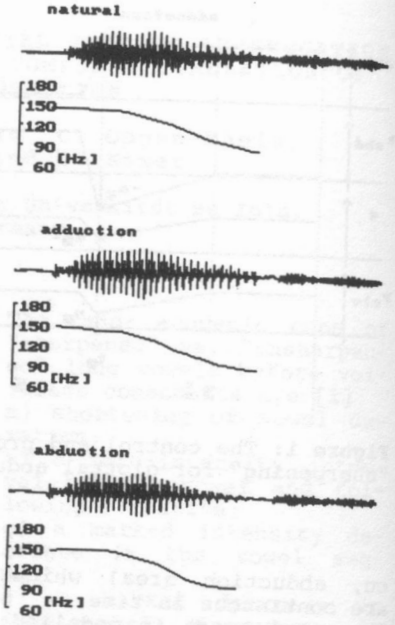


Figure 2: Resynthesis of "sharpening" in [ɔ:s] by glottal adduction and glottal abduction.

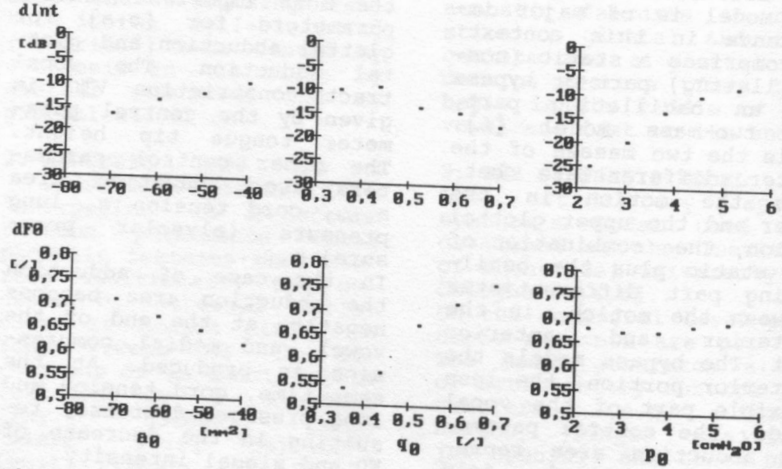


Figure 3: Acoustic parameters as function of production parameters. Case: adduction.

We measured the relative decrease in fundamental frequency dF0 and the relative decrease in intensity dInt. Fig. 3 shows the effect of the variation of the production parameters a_0, q_0 and p_0 on the acoustic parameters dF0 and dInt. (1) When the adduction becomes stronger (a_0 higher negative values) dInt decreases but dF0 increases. (2) Lowering of q_0 leads to a decrease in dF0 but to an increase in dInt. (3) Lowering of p_0 leads to a decrease in dInt while dF0 remains relatively unchanged.
So during strong adduction it is necessary to lower the cord tension and the lung pressure to get the acoustic features of the "sharpening".

5. DISCUSSION

Resynthesis of "sharpenend" syllables by glottal abduction or adduction was done. In the latter case, mechanical compression ($a_{abd} < 0$), low cord tension and a decrease in lung pressure are necessary. Similar physiological features were found for the Danish "stød" [7]. The production mechanisms are those of "creaky voice" phonation. Modelling "sharpening" by glottal abduction seems to be the less promising way since at the end of the vowel there is a relatively high air flow through glottis and vocal tract which may produce aspiration. But both production mechanisms generate the main acoustic cues characteristic of "sharpening".

6. REFERENCES

[1] HEIKE,G. (1962), "Suprasegmentale Merkmale der Stadtkölner Mundart. Ein Beitrag zur 'Rheinischen Schärfung'", *Phonetica 8*,147-165.
[2] HEIKE,G. (1988), *"Zur wortunterscheidenden Funktion der rheinischen Schärfung"*, Deutscher Wortschatz, 677-686, Walter de Gruyter: Berlin, New York.
[3] KRÖGER,B.J. (1989), *"Die Synthese der weiblichen Stimme unter besonderer Berücksichtigung der Phonation"*, Dissertation, Köln.
[4] HEIKE,G., GREISBACH,R., HILGER,S., KRÖGER,B.J. (1989), *"Speech synthesis by acoustic control"*, Eurospeech 89, Vol.2, 20-22, European Conference on Speech Communication and Technoloy, Paris.
[5] KRÖGER,B.J., OPGENRHEIN,C. (1990), "Das physiologisch akustische Modell AKUSYN des artikulatorischen Sprachsynthesesystems ASSCO", *IPKöln-Berichte 16*, 69-117.
[6] ISHIZAKA, K., FLANAGAN, J.L. (1972), "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Technical Journal 51*, 1233-1268.
[7] FISCHER-JØRGENSEN, E. (1989), "Phonetic analysis of the Stød in Standard Danish", *Phonetica 46*, 1-59.

DYNAMIQUE LINGUALE DES VOYELLES ORALES FRANÇAISES EVALUATION STATISTIQUE A PARTIR DE DONNEES CINERADIOGRAPHIQUES

Bernard FLAMENT

Institut Universitaire de Technologie - Heinlex - B.P. 420
44606 SAINT-NAZAIRE CEDEX - France.

ABSTRACT

In this present study we propose, through cinerediographic data, to extract by close evaluation the dynamics of the lingual articulator for French oral vowels in two sorts of actions : reinforced and unaccentuated.We intend to apply a statistical approach to lingual movements, at the place of articulation, using dispersion coefficients and confidence intervals. The degree of lingual articulator stabilization on the axe where maximum vocal tract constriction occurs is more important in the instances of reinforcement ; this is even more distinct if the length of vocal phonem is considered.

1-INTRODUCTION.LES TERMES DE LA PROBLEMATIQUE.

1.1 Le renforcement, qui peut affecter une partie de la chaîne parlée, tend à modifier de façon plus ou moins sensible le comportement articulatoire et par là les modalités acoustiques des réalisations phonématiques. Sur un plan spécifiquement articulatoire, les mouvements vélaires lors de la production des voyelles nasales renforcées sont notamment bien individualisés, l'abaissement du voile étant limité au maximum à la durée phonématique (certains cas de relèvement anticipé sont d'ailleurs relevés). En réalisation non-renforcée, des phénomènes d'extension ont lieu : le passage vélo-pharyngal non seulement présente un diamètre plus important, mais il peut se produire d'une manière à la fois récurrente et subséquente (B.FLAMENT I4I, I5I) introduisant bien entendu des modifications acoustiques des phonèmes contigus. Le positionnement vélaire y est en outre plus fluctuant, plus mobile alors qu'en réalisation renforcée, la stabilisation est plus marquée, le voile se maintenant en position plus proche de la paroi pharyngale sous l'influence d'une tension musculaire plus grande.

1.2 L'articulateur lingual semble bien soumis à ce même type de phénomène comportemental: dans un précédent travail (B.FLAMENT et A.-H.SOUBRA I6I), a été mise en évidence la plus grande stabilité des valeurs du diamètre de la constriction pour les voyelles nasales renforcées sur l'axe où se produit le rétrécissement maximal du conduit vocal. La tension neuro-musculaire qui se concentre dans la zone linguale participant à la définition du lieu d'articulation maintient la langue en position plus proche du palais, du voile ou de la paroi pharyngale suivant l'articulation et ce, de façon beaucoup plus stable que lors de la réalisation non-renforcée des mêmes phonèmes. Comme pour ce travail, nous évaluerons ici la dynamique de l'articulateur lingual, relative cette fois aux voyelles orales du français, en appliquant une approche statistique permettant de quantifier le degré de stabilisation de cet articulateur lors des deux types de réalisations phonématiques : renforcées et non-renforcées.

2- LES MODALITES D'EXPERIMENTATION.

2.1 Elles sont similaires à celles du travail déjà mentionné (I6I) : les voyelles, ici orales, du français dans leur quasi totalité, soit IaI IɛI IeI IɔI IoI IøI IiI IyI IuI sont insérées dans des monosyllabes de type CV, eux-mêmes placés dans de courts énoncés (5-7 syllabes) de manière à ce que les phénomènes d'accentuation soient limités numériquement. Ces monosyllabes sont tantôt soumis à une valorisation au sein d'une structuration du type "c'est ... qui / qu(e)" placée en position forte en début d'énonciation, tantôt ils sont inaccentués dans des énoncés qualifiables de neutres, sans procédé d'insistance.

2.2 Les prises de vues cinéradiographiques ont été effectuées au Centre Médico-chirurgical de STRASBOURG-SCHILTIGHEIM à la vitesse de 50 im./sec., ce qui permet une appréhension très fine des faits articulatoires. Le locuteur est un francophone -langue maternelle ; son français est dénué de toute trace de nuance régionale.

2.3 L'ensemble du contour lingual a été envisagé pour une juste définition du lieu d'articulation de la voyelle considérée. Des mesures précises quant au diamètre du passage buccal aussi bien dans la zone alvéolaire que dans les zones palatale, vélaire ou pharyngale ont été relevées. Le dynamisme de l'articulateur est étudié sur l'axe où se produit le rétrécissement maximal et ce, image par image, pour l'ensemble de la durée articulatoire.

3- TRAITEMENT STATISTIQUE DES DONNEES.

3.1 A partir des valeurs du diamètre du tractus relevées sur l'axe du lieu d'articulation, nous proposons un traitement statistique. Prenons le cas en effet de deux réalisations phonématiques bien contrastées sur le plan des données concernant la durée articulatoire : le IeI, le IøI et le IuI. En énonciation valorisée (E.V.), la durée phonématique est beaucoup plus accusée qu'en énoncé neutre (E.N.) dans lequel la voyelle est en position inaccentuée. A titre d'exemples, voici, pour ces 3 voyelles et dans les 2 types de réalisations, les valeurs (en cm) du diamètre du passage vocal sur l'axe retenu :

IeI E.N. 0,7 / 0,9 / 1 / 1,25
E.V. 0,7 / 0,75 / 0,75 / 0,75 / 0,75
0,8 / 0,85 / 0,85 / 0,9

IøI E.N. 1 / 1,15 / 1,25 / 1,3
E.V. 0,8 / 0,85 / 0,9 / 0,9 / 0,75 /
0,75 / 0,75 / 0,75

IuI E.N. 0,4 / 0,4 / 0,7 / 0,9 / 1
E.V. 0,5 / 0,5 / 0,5 / 0,5 / 0,6 / 0,8

Nous observons non seulement une grande disparité sur le plan de la durée phonématique (chaque valeur correspond à la durée d'une image radiologique), mais aussi une variance positionnelle plus importante dans le cas des réalisations en E.N. La langue se maintient peu dans la position-cible : une mouvance de cet articulateur intervient de façon plus marquée qu'en réalisation renforcée (E.V.) où une stabilisation se produit.

3.2 Evaluation des écarts-types.

3.2.1 L'écart-type d'une série de mesures comportant N nombres $x_1, x_2, \ldots x_N$ est noté par s et se définit par la formule suivante :

$$s = \sqrt{\frac{\sum_{j=1}^{N} (x_j - \bar{x})^2}{N}}$$

où $\bar{x}$ représente la valeur moyenne des nombres de la succession désignée par x.

3.2.2 Voici les valeurs de s pour les différentes voyelles orales considérées, en prenant en compte la totalité de la durée de l'articulation :

	N	s			
IaIE.N.7	0,0639	+			
E.V.8	0,0415	-		N	s
IɛIE.N.5	0,0374	+	IɛIE.N.4	0,1980	+
E.V.8	0,0348	-	E.V.9	0,0614	-
IɔIE.N.8	0,1871	+	IoIE.N.7	0,2279	+
E.V.9	0, 0774	-	E.V.8	0,1225	-
IøIE.N.4	0,1146	+			
E.V.8	0,0634	-			
IiIE.N.5	0,0245	+	IyIE.N.3	0,0471	+
E.V.5	0	-	E.V.3	0	-
IuIE.N.5	0,2481	+			
E.V.6	0,1106	-			

La dispersion des valeurs en E.V. par rapport à E.N. présente un déficit dans la totalité des cas : la stabilité articulatoire -linguale- y est plus importante, au lieu d'articulation, dans ce type de réalisation. Là où les valeurs de s sont déjà très faibles en E.N. (c'est le cas essentiellement des voyelles antérieures, de lieu d'articulation alvéolaire), celles-ci s'abaissent encore en E.V. Dans les autres cas, les écarts entre E.N. et E.V. sont parfois très sensibles avec des rapports (bilan E.N./E.V.) pouvant atteindre 1 à 2 et même quasiment 1 à 3. L'incidence des muscles linguaux n'est pas négligeable dans ce processus, notamment pour les réalisations postérieures :les muscles qua-

lifiés d'extrinsèques, en l'occurrence le palatoglosse, le styloglosse et l'hyoglosse modifient la position et la configuration de la langue dans la cavité buccale, de la zone palatale à la région pharyngale inférieure (sur l'activité linguo-musculaire, v. entre autres les travaux de W.-J.HARDCASTLE I7I, S.WOOD I9I, J.-P.ZERLING I10I, M.ROSSI I8I, pp.97-99).Les articulations postérieures IuI, IoI se caractérisent toutefois par une dispersion des valeurs plus accusée, imputable sans doute au fait que la masse linguale d'arrière est plus importante, et rend plus lent, plus aléatoire, le positionnement lingual au lieu d'articulation ; ceci est tout particulièrement sensible en E.N. où les phénomènes de coarticulation sont conséquents et où la stabilité y est donc moindre. En E.V., le diamètre du tractus diminue et la position-cible est davantage maintenue sous l'influence de la tension neuro-musculaire.

3.2.3 Les valeurs de s sont encore bien plus réduites pour les réalisations renforcées si l'on prend comme base temporelle la durée articulatoire des mêmes phonèmes , observée en E.N. :

IaI (N=7)	s = 0,0416		
IɛI (N=5)	s = 0,02	IeI (N=4)	s =0
IɔI (N=8)	s = 0,0415	IoI (N=7)	
IøI (N=4)	s = 0		s=0,0942
Ii I (N=5)	s = 0	IyI (N=3)	s=0
IuI (N=5)	s =0,04		

Une remarquable stabilité est observable pour les réalisations renforcées. La régulation du paramètre de la durée diminue le taux des éventualités de dispersion concernant les valeurs de $x_j - \bar{x}$, ce qui réduit encore la valeur de s . Il n'y a guère que pour IaI où celle-ci est quasiment la même, du fait d'une très faible augmentation de durée en E.V. (N=8) par rapport à E.N. (N=7) et pour IiI où sa valeur (nulle) reste inchangée.

3.3 Intervalles de confiance.

Afin de nous assurer de la validité des résultats, nous avons calculé l'intervalle de confiance pour chacun des phonèmes vocaliques (avec homogénéisation temporelle de la durée articulatoire E.V./ E.N.). Cet intervalle de confiance a été calculé en considérant un degré de confiance égal à 95% pour les différentes séries.

IaI	E.N.	N=7	0,0445	< σ	< 0,152
	E.V.	N=7	0,0290	< σ	< 0,0991
IeI	E.N.	N=5	0,0251	< σ	< 0,1202
	E.V.	N=5	0,0134	< σ	< 0,0643
IɛI	E.N.	N=4	0,1295	< σ	< 0,8526
	E.V.	N=4	0	< σ	< 0
IɔI	E.N.	N=8	0,1322	< σ	< 0,4071
	E.V.	N=8	0,0293	< σ	< 0,0902
IoI	E.N.	N=7	0,1586	< σ	< 0,5421
	E.V.	N=7	0,0656	< σ	< 0,2241
IøI	E.N.	N=4	0,0749	< σ	< 0,4932
	E.V.	N=4	0	< σ	< 0
IiI	E.N.	N=5	0,0164	< σ	< 0,0787
	E.V.	N=5	0	< σ	< 0
IyI	E.N.	N=3	0,1663	< σ	< 0,7974
	E.V.	N=3	0	< σ	< 0
IuI	E.N.	N=5	0,0301	< σ	< 0,3628
	E.V.	N=5	0,0268	< σ	< 0,1285

Les intervalles de confiance sont toujours compris entre 2 valeurs plus faibles -parfois de façon très nette - lorsque l'articulation considérée est en E.V. On observe même dans un certain nombre de cas des valeurs nulles, traduisant une constante stabilité articulatoire dans la zone d'articulation et ce, pour la durée concernée, dans ce type de réalisation.

4- RESULTATS.

1) Les valeurs de s en E.N. s'avèrent toujours supérieures en E.V. Ceci marque la relation directe entre renforcement articulatoire et stabilité -ou accroissement de la stabilité- de l'articulateur lingual au lieu d'articulation. La tension neuro-musculaire plus importante qui accompagne la réalisation de phonèmes vocaliques renforcés permet d'une part un positionnement lingual plus rapide dans la position-cible caractéristique de l'articulation ; d'autre part, elle favorise le maintien de cet articulateur dans cette position, provoquant ainsi une stabilisation linguale à l'endroit où s'effectue le rétrécissement maximal du conduit vocal.

2) Les voyelles de constriction nettement antérieure telles IiI IyI IɛI présentent des valeurs de s faibles en E.N. et encore plus réduites en E.V. (voire nulles). Ceci peut être imputable essentiellement sans aucun doute à la durée plus faible pour ce type de voyelles, donc à de moindres possibilités de dispersion des valeurs ; une autre raison découle de l'action du maxillaire inférieur : bien qu'il existe une complexité certaine des rapports entre les mouvements linguaux et ceux de la mandibule (A.BOTHOREL I1I, pp.119-120), l'influence stabilisatrice de cet articulateur apparaît le plus opérant s'agissant d'articulations vocaliques "réalisées par un mouvement de la partie antérieure de la langue" (A.BOTHOREL I2I, p.70). Pour ce qui est des voyelles de réalisation postérieure IɔI IoI IuI, le renforcement entraîne des réductions très marquées des valeurs de s (v. notre essai d'explication, §3.2.2). A l'instar de la postériorisation des articulations, la labialité intervient au détriment de la stabilité linguale, IoI et IuI présentant des valeurs de s plus élevées en E.N. par rapport à IɔI, et restant supérieures en E.V. pour IoI. L'adjonction, au plan distinctif, du trait de la labialité, perturbe le degré de stabilisation articulatoire (ceci va tout à fait dans le sens des observations d'A.BOTHOREL I2I, p.68).

3) La durée articulatoire n'est pas à considérer d'une façon absolue. Dans les cas de renforcement, la stabilisation linguale est en effet plus marquée alors que la durée articulatoire augmente très généralement ; encore convient-il de confronter les mêmes phonèmes dans les 2 types de réalisations -renforcées et non-renforcées. Ceci est à mettre en relation avec le débit d'élocution, plus rapide dans les énonciations neutres. En raison de l'accroissement de ce débit, la précision des gestes articulatoires est moins grande et la "cible" moins souvent atteinte (J.CAELEN I3I, p.129). Néanmoins, une brièveté phonématique n'implique pas nécessairement une plus grande instabilité articulatoire : les cas du IiI et du IyI sont plus que probants ; leur brièveté de réalisation -par nature- est largement compensée par l'augmentation de l'influence stabilisatrice du maxillaire inférieur. En fait, l'analyse des faits articulatoires est complexe : l'explication de ces faits passe bien souvent -sinon de façon inéluctable- par la prise en compte des imbrications d'articulateurs et de comportements musculaires.

4 - REFERENCES

I1I A.BOTHOREL (1983), "Contraintes physiologiques et indices articulatoires", *Speech Communication 2* , n°s 2-3, pp.119-122.

I2I A.BOTHOREL (1984), "Apport de la radiocinématographie à la recherche phonétique", *Etudes de Phonologie, Phonétique et Linguistique descriptive du français*, Buske Verlag, Hambourg, vol.1, pp.55-88.

I3I J.CAELEN (1985), "Introduction à la segmentation cinématique", *Actes des 14èmes J.E.P.*, G.A.L.F. et E.N.S.T., Paris, pp.129-132.

I4I B.FLAMENT(1988),"Positionnement vélaire en français sous l'effet de la tournure valorisante : présentatif+relative -Approche articulatoire sur les plans phonématique et interphonématique", *Actes du 7ème Symposium F.A.S.E.*, Edimbourg, vol.3, pp.875-882.

I5I B.FLAMENT (1989), "Traitement articulatoire -lingual et vélaire- des voyelles nasales en français sous l'effet de la valorisation", *Mélanges de phonétique générale et expérimentale offerts à P.SIMON*, Strasbourg, vol.1, pp.371-391.

I6I B.FLAMENT et A.-H.SOUBRA (1990),"Approche statistique de la dynamique linguale en français -Application aux voyelles nasales", *Actes du 1er Congrès français d'Acoustique*, I.C.P.I., Lyon, vol.I, pp.491-494.

I7I W.-J.HARDCASTLE (1976), *Physiology of Speech Production*, Academic Press, London / New York / San Francisco.

I8I M.ROSSI (1983), "Niveaux de l'analyse phonétique : nature et structuration des indices et des traits", *Speech Communication 2* , n°s 2-3, pp.91-106.

I9I S.WOOD (1977), "A radiographic analysis of constriction locations for vowels", *Working Papers 15*, Phonetics Laboratory, Lund University, pp.101-131.

I10I J.-P.ZERLING (1979), "Description de cinq voyelles orales du français en contexte et nouvelle classification articulatoire", *Verbum* , Nancy-II, 1, pp.55-87.

A FEEDFORWARD CONTROL STRATEGY CAN SUFFICE FOR ARTICULATORY COMPENSATIONS

Shinji Maeda

Département SIGNAL, CNRS URA-820
Ecole Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT

An articulatory model was used to analyze cineradiographic and labiofilm data. The variation in "target" values of two model parameters, the jaw and tongue-dorsum positions, during the production of the vowels, /i/ and /a/, was examined. The "target" values of these two parameters for the same vowel vary much more than the corresponding acoustic ones. The scattergram of each vowel exhibited a linear relationships which can be regarded as an indication of the coordination between the jaw and tongue. When the coordination effects are subtracted, the articulatory variability becomes comparable to that of the acoustic (F1/F2) one. Calculations with the model indicated that the coordination is used by speakers to achieve an acoustic compensation. These findings suggest that vowel production is compensatory and that compensation can be modelled effectively by a feedforward strategy.

1. INTRODUCTION

Bite-block vowel experiments have demonstrated a speaker's ability to compensate for the effects of blocked jaw position by readjusting the other articulators to produce specified vowels. Observing a speaker's ability to compensate immediately, Lindblom, Lubker and Gay have suggested that normal speech production itself is compensatory [3]. If this is the case, we should observe in normal speech a high degree of variability in the individual articulatory positions and a lower degree in the corresponding acoustic patterns, for example, in the formant patterns. Moreover, if compensation occurs in an arbitrary manner, it is not effective to specify vowel targets in terms of articulatory parameters. This appears to be one of reasons why the targets are often described by the vocal tract area function [2]. If compensation occurs in a lawful manner however, the vowel targets can be specified directly by the individual parameters with some calculations reflecting the laws. We shall investigate these questions by analyzing X-ray and labiofilm data with an articulatory model.

2. ARTICULATORY DATA AND MODEL

The data consist of more than 1000 digitized tracings of vocal tract shapes corresponding to 10 French sentences uttered by two female speakers, PB and DF [1]. Each of the data frames describing the vocal tract profiles from the glottis to the lip opening and the frontal lip shapes was obtained by manually tracing radiofilms and labiofilms shot simultaneously at a rate of 50 frames per second. The digitized version of the data has been kindly provided by the Phonetic Institute of Strasbourg, France.

The measured vocal tract shapes were analyzed statistically. A factor analysis has resulted in a linear articulatory model with seven parameters. In this study, we shall focus our attention on two parameters, the jaw and tongue-dorsum positions for two reasons: these two parameters are most important for specifying the tongue profiles and they can acoustically compensate for each other, specifically in the production of unrounded vowels, such as /i/, /e/, and /a/ [4].

3. ARTICULATORY VARIABILITY

With the linear model, the value of each parameter is calculated directly from the measured vocal tract shape. The articulation along a sentence can be described,

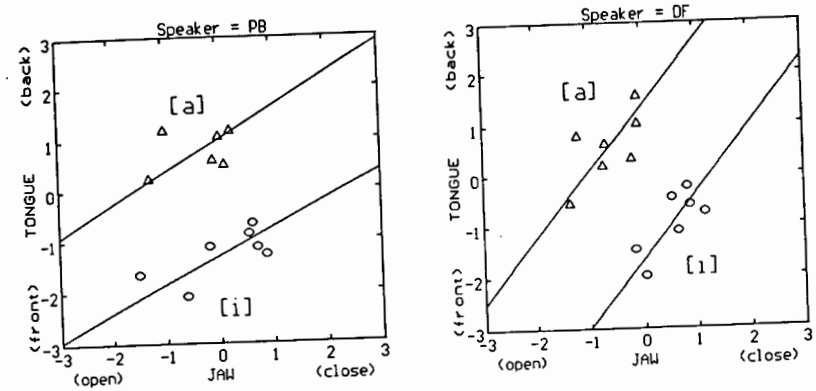


Fig.1 Scattergrams of jaw and tongue-dorsum parameters at the "articulatory targets" for the two vowels /i/ (indicated by the circles) and /a/ (by the triangles). The ordinate and abscissa have standardized units. Zero corresponds to the arithmetic mean calculated for all the utterances by each speaker. 1 (-1), 2 (-2), and 3 (-3) represent 1, 2, and 3 standard deviations, respectively, from the mean. Data for the two speakers, PB and DF are shown.

therefore, by the frame-by-frame variation of the calculated articulatory parameter values. The resultant data have indicated a considerable articulatory variability for the same vowel from different phonetic contexts. In order to assess the range of variability, trajectories of the two parameters, jaw and tongue dorsum, had been plotted on the jaw-tongue articulatory space. Then an articulatory "target" position was determined as the turning point on each trajectory. The result is shown in Fig.1.

The straight lines plotted on Fig.1 were determined by means of a principal component analysis of the scattergrams associated with each of the two vowels: they correspond to the first principal axis. Although the scattergrams exhibit a great degree of variations, the data points for /i/ and /a/ are distributed without overlap. Furthermore, each cluster is distributed roughly along the straight line. These straight lines can be regarded as linear approximations of the inter-articulatory coordination between jaw and tongue-dorsum. The observed variability, therefore, can be separated into a controlled context-determined variation and an unexplained residual, say, "true" variability. Since the proportion of the variance extracted by the first principal component varies between 65% (in the case of [a] uttered by speaker PB) and 88% ([i] by speaker DF), the true articulatory variability for jaw and tongue ranges from 35% to as small as 12% of the observed variance.

4. ACOUSTIC VARIABILITIES

The articulatory variability can be examined more meaningfully, if it is compared with the corresponding acoustic variability. In this study, the first (F1) and second formant (F2) frequencies, as the acoustic characteristics of the two vowels, were calculated using the articulatory model. The F1-F2 calculations were done only for speaker PB, since the data for DF lacks the lip section and thus F1 and F2 cannot be calculated. All seven parameter values were derived from the corresponding data frame. The area function and then formant frequencies were computed from model specified vocal tract shapes. The resultant F1/F2 plots are shown in Fig.2. The data points for the vowel /u/ are added to indicate the vowel space of speaker PB.

Comparing the articulatory target scattergrams in Fig.1 (for speaker PB) and the corresponding acoustic ones in Fig.2, it appears that the acoustic scattergram points are distributed more tightly than

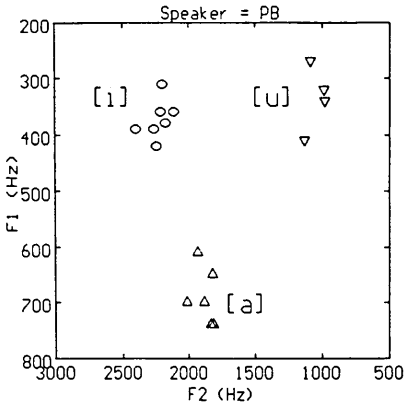


Fig.2 The first (F1) and second (F2) formant scattergrams corresponding to the articulatory target scattergrams shown in Fig.1 (for speaker PB). The scattergram for the vowel /u/ is also plotted to indicate the speaker's vowel space.

articulatory ones, *i.e.*, acoustic variability seems to be less than articulatory one. For a quantitative comparison, let us propose a variability index, υ (an averaged normalized variance), for two articulatory or two acoustic variables as follows:

$$\upsilon = 100 \times \sqrt{\frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_{1_{max}}^2} + \frac{\sigma_2^2}{\sigma_{2_{max}}^2}\right)} \quad (\%)$$

where σ_i^2 is the variance of variable i (= 1 or 2 in our case), and $\sigma_{i_{max}}^2$ is the possible maximum variance of variable i. Since a sufficient amount of data to determine the possible maximum variance is not available, we have assumed, as a gross approximation, that $\sigma_{i_{max}}$, of articulatory and acoustic data can be substituted by the values of half of the range of the individual variables. In the calculation of the articulatory variability index, $\sigma_{i_{max}} = 3$ is used for both jaw and tongue-dorsum data, corresponding to half the range, since parameter values rarely exceed the range from -3.0 to 3.0. The acoustic variability index is computed assuming that $\sigma_{1_{max}}$ (for F1) equals to 300 Hz, and $\sigma_{2_{max}}$ (for F2) to 1250 Hz. The calculated index values are listed in Table 1.

Table 1 Articulatory and acoustic variability indices (in %) for the two speakers

	PB		DF	
	/i/	/a/	/i/	/a/
Jaw/Tongue				
υ	21.7	16.2	17.0	18.4
$\upsilon_{residual}$	7.5	8.1	3.6	4.6
F1/F2				
υ	7.9	8.9	---	---

The index values span around 20% for the articulation and less than 10% for the acoustics. The residual articulatory variability indices are listed at the rows marked "$\upsilon_{residual}$" in Table 1, which are calculated from the proportion of variance corresponding to the residual. These index values are less than 10%, a value which is less than half of the corresponding total raw variability, and which compares well with the index calculated for the F1/F2 scattergrams of PB shown in Fig.2. For speaker DF, the true articulatory variability is four times less than the observed raw variability. The calculation have indicated that although the variability of the individual articulators is relatively great, if the coordination term is subtracted, the articulatory variability compares well with the acoustic one.

5. COMPENSATORY ARTICULATION

What mechanism lies behind this significant reduction of the variability from articulatory to acoustic by means of coordination? In our previous studies [4], we have already shown that in case of unrounded vowels such as /i/ and /a/, jaw and tongue-dorsum positions can acoustically compensate for each other, as mentioned earlier. The compensation means that a deviation in the position of one articulator can be compensated by a readjustment of other articulator(s) to keep the deviation in the acoustic pattern to a minimum. It is reasonable, then, to hypothesize that the inter-articulatory coordination, in fact, results in the acoustic compensation of the type just described above. If this is the case, the principal axis representing the coordination in Fig.1, is also an acoustical "equi-line", *i.e.*, changes in the values of the two parameters along these lines result in relatively invariant acoustic patterns that depend only on the vowel identity.

In order to demonstrate the acoustic equivalence for the two vowels, F1 and F2 values were calculated at different jaw positions from -2.0 (low) to 1.0 (high) for /i/ and from -3.0 to 0.0 for /a/, with 1.0 step size. The corresponding tongue positions were determined by their linear relationships. Note that a change in jaw position influences not only the tongue shape, but also the lip aperture and, to some extent, the larynx position. The values of the remaining five parameters were kept fixed at those originally determined from the corresponding vocal tract data frame. The results are listed in Table 2. The index related to the equi-line of /a/ is 3.2%, which is much smaller than observed acoustic variability. As far as the vowel /i/ is concerned, the index becomes extremely small, about 1%, indicating that the equi-line produces an almost invariant F1-F2 pattern.

Table 2 F1/F2 variability indices calculated along the equi-lines of PB in Fig.1.

	range	index
/i/	-2.0 ⇔ 1.0	1.1 %
/a/	-3.0 ⇔ 0.0	3.2 %

Although the acoustic compensation along the equi-lines is not perfect, it is safe to state that articulatory manoeuvres along an equi-line tend to result in fairly invariant acoustic patterns around the target vowel. It should be emphasized here that the equi-lines are derived from the observation of data. It is tempting to speculate then that the speakers have integrated these equi-lines in their mental process and exploit them to place individual articulator positions differently but appropriately for particular phonetic contexts, yet producing relatively invariant acoustic targets.

It may be noteworthy to mention that the coordination does not necessarily always means compensation. In the case of /u/ for example, the raw variabilities of the jaw and lip parameters (height and protrusion) were relatively small, less than 10%. In detail however, scattergrams indicated that closing the jaw, and narrowing and protruding of the lip opening occur concomitantly, enhancing together a narrow and long lip tube. The acoustic consequences of this kind of coordination would be exactly in the opposite of compensation.

6. CONCLUDING REMARKS

It has become clear that the apparently large variability of the individual articulator positions during the same vowel but from different contexts can be explained, at least in part, by the inter-articulator coordination. Moreover, the coordination is such as to achieve an acoustic compensation which results in the realization of a relatively invariant acoustic target, thus supporting the idea of speech production as a compensatory process [3]. Surprisingly, the coordination and thus the compensation can be specified directly in terms of articulatory parameters. The implication of this is important. If the relationship is well defined in such a simple fashion, it is not unreasonable to assume that speakers know exactly how to coordinate in advance. Then a feedforward control mechanism can be assumed for the compensatory articulation, without resorting to acoustic or to sensory feedback.

7. REFERENCES

[1] Bothorel, A., Simon, P., Wioland, F. and Zerling, J-P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français*. Travaux de l'Institut de Phonétique de Strasbourg.

[2] Gay, T., Lindblom, B. and Lubker J. (1981). "Production of bite-block vowels: Acoustic equivalence by selective compensation", J. Acoust. Soc. Am., 69(3), 802-810.

[3] Lindblom, B., Lubker, J. and Gay, T. (1979). "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", J. Phonetics, 7, 147-161.

[4] Maeda, S. (1990). "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model". In *Speech Production and Speech Modeling* (W.J. Hardcastle & A. Marchal, editors), pp. 131-149. Kluwer Academic Publishers.

TOWARDS THE SPECTRAL CHARACTERISTICS OF FRICATIVE CONSONANTS

Christine H. Shadle[1], Pierre Badin[2], André Moulinier[1]

University of Southampton, UK[1]
ICP, INPG, Grenoble, France[2]

ABSTRACT

Articulatory data and the all-pole transfer functions for sustained fricative consonants [s, ʃ, ç] were used to identify the cavity affiliation of peaks and troughs in the far-field spectra. This identification then allowed an analysis of the differences between fricatives, and across subjects within fricatives, necessary steps towards the establishment of distinguishing acoustic cues.

1. INTRODUCTION

It has long been established that [f,θ] are distinguished chiefly by the transitions of the vowels on either side, while [s,ʃ] are distinguished by their spectral characteristics [3]. But establishing the particular spectral cues that distinguish [s, ʃ] from each other, or from other fricatives, is more difficult. Many authors report consistency within a speaker, but high variability across speakers [4,7]. Perhaps as a result, efforts to phrase distinguishing cues in terms of the frequency range of the highest intensity levels, or in terms of relative intensity levels, seem to work well within a speaker but poorly across speakers (e.g. the frequency ranges overlap so much as to be useless.) [7].

In this study we explore variability in the spectra of sustained fricatives. First, we need to establish which aspects of the spectrum are consistent within a speaker-fricative combination. Then where possible we identify the articulatory parameters that control these consistent features of the spectrum. Finally we use this articulatory-acoustic mapping to explain some of the across-subject differences. This sequence should lead to a set of paired articulatory-acoustic cues that can then be tested for their perceptual importance.

2. METHOD

2.1 Corpus and Speakers

The corpus used in this paper is the result of a larger study (Leeds, Grenoble, Southampton). It includes articulatory, aerodynamic and acoustic measurements made of two speakers. The corpus includes 13 fricatives [f,v,θ,ð,s,z,ʃ,ʒ,ç,j,x,ɣ,h] produced in several ways. This study refers only to the sustained corpus, in which the set of 13 fricatives was said six times; in each set, the order of the 13 fricatives was randomized. Two different recordings of the sustained fricatives were used in this study, as detailed below.

The two speakers used for the corpus are the first two authors of this paper, and will be referred to as CS, a woman speaker of General American English, and PB, a man speaker of French. Although the list of fricatives recorded includes several that are not native to either speaker, these were included deliberately to obtain further examples of place variation for the same vocal tracts.

In addition to measurements made while speaking, X-ray data and dental impressions were available for each subject. Together with EPG data and external photographs, these were used to construct an area function for each unvoiced fricative for each speaker [6].

2.2 Acoustic Analysis

Data shown in this paper were recorded under high-fidelity conditions: the subject was seated in a chamber anechoic above 170 Hz, with a B&K 4165 ½" microphone located 1m in front of the subject's mouth. Recordings were made with a Sony PCM system at 16 bits with a sampling frequency of 44.1 kHz. A calibration signal was recorded to allow absolute sound pressure level to be retained.

An average power spectral density function was computed by averaging 25 spectra in the center of the 3s fricative. Each spectrum was computed using a 20ms Hanning-window.

2.3 Determination of Transfer Function

In this experiment, the subject assumed the position for a fricative, but without actual speech production (glottis held closed). The vocal tract was excited by a small loudspeaker fed with white noise and pressed against the neck just above the thyroid cartilage. A microphone located 2cm from the mouth detected the (very weak) noise signal after filtering by the vocal tract. This signal was essentially the all-pole transfer function of the tract, up to about 5 kHz.

The area functions derived from articulatory data were then used to predict the all-pole transfer function for each fricative. Comparison of predicted and measured all-pole functions then enabled identification of the cavity affiliation of each pole. Further details are given in [2].

3. RESULTS AND DISCUSSION

Figure 1 shows three of the fricatives analyzed, with all six tokens shown on each graph. Note first the consistency apparent within each graph, i.e. within each fricative-subject combination. This consistency makes it easier to evaluate the variability across speakers, and across fricatives. For [s,ç] the overall spectral shapes are similar but the frequencies at which particular peaks occur differ between the two speakers. For [ʃ], even the overall shape differs: both speakers have a region of high energy, between 1.5 and 6 kHz for PB, and 2.5 to 7 kHz for CS. However, for PB there is an abrupt drop in amplitude of some 10 dB at 6 kHz and the spectrum is approximately level above that frequency; for CS, there is no abrupt drop. Instead the level falls off steadily, decreasing 20 dB between 7 and 12 kHz. Can we make sense of these differences?

Badin's results [2] indicate that for CS's [ʃ], F1 is a Helmholtz resonance of back cavity and constriction; F2 and F3 are back cavity resonances; and F4 is a front cavity resonance. A series pressure source in the front cavity would result in zeros cancelling the back cavity resonances, plus two free zeros: one corresponds to a Helmholtz resonance of the constriction and the part of the front cavity between the constriction exit and the source. The other corresponds to the half-wavelength resonance of the same part of the front cavity.

Since CS has smaller vocal tract dimensions, her formant frequencies are predicted to be higher, and in fact they are. However, a much more obvious difference is that for CS the first four formants are approximately evenly spaced, while PB has F2, F3 and F4 clustered together. With the zeros interspersed, these small differences in formant frequency make a big difference in formant amplitude: for PB, the second formant is boosted and becomes the lowest high-amplitude peak, while for CS, F3 takes on that role. This means that the lower edge of the high-amplitude region differs by 1 kHz, even though F2 differs by only 100-200 Hz.

Above 5 kHz we have less information to work with. However, the differences in spectral amplitude and slope could be explained if the free zero were at a significantly lower frequency for PB than for CS, e.g. 7 and 12 kHz respectively. This zero frequency should be inversely proportional to l_0, the teeth-constriction distance, and in fact l_0 is significantly longer for PB, as evidenced from X-ray and direct palatography. This is surprising since

the vocal tract dimensions in the anterior part of the mouth cavity, obtained from measurements of the two subjects' dental impressions, are quite similar. Since the phoneme is native to each subject, and the spectral differences noted are consistent within each subject, more subjects are needed to establish why the articulatory differences exist.

The fricative [s] is more similar for the two subjects. Since the front cavity is smaller than for [ʃ], the corresponding resonances are higher. For CS, it appears from transfer function simulations that the lowest front cavity resonance is F6 (see Fig. 1); F2, F3, F4 and F5 are the lowest resonances of the back cavity (harmonics of the half-wavelength mode), and are accompanied by bound zeros. For PB the lowest front cavity resonance is F5, and F2, F3, and F4 are the back cavity resonances [1]. The differences between these resonances are consistent with the articulatory data. The amplitude of the plateau above the front cavity resonance relative to the spectral level of this resonance varies noticeably between the two subjects, and again the free zero may be lower for PB (approximately 11.5 kHz) than for CS (well above 12 kHz).

The fricative [ç] is not native to either speaker, and so might be expected to be more variable. In fact, it looks consistent for each speaker, and the overall spectral shape is similar. For both speakers, the lowest front-cavity resonance is the lowest high-amplitude formant. This corresponds to F4 for CS, F3 for PB. Although the front cavity is longer for [ç] than for [ʃ], this front-cavity resonance is not significantly lower. A possible explanation is that for extremely short front cavities, the resonance frequency is related to the volume or possibly vertical dimension. Thus the exact shape of the sublingual cavity becomes important for [s,ʃ]. As for [ʃ], the spectral shape at high frequencies differs, and could be explained in part by a difference in source-constriction distance. The likelihood that the source is distributed [5] complicates the issue by blurring the free zero, but in any case a lower-frequency free zero would reduce the overall amplitude relative to [ʃ].

5. CONCLUSION

The search for acoustic cues distinguishing fricative consonants must begin with a study of the variability present in fricative production. By using subjects for whom much articulatory data is available, it has been possible to locate low-amplitude but consistent spectral peaks, and to discover their cavity affiliation and controlling parameters. Although vocal tract dimensions influence peak frequencies, the added complications introduced by zeros mean that simple measures such as frequency range for high-amplitude regions are likely to be highly variable.

6. ACKNOWLEDGEMENT

This work was funded in part by a collaborative EC SCIENCE award, CEC-SCI*0147C(EDB).

7. REFERENCES

[1] BADIN, P. (1989) "Acoustics of voiceless fricatives: production theory and data," STL-QPSR 3/1989, 33-55.

[2] BADIN, P. (1991) "Fricative consonants: acoustic and X-ray measurements", *J. Phonetics*, in press.

[3] HARRIS, K. (1958) "Cues for the discrimination of American English fricatives in spoken syllables", *Language and Speech* 1, 1-7.

[4] HUGHES, G.W. & HALLE, M. (1956) "Spectral properties of fricative consonants", *J. Acoust. Soc. Am.* 28:2, 303-310.

[5] SHADLE, C.H. (1990) "Articulatory-acoustic relationships in fricative consonants", in *Speech Production and Speech Modelling*, eds. W.J. Hardcastle and A. Marchal, Kluwer Acad. Pub., 187-209.

[6] SHADLE, C.H. (1991) "The effects of geometry on source mechanisms of fricative consonants", *J. Phonetics*, in press.

[7] STREVENS, P. (1960) "Spectra of fricative noise in human speech", *Language and Speech* 3, 32-49.

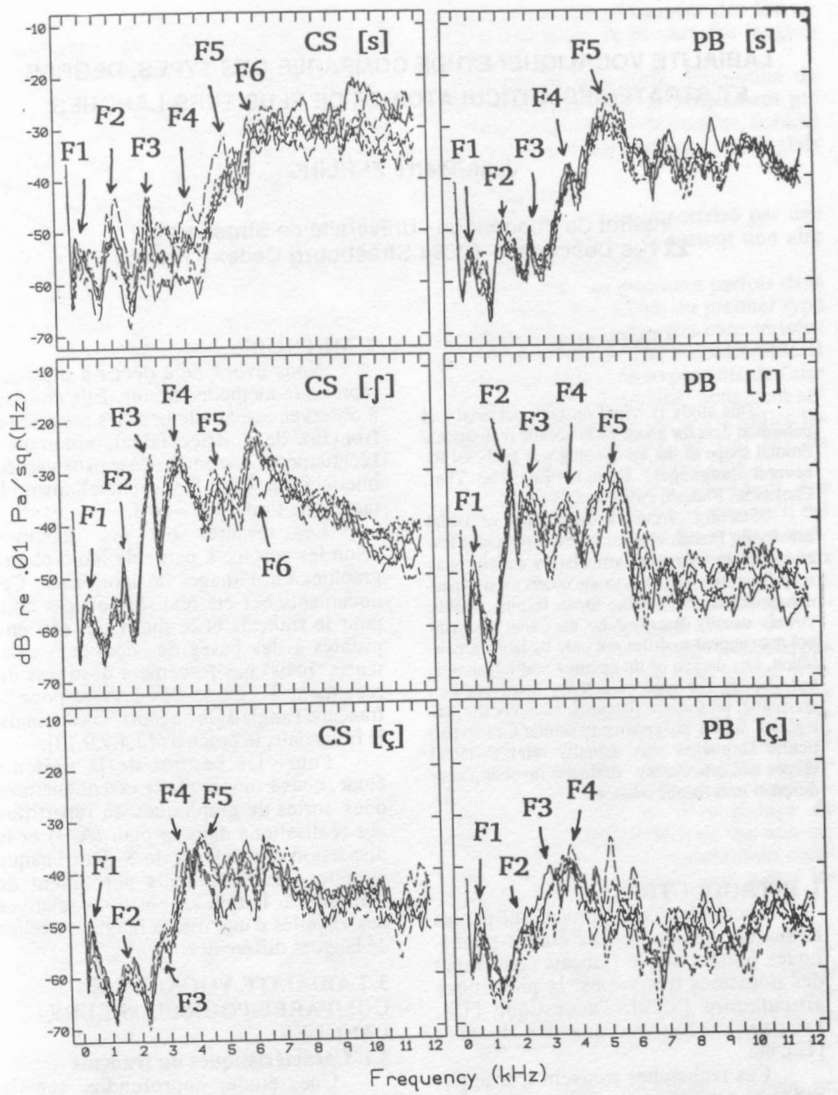


Fig. 1. Averaged power spectral densities for sustained productions of the fricatives [s, ʃ, ç]. In each graph the six curves shown correspond to the six tokens uttered by each subject. Subject PB is male, native French speaker; CS is female, native General American speaker.

LABIALITE VOCALIQUE: ETUDE COMPAREE DES TYPES, DEGRES ET STRATEGIES ARTICULATOIRES DE PLUSIEURS LANGUES

Jean-Pierre ZERLING

Institut de Phonétique - Université de Strasbourg II
22 rue Descartes - 67084 Strasbourg Cedex - France

ABSTRACT

This study is based on both personal and published data for about 2500 vocalic realisations. Frontal shape of the lip opening was analysed for several languages: French, English, Thai, Cantonese, Finnish, Polish and Swedish.

Starting from a description of labial activity for French vowels, we show through a few examples in what way articulatory activity may vary from one language to the other, even if phonological features are the same. In other words, vowels usually described by the same IPA symbol may appear to differ not only by labial articulation, i.e., degree of lip opening and labial area, but also by the type of activity involved, i.e., spreading, protrusion, flatening, etc. Vocalic categories which are apparently similar for two particular languages may actually refer to labial shapes and articulatory strategies involving very different intervocalic relations.

1. INTRODUCTION

On sait le rôle important joué par les lèvres en phonation. Les études phonétiques portant sur le français concernent des domaines très variés: la phonétique articulatoire [2,22], l'acoustique [15], l'acquisition automatique des données [12], etc.

Ces recherches trouvent leur application en phonétique et en phonologie [5], en modélisation et en synthèse de la parole[1,14], en perception visuelle [8,11,12,18], etc.

La présente étude vient se greffer à diverses autres que nous avons nous même menées, toujours à propos de la labialité, mais concernant en plus la coarticulation [19], les stratégies articulatoires [20,22], la variabilité [21] et les comparaisons interlangues [22,23].

2. METHODE

Nous avons déjà décrit à plusieurs reprises la méthode retenue. Elle consiste à observer essentiellement les paramètres frontaux de l'orifice labial, notamment l'écartement horizontal A, l'espace vertical interlabial B, l'aire S et éventuellement le facteur de forme K2 = A/B.

Les mesures ont été obtenues, selon les corpus, à partir de labio-photographies ou d'images de labiofilms. Ces documents ont été réalisés soit par nous pour le français et le thaï [22], soit empruntés à des bases de données existantes mais pas forcément destinées directement à l'étude des lèvres, pour le français, l'anglais, le suédois, le cantonais, le finlandais, le polonais [3,4,7,9,13].

Pour les besoins de la présente étude, nous avons exploité essentiellement deux sortes de graphiques: la répartition des réalisations dans le plan (A,B) et la dispersion des valeurs de S pour chaque voyelle. Ces documents permettent de comparer à la fois les positions relatives des voyelles d'une même langue et celles de langues différentes.

3. LABIALITE VOCALIQUE COMPAREE POUR PLUSIEURS LANGUES

3.1. Caractéristiques du français

Une étude approfondie sur le français [22], nous a permis de conclure, au moins pour notre langue, que les deux degrés de labialité phonologique ne correspondaient pas de manière biunivoque à deux degrés de labialité articulatoire. De plus, à l'abaissement relativement progressif de la langue lors de la prononciation des voyelles d'une même série ne correspond pas nécessairement une ouverture progressive de l'orifice labial. Statistiquement parlant, et aussi bien pour les réalisations tenues que pour celles en contexte, il apparaît que les voyelles se regroupent selon trois classes de labialité distinctes que nous avons nommées pour le français: [-lab], [+lab] et [++lab] (Fig.1). Il est à remarquer que toutes les voyelles d'une même classe peuvent adopter la même forme frontale d'orifice labial, indépendamment de l'articulation linguale. On note encore que deux voyelles labiales peuvent s'opposer par la labialité: [ɑ,ɔ], ce qui contredit l'habituelle opposition +/-rond.

3.2. Comparaison à d'autres langues

Les langues étudiées sont généralement caractérisées par une opposition phonologique binaire de labialité vocalique. C'est-à-dire qu'elles possèdent des voyelles de même nature articulatoire s'opposant essentiellement par leur degré de labialité. Dans la mesure du possible, les systèmes choisis comportent des voyelles antérieures à la fois labialisées et non-labialisées.

Partant d'une classification articulatoire générale de la labialité vocalique que nous avons définie par ailleurs [22], nous passerons en revue les caractéristiques labiales proposées en illustrant par des exemples précis leur bien-fondé ou leur variation d'une langue à une autre.

3.2.1. Mode de labialisation

Cette caractéristique est généralement binaire: labialisé vs non-labialisé. Elle est valable pour toutes les langues observées et son choix relève de la description phonologique.

L'observation révèle que pour un même mode de labialité, divers types de labialisation peuvent être utilisés, qui génèrent néanmoins une même aire labiale, et permettent donc d'aboutir à un même résultat acoustique.

3.2.2. Type de labialisation

Pour le moment, nous en retenons cinq. Ils sont le reflet direct de l'activité musculaire:
- labialisé arrondi protrus
- labialisé écrasé non-protrus
- non-labialisé écrasé
- écarté, avec recul latéral des commissures
- neutre, contrôlé par les mouvements du maxillaire.

Un même mode peut donc être obtenu de diverses manières, selon son type:
Mode labial:
- le plus fréquemment réalisé par une protrusion et un arrondissement des lèvres; c'est le cas pour la plupart des langues observées,
- mais parfois une forte diminution de l'aire labiale est obtenue simplement par un écrasement vertical de l'orifice, comme pour certaines prononciations de l'anglais (Fig.2).

Mode non-labial:
- il est généralement caractérisé par une forme assez variable et surtout une aire relativement importante;
- en revanche, on rencontre parfois dans une même langue, à côté du premier type de non-labiales, des voyelles caractérisées par un orifice très écrasé verticalement. Il s'en suit une réduction importante de l'aire mais qui reste néanmoins nettement supérieure à celle subie par les voyelles labialisées: [i,ɨ] polonais [4], [ɜ:] anglais [3], [i,ɔ] thaï (Fig.3), anglais (Fig.2).

Certaines langues possèdent des voyelles labialisées très ouvertes qu'il est difficile de classer avec les types précédents, par exemple: [ɑ:] suédois [13] et en thaï (Fig.3). Il nous paraît judicieux de les appeler "neutres" et de les considérer plutôt comme non-labialisées.

Bien que notre étude ne traite pas directement de l'activité musculaire, nous avançons l'hypothèse que chacun de ces types de labialisation pourrait constituer un "axe labial naturel" [16], caractérisé par un ensemble de voyelles impliquant l'activité progressive non contradictoire d'un muscle ou d'un groupe de muscles ; par opposition à un axe non-naturel impliquant une réorganisation complète de l'activité musculaire entre les voyelles.

3.2.3. Degré de labialisation

Il reflète globalement l'ouverture labiale, c'est-à-dire à la fois l'espace vertical inter-labial B et l'aire aux lèvres S. Il peut être contrôlé par les mouvements du maxillaire inférieur ou par ceux des lèvres.

On peut rencontrer, selon les langues, un même degré de labialisation pour plusieurs voyelles appartenant à une même série, par exemple en français pour [u,o] ou [y, ø], ou au contraire des degrés différents variant parallèlement à l'abaissement de la langue, comme en thaï.

3.2.4. Stratégie de labialisation

Elle gère les degrés respectifs de labialisation des différentes voyelles d'une

même catégorie en les rendant, par exemple:
- dépendants de l'aperture intra-buccale de la voyelle, et donc en général soumis à une variation graduelle, comme en suédois, en finlandais [13] et en thaï (Fig. 3).
- indépendants de la voyelle, qui adopte alors un degré soit relativement constant: voyelles françaises [y,ø,u,o] (Fig.1), soit au contraire aléatoire: voyelles [i,e,ɛ,a] du français (Fig.1) ou [i,ɛ,a] du cantonais (Fig.4).

3.2.5. Catégorie labiale

Partant de là, nous appelons catégorie labiale un regroupement de voyelles ou de réalisations vocaliques de même mode et de même type gouvernées par une même stratégie, celle-ci indiquant comment est géré le degré de labialisation au sein de la catégorie. Une catégorie est donc obtenue moyennant le respect d'un ensemble de contraintes articulatoires et bio-mécaniques, et de stratégies motrices.

Le nombre de catégories peut différer pour deux langues même si, pour des séries apparemment identiques, celles-ci comportent les mêmes voyelles ou du moins utilisent les mêmes symboles vocaliques. Par exemple, nous dirons que le français possède deux catégories, ou deux degrés différents de labialité pour les voyelles labiales (moyennement et fortement labialisées : [+lab] et [++lab]), alors que l'anglais et le thaï n'en possèdent qu'une (Fig.1,2,3). En revanche, le thaï possède deux catégories de non labialisées: des écartées et des écrasées, à l'opposé du français, qui ne possède qu'une catégorie d'écartées.

3.2.6. Enfin, à ces différentes caractéristiques relevant directement de la langue parlée, et donc de sa "base articulatoire" labiale [17], s'ajoutent évidemment des **comportements individuels** qui relèvent à la fois des habitudes articulatoires et coarticulatoires de chacun.

4. REFERENCES

[1] ABRY C. & BOË L.J. (1983) "L'encodage labial des voyelles du français", *Speech Communication*, 2, 123-128.

[2] ABRY C., BOË L.J., CORSI P., DESCOUT R., GENTIL M. & GRAILLOT P. (1980) *Labialité et Phonétique*. Institut de Phonétique de Grenoble, 304 p.

[3] BOLLA K. (1989) *A Phonetic Conspectus of English*. Magyar Fonetikai Füzetek, Hungarian Papers in Phonetics, 20, Budapest, 402 p.

[4] BOLLA K. & FOLDI E. (1987) *A Phonetic Conspectus of Polish*. Magyar Fonetikai Füzetek, Hungarian Papers in Phonetics, 18, Budapest, 400 p.

[5] BONNOT J.F. & BOTHOREL A. (1989) "Co-dépendance des traits phonétiques, sensibilité au contexte et variabilité paramétrique", in *Mélanges de Phonétique Générale et Expérimentale offerts à Péla SIMON*, Inst. de Phon.de Strasbourg, 95-116.

[6] BONNOT J.F., CHEVRIE-MULLER C., GREINER G., MATON B. & GUIDET C. (1983) "Etude de l'encodage moteur des traits de nasalité et de labialité à partir de l'activité EMG des muscles orbiculaires (OO) et élévateur du voile (LP)", *11th Int. Cong. Acous.*, Toulouse, p.76.

[7] BOTHOREL A., SIMON P., WIOLAND F. & ZERLING J.P. (1986) *Cinéradiographie des voyelles et consonnes du français*. Publication de l'Institut de Phonétique de Strasbourg, 298 p.

[8] CATHIARD M.A. (1989) "La perception visuelle de la parole: aperçu de l'état des connaissances", *Bulletin de l'Institut de Phonétique de Grenoble*, vol. 17/18, 109-193.

[9] FROMKIN V.A. (1964) "Lips positions in American English vowels", *Language and speech*, 7, 215-225.

[10] GENTIL M. (1980) *Labialité en français: étude phonétique et aspects physiologiques des lèvres*, Thèse de 3e cycle, Université de Grenoble III, 440 p.

[11] GENTIL M. (1981) "Etude de la perception de la parole: lecture labiale et sosies labiaux", *Rapport IBM*, France, personal communication.

[12] LALLOUACHE M.T. (1991) *Traitement et analyse des images "visage-parole", Thèse de l'I.N.P.G.*, à paraître.

[13] LINKER W. (1982) "Articulatory and acoustic correlates of labial activity in vowels: a cross-linguistic study", *UCLA Working Papers in Phonetics*, 56, 134p.

[14] MAEDA S. (1989) "Articulation compensatoire des voyelles: analyse de données cinéradiographiques avec un modèle linéaire", in *Mélanges de Phonétique générale et expérimentale offerts à Péla Simon*, Publi. de l'Inst. de Phonétique de Strasbourg, 545-562.

[15] MAJID R., ABRY C., BOË L.J. & PERRIER P. (1987) "Contribution à la classification articulatori-acoustique des voyelles: étude des macro-sensibilités à l'aide d'un modèle articulatoire", *11th Int. Cong. Phon. Sc.*, Tallin, USSR.

[16] ROSSI M. (1983) "Niveaux de l'analyse phonétique: nature et structuration des indices et des traits", *Speech Communication*, 2, 91-106.

[17] STRAKA G. (1989) "Base articulatoire. Essai d'une mise au point", in *Mélanges de Phonétique générale et expérimentale offerts à Péla Simon*, Publication de l'Institut de Phonétique de Strasbourg,757-768

18] TSEVA A. (1990) "L'arrondissement dans l'identification visuelle des voyelles du français. Premiers acquis", *Bull. LCP Grenoble*, 3, 149-186.

[19] ZERLING J.P. (1980) "Coarticulation labiale et aire aux lèvres dans des groupes occlusives-voyelles en français", *Séminaire International Labialité*, GALF, Lannion, fév. 1980, 12p.

[20] ZERLING J.P. (1989,a) "Les trois degrés de labialisation des voyelles isolées en français. Etude pour 105 locuteurs", in *Mélanges de Phonétique Générale et Expérimentale offerts à Péla SIMON*, Inst. de Phon. de Strasb., 807-831.

[21] ZERLING J.P. (1989,b) "Stratégies labiales vocaliques en français. Variabilité des paramètres frontaux.", *Sém. Variab. et spéc. des locuteurs*, SFA-GCP, Marseille, 20-21 juin 89, 116-119.

[22] ZERLING J.P. (1990) *Aspects articulatoires de la labialité vocalique en français. Contribution à la modélisation à partir de labiophotographies, labiofilms, et films radiologiques*, Thèse d'Etat, Université de Strasbourg II, 600p.

[23] ZERLING J.P. (1991) "Frontal lip shape for French and English vowels: a cross-linguistic study", *2nd Seminar on Speech Production: Models and data*, Leeds, May 13-15th 1990, et in *Journal of Phonetics*, à paraître, juillet 1991.

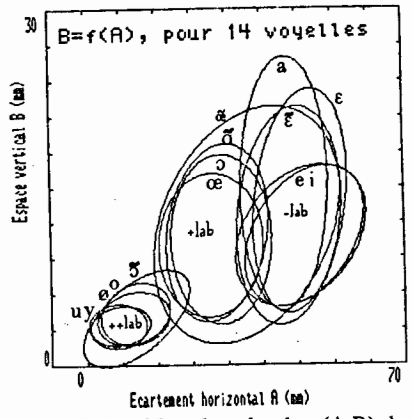


Fig.1 Répartition dans le plan (A,B) de 14 voyelles du français, 105 sujets, 1238 réalis., ell. à 90%, d'après [22, p.143]

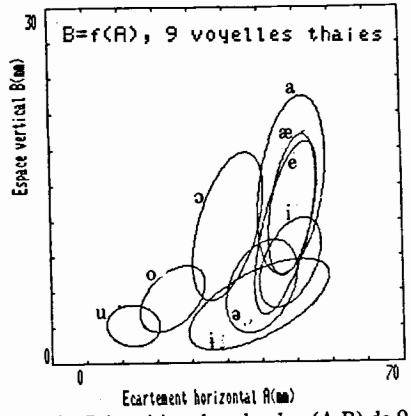


Fig.3 Répartition dans le plan (A,B) de 9 voyelles brèves et longues du thaï, 6 sujets, 108 réalis., ell. à 75% d'après [22, p.483]

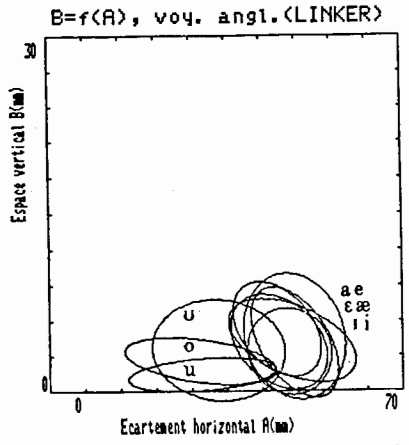


Fig.2 Répartition dans le plan (A,B) de 9 voyelles de l'anglais, 8 sujets, 72 réalis., ell. à75%, d'après [13]

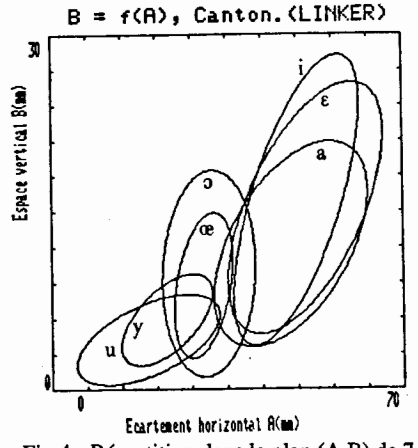


Fig.4 Répartition dans le plan (A,B) de 7 voyelles du cantonais, 8 sujets, 56 réalis., ell. à 75% d'après [13]

RESISTANCE DU [ɛ] SIMPLE ET DU [ɛ ɛ] DOUBLE A LA COARTICULATION MANDIBULAIRE EN ARABE MAROCAIN

N. Rhardisse R. Sock

Institut de la Communication Parlée, CNRS URA 368
Grenoble, France

ABSTRACT

This investigation attempts to characterize *coarticulation resistance* to speech rate for simple [l] and double [ll], through [i→a] and [a→i] vowel-to-vowel transitions, in Moroccan Arabic. Results show that the raising gesture for [i] has basically more coarticulatory influence on the consonantal gesture than the lowering gesture for [a], a trend which is especially evident for *geminates* at *normal rate*. These findings are discussed in the frame of speech motor control theories of anticipation.

1. INTRODUCTION

Le but de ce travail est de tester la résistance à la coarticulation mandibulaire des consonnes [l] vs. [ll], dans un système phonologique spécifique : l'arabe marocain; et ceci selon le contexte vocalique et face à la variation de la vitesse d'élocution. Une attention particulière sera donnée au paradigme expérimental de la *résistance coarticulatoire*, proposé par [5] pour une variante de l'anglais britannique. Nous utiliserons les concepts et les outils statistiques liés au contrôle moteur [7, 14]. L'approche relative de la durée (*timing* relatif) sera un moyen efficace pour apprécier les changements qui interviennent lorsqu'on fait varier le débit [6, 11]. Nous suivrons les principes méthodologiques sur la recherche en parole proposés par [2] qui se sont avérés rentables pour la mise en évidence du *timing* des oppositions phonologiques.

Enfin, nous essayerons de confronter nos résultats, sur la coarticulation anticipante, aux modèles *look-ahead* [9] et *time-locked* [3]. L'articulateur choisi dans cette étude est la mandibule : en effet, celle-ci joue un rôle important dans l'organisation temporelle, en régulant l'ouverture et la fermeture du tractus vocal.

2. METHODE

2.1.Corpus

Le corpus établi est tiré de la langue marocaine de la région de Fès, où le contrôle de la quantité consonantique [l] vs. [ll] est phonologique. Cela nous permet d'examiner la manière dont les consonnes simple [l] et double [ll] régulent le timing d'une trajectoire mandibulaire haut-bas dans une transition [i→a] et bas-haut dans une transition [a→i]. Les oppositions choisies sont les suivantes : [ala] "unité de mesure", [alla] "il a fait mijoter", [ali] "il faisait frire", [alli] "fais mijoter!", [ili] "médisance", [illi] logatome, [ila] "si", [illa] "sauf".
Ces items, insérés, dans une phrase porteuse assertive, ont été réalisés par une locutrice marocaine de Fès. Chaque item a été répété 12 fois dans un ordre aléatoire. La première série a été enregistrée en débit normal (conversationnel); la seconde série avec une exigence de débit rapide.

2.2.Mesures

Le signal acoustique, numérisé à 8kHz, a été étiqueté manuellement en événements [1] à l'aide d'un éditeur de signal [4]. Les signaux de déplacement vertical de la mandibule ont été recueillis à l'aide d'un kinésiographe mandibulaire (K5AR) et échantillonnés à 160 Hz. Les événements articulatoires ont été repérés sur les signaux de vitesse et d'accélération. Ces derniers ont été obtenus par dérivation du signal de position lissé par des fonctions splines cubiques.
Les événements acoustiques : VVO ou début du voisement vocalique (correspondant ici à la détente des [l]) et VVT ou fin du voisement vocalique (closion des [l]), nous fournissent la base temporelle VVT-VVO (tenue des [l]) comme domaine d'étude de la gémination consonantique et de la *résistance coarticulatoire*.
Sur le plan articulatoire nous avons repéré les événements :
- ACC : défini comme l'accélération maximale du geste d'abaissement mandibulaire pour la réalisation de la voyelle subséquente [a].
- DEC : défini comme la décélération maximale du geste d'élévation mandibulaire pour réaliser les consonnes [l] simple ou [ll] double.

Nous avons retenu, à partir de ces événements articulatori-acoustiques, les deux phases temporelles suivantes :
- LON (ACC-VVO) ou *Lowering Acceleration re : [a] Vowel Onset*, qui va du geste de l'accélération maximale de l'abaissement mandibulaire à l'établissement vocalique du [a] subséquent. Exprimée en pourcentage de la base VVT-VVO, elle nous donne le degré de coarticulation ou de ("pénétration") du [a] dans les consonnes.
- RON (DEC-VVO) ou *Raising Deceleration re : [i] Vowel Onset*, qui a pour bornes, la décélération maximale du geste de l'élévation mandibulaire pour [l] simple ou [ll] double et l'établissement de la structure formantique de la voyelle suivante [i]. Exprimée en pourcentage de VVT-VVO, elle nous donne le degré de coarticulation ou ("pénétration") du [i] dans les consonnes.
Pour les réalisations [ili] et [illi] le degré de coarticulation est à 100% (par défaut), étant donné que le geste d'élévation mandibulaire se réalise bien en amont de notre base temporelle.

3. RESULTATS

3.1.Oppositions

L'opposition entre [ala] et [alla] (fig. 1), se fait en débit normal aussi bien par la phase que par la base temporelle avec une différence de 22% en moyenne pour la phase LON et 83 ms en moyenne pour la base (respectivement, $t = 16.23$ et $t = 30.15$, significatifs à $p \leq 0,05$; même seuil pour les suivants). Lorsqu'on augmente le débit, cette opposition de phase et de base se maintient avec une différence de 24% en moyenne pour la phase et de 67 ms pour la base temporelle (respectivement, $t = 6.57$ et $t = 23.80$).
Quand on oppose [ila] à [illa] (fig. 2) en débit normal, on remarque une différence sur la phase de 47% en moyenne ($t = 8.12$). La différence, en moyenne de 82 ms, entre les deux bases temporelles est, bien entendu, significative ($t = 20.59$). En débit rapide, cette opposition des classes phonétiques ne se fait plus que sur la base temporelle, avec une différence en moyenne de 55 ms ($t = 21.80$).
En débit normal, l'opposition entre les classes [ali] et [alli] (fig. 3) se produit sur la phase RON et sur la base temporelle : la différence est, en moyenne, de l'ordre de 29% sur la phase et de 82 ms sur la base (respectivement, $t = 6.51$ et $t = 18.75$). L'opposition n'est maintenue en débit rapide que grâce à la base temporelle, avec une différence de 55 ms en moyenne entre les deux classes ($t = 17.89$).
L'opposition entre les classes [ili] et [illi] (fig. 4) se réalise seulement sur la base temporelle. En débit normal, on constate une différence de 70 ms en moyenne entre les deux classes phonétiques ($t = 17.73$). Cette différence est réduite en débit rapide : elle n'est plus en moyenne que de 40% ($t = 12.38$).
Une tendance générale est que *les classes des géminées dérivent vers les simples lorsque la tâche devient plus complexe* (voir fig. 1,2,3). Il est bien connu – dans le cadre des transitions de phase ou du paradigme de la Synergétique réactualisé par [8] –, que les structures complexes tendent vers des structures simples lorsque la tâche devient plus difficile (ici l'augmentation de la vitesse d'élocution). C'est un processus de simplification, démontré par les changements historiques (CC→C, CVC→CC, VCV→VV), que nous avions pu déjà – pour d'autres géminées de l'arabe marocain [13] – mettre en évidence sur le plan acoustique.

3.2.Coarticulation

L'analyse des résultats nous montre qu'on ne peut pas systématiser une seule stratégie de coarticulation pour la réalisation de l'opposition simple vs. double face à la variation du débit. Lorsqu'on oppose [ala] à [alla] (fig. 1) en débit normal, on constate que les doubles

sont plus coarticulées que les simples (≈ 50% vs. ≈30%), ce qui signifie que [ll] double est moins résistant en débit normal. En débit rapide, la tendance reste structurellement la même (≈ 30% vs. ≈ 50%).

Pour l'opposition [ila] ~ [illa] (fig. 2), nous constatons le phénomène inverse. En débit normal, les simples sont plus coarticulées, donc moins résistantes, que les doubles (≈60% vs. ≈20%). En débit rapide, les deux classes ont des pourcentages comparables de coarticulation : elles se confondent à environ 70%, ce qui est un taux assez important de coarticulation.

Mais l'examen des classes [ali] ~ [alli] (fig. 3) nous révèle, qu'en débit normal, ce sont les doubles qui sont le plus coarticulées (à environ 65%). Ici, comme pour les classes [ala] et [alla] (fig. 1), les doubles sont moins résistantes à la coarticulation en débit normal. Cependant en débit rapide, le degré de coarticulation est semblable pour simples et doubles, avec un taux de résistance moindre (≈45% vs. ≈35%).

Enfin, on peut dire de manière générale que les classes [ili] et [illi] (fig. 4) sont de loin les plus coarticulées, avec un minimum de 100% de coarticulation (par défaut, cf. *supra*).

D'après ces données nous pouvons poser que le [i] a intrinsèquement une puissance de "pénétration" plus élevée que le [a], ce que révèle particulièrement le comportement des *géminées* en débit *normal*. Nous pensons donc, comme [10] et [12], que la consonne [l] semble mieux épouser la hauteur mandibulaire de la voyelle [i], ce qui expliquerait son taux élevé de coarticulation.

4. CONCLUSION

En conclusion, nous pouvons souligner que le phénomène de *résistance à la coarticulation anticipante* des [l] en arabe marocain peut comporter aussi bien une composante largement partagée par d'autres langues qu'une autre plus ou moins spécifique.

C'est ainsi, en commençant par l'aspect spécificité, que l'on peut valider, d'une part, le modèle *time-locked*; mais seulement en *timing relatif*, avec ceux de nos résultats qui montrent une stabilité des phases malgré la variation de la vitesse d'élocution. C'est le cas pour la classe [alla], où l'accélération maximale se produit à intervalle *proportionnellement fixe* par rapport au début acoustique de la voyelle suivante.

Mais d'autre part, nous pouvons évoquer le comportement "orthodoxe" – par rapport au modèle *look-ahead* – de la classe phonétique [illi] : elle n'invaliderait pas un tel modèle, car la coarticulation anticipante est maximale, que la consonne soit simple ou géminée.

Remerciements à G. Feng pour son aide en traitement des signaux; à C. Abry pour ses commentaires.

REFERENCES

[1] ABRY, C. BENOIT, C. BOE, L.J. & SOCK, R. (1985), "Un choix d'événements pour l'organisation temporelle du signal de la parole", *14èmes JEP du GCP du GALF*, 133-137.

[2] ABRY, C. ORLIAGUET, J.P. & SOCK, R. (1990), "Patterns of Speech Phasing. Their Robustness in the Production of a Timed Linguistic Task : Single vs. Double (abutted) Consonants in French", *Europ. Bull. of Cog. Psych.* : 10, 269-288.

[3] BELL-BERTI, F. & HARRIS, K.S. (1979), "Anticipatory Coarticulation : Some Implications of Lip Rounding", *J. Acoust. Soc. Am.* : 65, 1268-1270.

[4] BENOIT, C. (1984), "EDISIG : Encore un Editeur de signal!?", *13èmes JEP du GALF*, 211-213.

[5] BLADON, R.A. & AL-BAMERNI, (1976), "Coarticulation Resistance in English /l/", *J. of Phon.* : 4, 137-150.

[6] FOWLER, C.A. (1980), "Coarticulation and Theories of Extrinsic Timing", *J. of Phon.* : 8, 113-133.

[7] GENTNER, D.R. (1987), "Timing of Skilled Motor Performances. Tests of the Proportional Duration Model", *Psych. Rev.* : 82, 225-260.

[8] HAKEN, H. KELSO, J.A.S. & BUNZ, H. (1985), "A Theorerical Model of Phase Transitions in Human Hand Movements", *Biol. Cyb.* : 51, 347-356.

[9] HENKE, W.L. (1967), "Preliminaries to Speech Synthesis Based on an Articulatory Model", *Speech Conference, Boston*, 170-177.

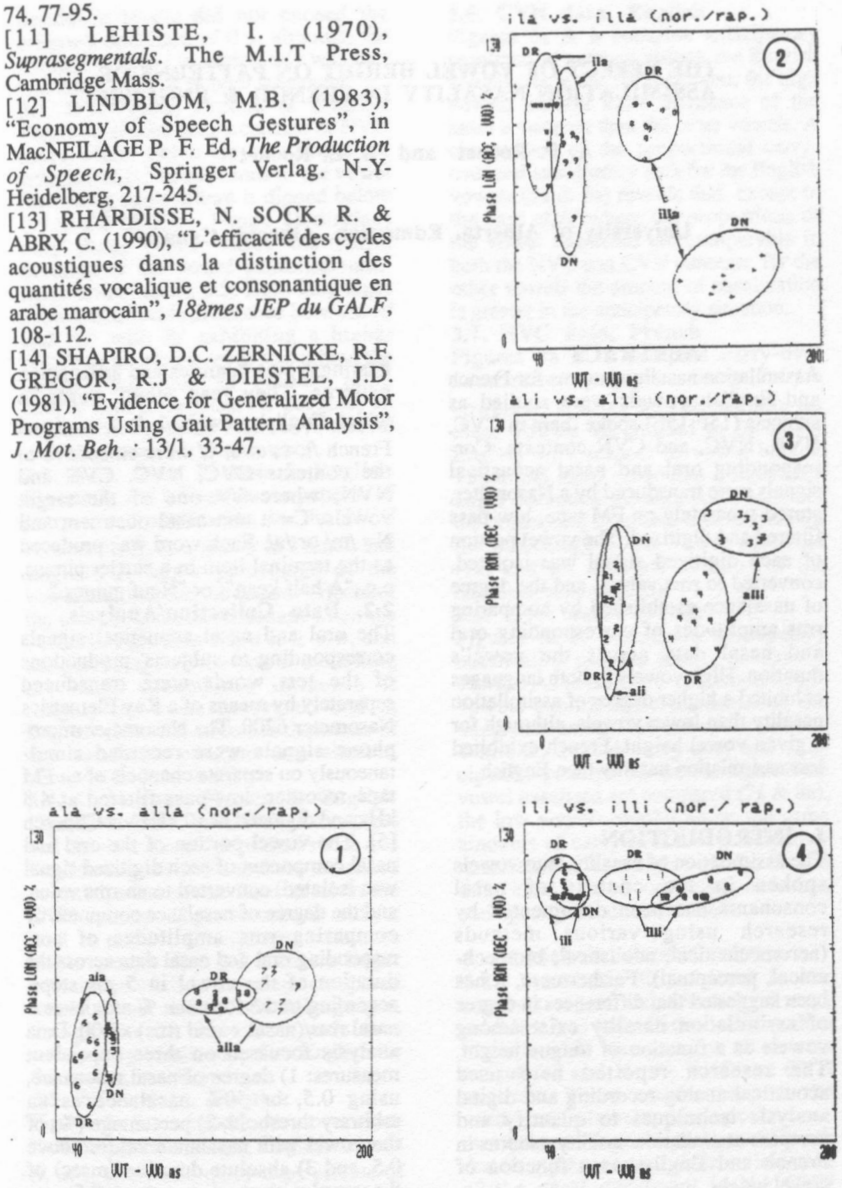
[10] KEATING, P.A. LINDBLOM, B. LUBKER, J. & KREIMAN, J. (1990), "Jaw Position in English and Swedish VCVs", *Working Papers in Phonetics* : 74, 77-95.

[11] LEHISTE, I. (1970), *Suprasegmentals*. The M.I.T Press, Cambridge Mass.

[12] LINDBLOM, M.B. (1983), "Economy of Speech Gestures", in MacNEILAGE P. F. Ed, *The Production of Speech*, Springer Verlag. N-Y-Heidelberg, 217-245.

[13] RHARDISSE, N. SOCK, R. & ABRY, C. (1990), "L'efficacité des cycles acoustiques dans la distinction des quantités vocalique et consonantique en arabe marocain", *18èmes JEP du GALF*, 108-112.

[14] SHAPIRO, D.C. ZERNICKE, R.F. GREGOR, R.J & DIESTEL, J.D. (1981), "Evidence for Generalized Motor Programs Using Gait Pattern Analysis", *J. Mot. Beh.* : 13/1, 33-47.



Figs. 1-4. Ellipses de dispersion (à 90%) pour l'ensemble des classes phonétiques dans les deux débits, normal (DN) et rapide (DR). En abcisse : durée en ms de la base temporelle (VVT-VVO) ; en ordonnée : pourcentage des phases (LON ou RON) en fonction de la base temporelle (cf. texte).

THE EFFECT OF VOWEL HEIGHT ON PATTERNS OF ASSIMILATION NASALITY IN FRENCH & ENGLISH

A. P. Rochet and B. L. Rochet

University of Alberta, Edmonton, Alberta, Canada

ABSTRACT

Assimilation nasality patterns for French and English vowels were studied as subjects (15F;15E) spoke them in CVC, NVN, NVC, and CVN contexts. Corresponding oral and nasal acoustical signals were transduced by a Nasometer, stored separately on FM tape, low-pass filtered and digitized. The vowel portion of each digitized signal was isolated, converted to rms values, and the degree of nasalance established by comparing rms amplitudes of corresponding oral and nasal data across the vowel's duration. High vowels in both languages exhibited a higher degree of assimilation nasality than lower vowels, although for a given vowel height, French exhibited less assimilation nasality than English.

1. INTRODUCTION

The assimilation of nasality onto vowels spoken in the context of nasal consonants has been documented by research using various methods (aeromechanical, acoustical, biomechanical, perceptual). Furthermore, it has been suggested that differences in degree of assimilation nasality exist among vowels as a function of tongue height. The research reported here used acoustical analog recording and digital analysis techniques to quantify and compare assimilation nasality patterns in French and English as a function of vowel height.

2. PROCEDURES

2.1. Subjects/Speech Sample

Subjects were 30 young adults, 15 native speakers of Standard French and 15 of Canadian English, with normal hearing, voice qualities and articulation patterns. They read aloud words in which English vowels /i, I, ɛ, a, u/ and French /i, ɛ, a, u, y/ were embedded in the contexts CVC, NVC, CVN and NVN, where V= one of the target vowels, C= a non-nasal obstruent and N= /m/ or /n/. Each word was produced as the terminal item in a carrier phrase, e.g.,"A half keen"; or "Neuf quines."

2.2. Data Collection/Analysis

The oral and nasal acoustical signals corresponding to subjects' productions of the test words were transduced separately by means of a Kay Elemetrics Nasometer 6200. The Nasometer microphone signals were recorded simultaneously on separate channels of an FM tape recorder, low-pass filtered at 4.8 kHz and digitized at 10 kHz via CSpeech [5]. The vowel portion of the oral and nasal component of each digitized signal was isolated, converted to an rms value, and the degree of nasalance computed by comparing rms amplitudes of corresponding oral and nasal data across the duration of the vowel in 5 ms steps, according to the formula: % nasalance = nasal rms/(nasal + oral rms) x 100. Data analysis focussed on three dependent measures: 1) degree of nasal resonance, using 0.5, or 50% nasalance as an arbitrary threshold, 2) percentage (%) of the vowel with nasalance values above 0.5, and 3) absolute duration (msec) of the vowel with nasalance above 0.5.

3. RESULTS

3.1. CVC data, French & English

Figure 1 depicts the percentage of CVC cases without significant nasalance (i.e., <0.5). In the majority of cases, nasalance levels did not exceed the arbitrary threshold of 0.5, although the number of cases in which this was true was smaller for /i/ in both languages.

3.2. NVN data, French & English

Figure 2a graphs the percentage of NVN cases where nasalance was above the criterion of 0.5 at both ends of the vowel (including cases where it dipped below 0.5 in the middle). Figure 2b displays only those cases where the entire duration of the vowel exhibited nasalance levels above 0.5. Both languages show a noticeable difference between /i/ and /a/, with /i/ exhibiting a higher sustained nasalance level throughout the vowel's duration. In French, more clearly than in English, /u/ occupies an intermediate position between /i/ and /a/ with respect to this phenomenon.

3.3. NVC data, French & English

Figures 3a and b illustrate the patterns of carry-over nasalization in the NVC context for /i, u, ɛ and a/ in French and English. A larger percentage of the vowel exhibits the carry-over effects of the preceding nasal consonant when the vowel is high than when it is low, and the percentage of the vowel exhibiting the nasal consonant's influence is roughly the same in French and in English (3a). The absolute durations of the nasalized portions are shorter, however, in French (3b).

3.4. CVN data, French & English

Figures 4a & b illustrate the percentages and absolute durations of the French and English target vowels that are nasalized in anticipation of the final nasal in the CVN context. High vowels /i/ and /u/ tend to exhibit anticipatory nasalance levels greater than 0.5 across a larger percentage of their durations compared to mid or low vowels in both languages, though French always reveals less anticipatory nasalization than English for the vowels considered.

3.5. NVC data, English

Figures 5a & b compare carry-over nasalization patterns among English vowels /i, u, I, ɛ, a/. The carry-over effects of the initial nasal consonant influence a larger portion of the high vowels than of the others, and the effect is consistent whether one considers the percentage of vowel nasalized (5a), or the absolute duration of the nasalized segment (5b).

3.6. CVN data, English

Figures 6a & b compare anticipatory nasalization patterns among the English vowels. As in the NVC context, the high vowels exhibit more influence of the nasal consonant than the other vowels. A comparison of the proportional carry-over and anticipatory data for the English vowels (5a & 6a) reveals that, except in the case of /i/ where the proportions of the vowel nasalized are comparable in both the NVC and CVN contexts, for the other vowels the amount of nasalization is greater in the anticipatory situation.

3.7. NVC data, French

Figures 7a & b compare carry-over nasalization patterns among French vowels /i, y, u, ɛ, a/. The patterns for these vowels are similar to those for English with respect to vowel height: The initial nasal consonant influences a larger portion of the high vowels than of the others, and the effect is consistent for the percentage of the vowel nasalized (7a) and the absolute duration of the nasalized segment (7b).

3.8. CVN data, French

Figures 8a & b compare anticipatory nasalization patterns among the French vowels. As in the NVC context, the high vowels exhibit more influence of the nasal consonant than the other vowels. When the French carry-over and anticipatory patterns for percentage of the vowel nasalized are compared (7a & 8a), the low vowels exhibit about the same amounts of carry-over and anticipatory nasalization. The high vowels, on the other hand, exhibit more carry-over than anticipatory nasalization effects.

4. SUMMARY/DISCUSSION

4.1. For these 30 subjects in the contexts examined, French always exhibited less assimilation nasality than English. This was true for all vowels considered and for the NVC and CVN contexts. These results support the validity of Delattre's pedogogical recommendation to English speakers of French that they prevent premature anticipation of the nasal consonant in the CVN context in order not to nasalize the vowel [3]. These data do not, however, support Delattre's assertion that French vowels followed by a nasal consonant remain oral throughout their duration.

4.2. The degree of vowel nasalization in a nasal consonant context varied with the height of the vowel. High vowels exhibited more assimilation nasality than low vowels. This correlation is very systematic in the French vowel data; it also applies to the English vowel data, although less systematically. The apparent contradiction between these results and those of Clumeck [1] may be related to his use of the term "nasalized" to describe articulatory gestures of the velum, and the fact that the biomechanical behavior of the velopharynx cannot be assumed to be monotonically related either to the perception of nasal resonance or to the acoustical consequences of nasal coupling during speech production. The perception or measurement of nasal resonance is ultimately a function of the relative acoustical impedances of the oral and nasal cavities, as well as the formant frequency values of the vowel in question. The spectral envelopes of /i/ and /u/ are markedly affected by small nasal coupling, whereas vowels with a more open tract configuration are much less affected by small degrees of coupling [2]. This is consistent with listeners' judgements that the amount of nasal coupling necessary for the perceptual identification of nasalization was almost three times as much for low vowels as for high ones [4].

5. CONCLUSIONS

5.1. The difference in the degree of nasalance between French and English may be related to the fact that English does not have phonemic nasal vowels and therefore can "tolerate" higher levels of assimilation nasality.

5.2. The higher levels and longer durations of assimilation nasality observed for the high vowels in both French and English are related to the acoustical impedance of the vocal tract for the production of these vowels. There is no obvious articulatory or physiological reason for the earlier lowering of the velum observed by Clumeck [1] for low vowels in the CVN context. It may simply be that such lowering does not have an undesirable acoustical effect, and does not lead to excessive perceptible nasalization of these vowels. Later lowering of the velum for high vowels, however, may ensure that their spectral envelopes are not too drastically affected by extraneous nasal resonance.

5.3. Further research on assimilation nasality is recommended by means of simultaneous multidimensional sampling methods that could consider biomechanical, perceptual and acoustical parameters of vowel production without losing sight of the phonemic characteristics of the languages sampled.

6. REFERENCES

[1] CLUMECK, H. (1976), "Patterns of soft palate movement in six languages" *Journal of Phonetics 4*, 337-351.

[2] CURTIS, J. (1968), in "*Cleft Palate and Communication,*" D. Spriestersbach & D. Sherman (eds.) New York: Academic.

[3] DELATTRE, P. (1951) "A l'usage des étudiants anglo-américains" 2nd ed. *Principes de Phonétique Française*, Middlebury, VT: Middlebury College.

[4] HOUSE, A. and STEVENS, K. (1956) "Analog studies of the nasalization of vowels" *Journal of Speech & Hearing Disorders 21*, 218-232.

[5] MILENKOVIC, P. (1990) Department of Electrical & Computer Engineering, University of Wisconsin, Madison, WI 53705 U.S.A.

FIGURE 1: % /CVC/ CASES WITHOUT SIGNIFICANT NASALANCE (<0.5)

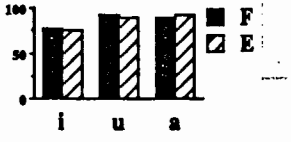


FIGURE 2a: % /NVN/ CASES WITH NASALANCE >0.5 AT BOTH ENDS

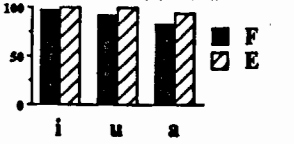


FIGURE 2b: % /NVN/ CASES WITH NASALANCE >0.5 OVERALL

FIGURE 3a: /NVC/ FRENCH & ENGLISH PERCENTAGE OF VOWEL NASALIZED

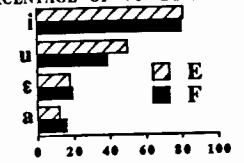


FIGURE 3b: /NVC/ FRENCH & ENGLISH DURATION OF NASALIZATION (MSEC)

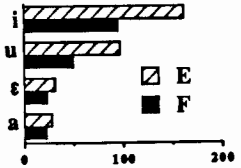


FIGURE 4a: /CVN/ FRENCH & ENGLISH PERCENTAGE OF VOWEL NASALIZED

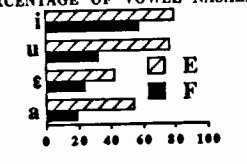


FIGURE 4b: /CVN/ FRENCH & ENGLISH DURATION OF NASALIZATION (MSEC)

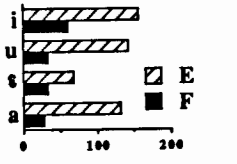


FIGURE 5a: /NVC/ ENGLISH PERCENTAGE OF VOWEL NASALIZED

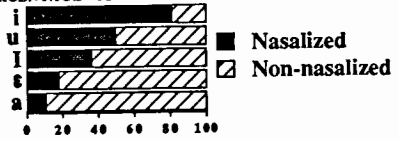


FIGURE 5b: /NVC/ ENGLISH DURATION OF NASALIZATION (MSEC)

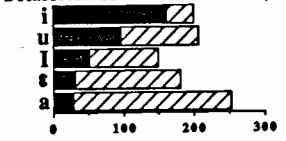


FIGURE 6a: /CVN/ ENGLISH PERCENTAGE OF VOWEL NASALIZED

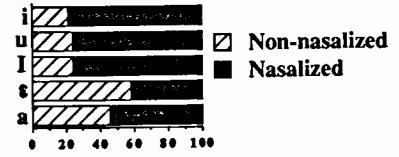


FIGURE 6b: /CVN/ ENGLISH DURATION OF NASALIZATION (MSEC)

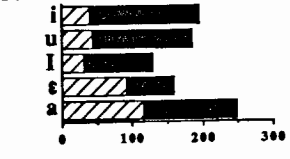


FIGURE 7a: /NVC/ FRENCH PERCENTAGE OF VOWEL NASALIZED

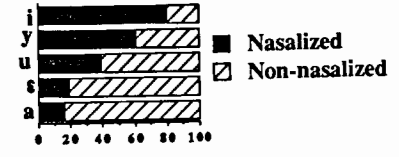


FIGURE 7b: /NVC/ FRENCH DURATION OF NASALIZATION (MSEC)

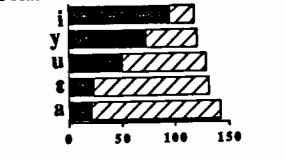


FIGURE 8a: /CVN/ FRENCH PERCENTAGE OF VOWEL NASALIZED

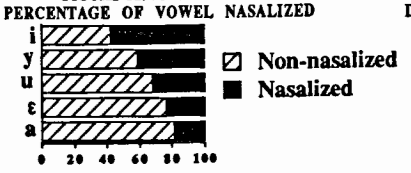
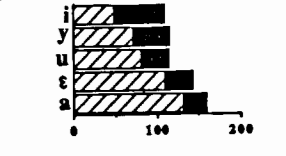


FIGURE 8b: /CVN/ FRENCH DURATION OF NASALIZATION (MSEC)



FRICATIVE CONSONANTS AND THEIR ARTICULATORY TRAJECTORIES

Celia Scully*, Esther Georges*, Eric Castelli*

University of Leeds, UK*
ICP,INPG, Grenoble, France*

ABSTRACT

Articulatory paths relevant to the production of some fricatives are related to the changing acoustic patterns as seen on spectrograms. Two techniques are used: aerodynamically derived area (A) traces and electropalatography (EPG) contacts. [pVCV] sequences are analysed.

1. INTRODUCTION

Better descriptions are needed for the production of fricative consonants. Fricatives in speech-like sequences are characterised not only by their quasi-static spectra but also by rapidly changing acoustic patterns. The latter seem to be essential for the identification of /f/ versus /θ/ and /v/ versus /ð/ [2], [3]. Phonemically, fricative consonants and adjacent vowels are considered as separate entities but in the processes of speech production there is no clear boundary between them. Between segments clearly associated with either consonants or vowels there are regions of rapid change: in these, there are changing combinations of the acoustic sources - voice, aspiration noise (generated just above the glottis) and frication noise (generated downstream from a vocal tract constriction) - as well as rapidly changing formant frequencies. Inter-articular coordination and the form of transitions for individual articulators are both important in determining how sources and filters covary across this boundary region. The study reported here is an exemplification of part of a larger study (Grenoble, Southampton, Leeds), based on multiple analyses for two speakers. A studio recording made by the speaker provides cueing, so as to match the speaking style and rate across data gathered on different occasions and in different laboratories. Articulatory paths in the natural speech are to be copied in models of speech production [4], [1], [6]. Analysis-by-synthesis will be used to obtain good aerodynamic and acoustic matches between the natural and the simulated speech. The aim is to characterise, as general rules, the production and acoustics of the speakers' fricatives.

2. APPROACH OF THE STUDY

This paper focusses on [s] produced in phonetically controlled [V-V] contexts by one of the speakers, a woman speaker of General American English. Sequences such as [pi'sipi'sipi i...] produced on a single expiratory breath allowed subglottal pressure (PSG) to be estimated.

Two techniques for the estimation of vocal tract articulation relevant to the production of alveolar fricatives are included here. First, aerodynamic parameters are used to give an Area (A) trace, indicating the cross-section area of the alveolar constriction of the vocal tract; secondly, electropalatography (EPG) is used to show the regions of contact between the tongue and the hard palate. Articulatory paths estimated by each method in turn are time matched to spectrograms from simultaneously made recordings. In this way, part of the detailed articulatory-to-acoustic mapping is studied. An additional aim is to demonstrate that the two methods are consistent and complementary.

3. AREA (A) TRACE

3.1. Method

This is a parameter for one of the two major constrictions of the respiratory tract, the other one being the glottal area. Volume flowrate of air through the mouth, U (in cm^3/s) and oral air pressure, P (in cmH_2O) are combined, using the orifice equation with an empirical constant k = 0.00076 to give:

$A = k.U/P^{0.5}$

The methods have been described elsewhere [5], [7]. The A trace is not the true value of the minimum cross-section area for the alveolar ridge portion of the vocal tract, and may be expected to depend on the taper angle into and out of the constriction and its length. Total airflow through nose as well as mouth is recorded, but checks showed that all the airflow was through the mouth for the sequences analysed. Oral air pressure is taken as pressure drop across the constriction but actually includes any pressure drop across the teeth and lips also. The method has the advantage that it links articulation and aerodynamics in a way that is internally consistent and consistent with the descriptive framework of one of the composite models [7] in which it is to be used.

3.2. Results

Figure 1 shows an example of the aerodynamic traces, with some of the simultaneously obtained acoustic traces.

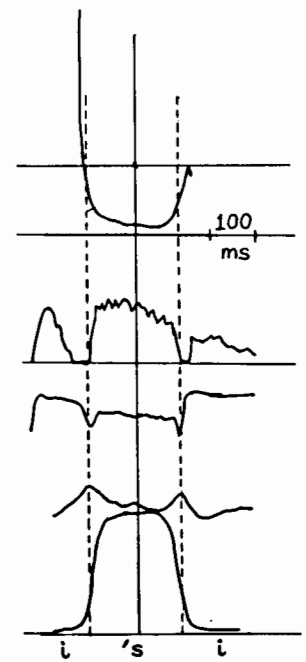


Figure 1. Articulatory, aerodynamic and acoustic traces as functions of time, for the third in a series of [(p)i'si(p)]. from top to bottom: Area A, with a line at 0.2 cm^2; I.L. H.P. filtered at 3.9 kHz, I.L. H.P. filtered at 500 Hz; Oral (total) volume flowrate of air U; oral air pressure P.

An auditory check confirmed that this

was an acceptable example of [i'si]. Figure 2 shows the relationship between the articulatory trace and the acoustics.

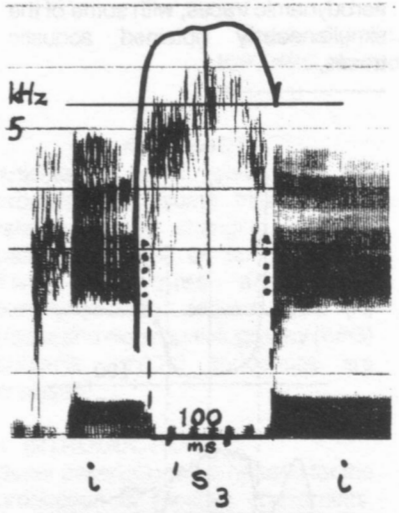


Figure 2. Area (A) trace from Figure 1, inverted, combined with spectrogram (frequency versus time) pattern features of the identical utterance. Time scales are matched with voice offset and onset aligned.

It may be seen that the 0.2 cm^2 threshold for the A trace goes beyond the domain of frication noise to include the boundary regions. The inverted form of the A trace may match the changing spectral pattern for the frication noise, but the evidence is not conclusive.

The peaks of airflow seen in Figure 1 and shown by the dotted lines in Figure 2 almost coincide with voice offset and voice onset shown by the dashed lines. The airflow peaks are located at the boundary region, between the frication noise segment and the vocoid, where the acoustic sources including voice and aspiration noise are changing rapidly.

4. EPG DATA

4.1. Method

EPG data for the fricatives in the same [...pV'-Vp...] context are analysed as follows: the number of contacts is determined for: the first, second and third lines of contact (*front*, shown by a solid line) and the fourth to the eighth line (*back*, shown by a dotted line). The results are plotted on a grid which shows time vs number of contacts. The two resulting traces represent changes in the amount of contact between tongue and palate for the front and the back of the mouth; transitions from and to vowels are investigated.

4.2 Results

Figure 3 shows the relationship between the articulatory traces and the acoustics for a representative example of [i'si].

A threshold (indicated by the arrow in Figure 3) was chosen to define the quasi-static segment observed for all of the fricatives analysed so far. It was found that this threshold corresponds rather closely with the frication noise segment. The match is excellent in this example.

The contact for the back portion of the tongue, however, decreases during the frication noise segment. This may perhaps indicate a lowering of the tongue dorsum similar to that observed by Wood [8] on X-ray traces. As Wood suggests, this may enhance the acoustic separation of front and back cavities. Referring back to the noticeable change in spectrum during this portion of a different production of the same fricative, seen in Figure 2, this explanation seems a plausible one. The speaker appears to be tuning the front cavity resonance.

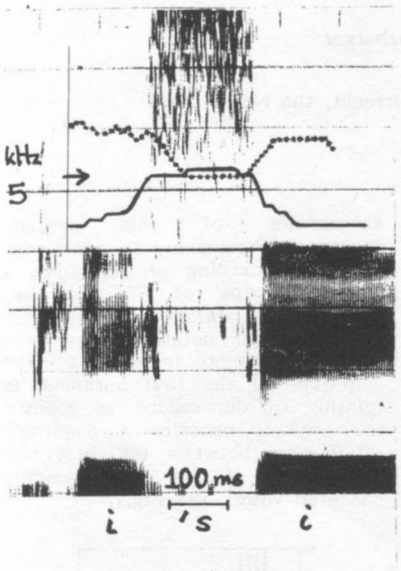


Figure 3. EPG contact traces combined with spectrogram pattern features (solid line for front contacts, dotted line for back contacts).

5.CONCLUSION

The area trace seems to define a wider domain than does the front contact shown by the EPG data. The area trace may reflect changes in the overall tongue configuration such as that discussed above, and possibly mouth outlet shape also. This interpretation of the two kinds of articulatory data will be tried out in the modelling.

6. ACKNOWLEDGEMENT

The work was funded in part by a collaborative EC SCIENCE award: CEC-SCI * OI47C (EDB).

7. REFERENCES

[1] CASTELLI, E. (1989) "Caractérisation acoustique des voyelles nasales du français", Ph.D. thesis, INPG, Grenoble.

[2] HARRIS, K.S. (1958) "Cues for the discrimination of American English fricatives in spoken syllables", *Language and Speech, 1,* 1-7.

[3] HEINZ & STEVENS (1961) "On the properties of voiceless fricative consonants", *J.Acous.Soc.Am., 33,* 589-596.

[4] MAEDA, S. (1990) "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model" in *Speech Production and Speech Modelling*, W.J. Hardcastle & A. Marchal, eds., 131-149.

[5] SCULLY, C. (1986) "Speech production simulated with a functional model of the larynx and the vocal tract", *J.Phonetics, 14,* 407-413.

[6] SCULLY, C. (1987) "Linguistic units and units of speech production", *Speech Comm., 6,* 77-142.

[7] SCULLY, C., CASTELLI, E., BREARLEY, E. & SHIRT, M. (1991) "Articulatory paths and aerodynamic patterns for some fricatives", *J.Phonetics*, in press.

[8] WOOD, S.A.J. (1991) "Crosslinguistic X-ray data on the temporal coordination of speech gestures", *J.Phonetics*, in press.

BREATHINESS IN MALE AND FEMALE SPEAKERS

D. Günzburger

Institute of Phonetics, Utrecht, the Netherlands

ABSTRACT

The present study provides data on degree of breathiness produced by Dutch male and female speakers in a neutral and an emotive context. The acoustically defined parameter DH indicates significant differences between male and female speakers in both contexts. There is an increase in breathiness for either population from neutral to emotive context. Analysis of average F0 and average intensity levels show decreased values for both male and female speakers in the emotive condition as opposed to the neutral condition.

1. INTRODUCTION

Breathiness can be defined in various domains. In **articulatory** terms breathy phonation arises from an incomplete adduction of the vibrating vocal folds and can lead to an increase of the average airflow of up to 60% in comparison to non-breathy (vowel) production. Extreme breathiness can be indicative of pathological speech and function as a **perceptual** marker of various laryngeal disorders. On a less extreme note breathiness can impair the general perceptibility and understandibility of speech and convey the impression of increased monotony. Various **acoustic** correlates can account for breathiness [1I,13I. Due to the incomplete vocal fold closure during phonation of a breathy vowel there is considerable leakage of air through the glottis which causes interspersed noise at higher frequencies of the acoustic spectrum. Presumably in connection with a slackening of the folds a slight lowering of F0 has been observed for breathy vowels and, probably most notable, there is a fairly consistent increase of the amplitude of the first harmonic in relation to the second as opposed to an opposite amplitude relationship between the first two harmonics of a non-breathily phonated vowel, see Fig.1.

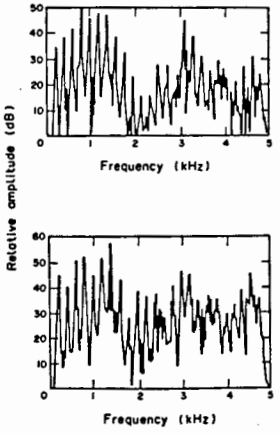


Fig.1: Spectra of non breathy vowel (top) and breathy vowel (bottom). Figure from Bickley (1982).

Since our investigation of breathiness of male and female speakers partly follows the one by Henton and Bladon [2I, we take the amplitude relationship between the two lowest harmonics as an operational means to define breathiness: H1 - H2 or DH (delta H). In addition, this acoustic cue seems to correlate fairly well with listeners' judgements about the breathiness of perceived vowels.

2. OCCURRENCE OF BREATHINESS AND EXPERIMENTAL OUTLINE

In numerous languages breathiness is used to form phonemic contrasts (for references see [2I). These languages, however, are not our present area of interest, nor is breathiness as a marker of pathological speech. Evidence has been adduced (ibid) that female speakers of two British accents consistently used a more breathy voice quality than men in ordinary speech. Although breathiness may be considered an inefficient way of voice production with a number of communicative limitations, the claim has been put forward that, consciously or unconsciously, women use breathiness as a means of communicating arousal, intimacy, or, in other words, to sound more "sexy".

Our present experiment was set up to a) compare Dutch breathiness data of male and female speakers in an ordinary speaking mode those of English and b) to investigate whether in an emotive context there would be an increase in breathiness by either speaker sex.

As corollary variables to DH, average F0 and average intensity of the vowels under investigation will be considered as well.

3. EXPERIMENTAL PROCEDURE

13 male and 13 female speakers participated in the experiment. Their ages varied between 19 and 38 years. They belonged - either as staff or students - to the University of Utrecht and were speakers of the Dutch equivalent of RP.

The vowel to be analysed was decided to be an /a/ since this open vowel's first formant is high enough to be of no influence on the first harmonic. Monosyllabic words containing the vowel /a/ were embedded in unpretentious sentences which, on their part, were combined to form an unpretentious piece of running prose. Due to their monosyllabicity all stimulus words carried lexical stress. In addition a semantically intimate passage containing numerous /a/'s was selected from a sultry-romantic piece of fiction in order to simulate an emotive context.

Speakers were instructed to read the first text in an ordinary and the second text in a sexually charged way.

Recordings were made individually in a sound-proof room using a Revox B77 mkII tape recorder and a Sennheiser microphone. A mouth-to-microphone distance of 30 cm was used. The input volume control was held constant and subjects were given some practice time.

Data were further processed digitally. Per reading mode and subject 12 35 ms steady state portions of the /a/ vowels were excised and relative amplitudes of the first two harmonics, and F0 and amplitude of the steady states were established.

4. RESULTS

4.1. RELATIVE AMPLITUDE OF HARMONICS

Table I shows the average values and corresponding SD's of DH produced by male and female speakers in the two reading modes. A negative DH value indicates that the amplitude of the first harmonic is lower than that of the second harmonic and v.v. According to Bickley (1982) a negative DH value is the consequence of breathy phonation.

Figure 2 represents per speaker the DH values in the ordinary and the emotive reading mode.

From table I and figure 2 it can easily be seen that female DH values are higher than male values and that DH values of both sexes are higher in the emotive context than in the ordinary context. Statistical analysis shows the between-sexes difference to be significant in either reading mode ($p<.05$ and $p<.01$ resp.) and the between-reading mode difference to be not more than a strong tendency. Moreover it was shown that there is no significant difference in breathiness between the female-ordinary condition and the male-emotive condition which means that female speakers used the same degree of breathiness in reading the ordinary text as did male speakers in reading the emotive text.

4.2. FUNDAMENTAL FREQUENCY

Results of the F0 analysis of the measured /a/ steady states are shown in table II.
As can be seen there is a decrease in fundamental frequency for the emotive reading text for either sex; differences, however, do not reach the level of significance.

Interindividually, however, a significant positive correlation exists between F0 and DH.

Table I: Average DH in dB for ordinary and emotive reading mode.

	female speakers (n=13)		male speakers (n=13)	
	ordinary	emotive	ordinary	emotive
x	+3.9	+5.4	-0.6	+1.1
SD	5.2	4.6	3.0	2.6

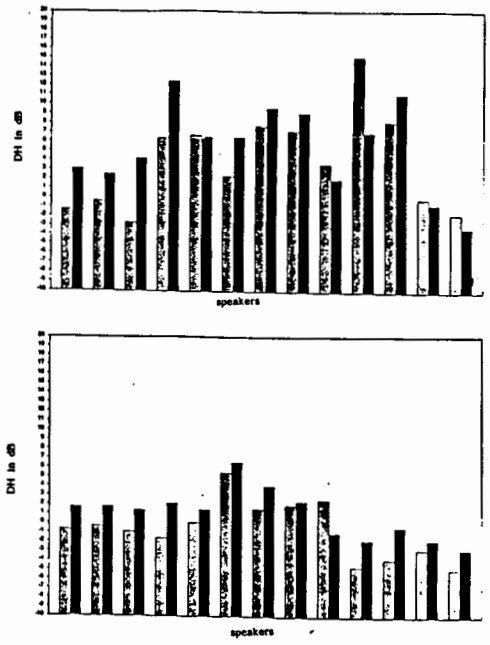


Figure 2: DH in dB for ordinary reading mode (light bars) and emotive reading mode (dark bars) for female speakers (above) and male speakers (bottom).

Table II: F0 in Hz for ordinary and emotive reading mode

	female speakers (n=13)		male speakers (n=13)	
	ordinary	emotive	ordinary	emotive
x	213	204	131	122
SD	25	25	20	18

4.3. AMPLITUDE

For either sex there is a slight decrease of amplitude values for the emotive context in comparison with the ordinary context (female: 2.9 dB; male: 1.9 dB). Differences are insignificant.

5. DISCUSSION AND CONCLUSION

Female speakers produce significantly more breathiness in comparison with male speakers in the ordinary reading mode as well as in the emotive reading mode. The degree of breathiness produced by female speakers in the ordinary context turns out to be even equivalent to that of male speakers in the emotive context. We found a decrease of both F0 values and amplitude values for the emotive context for male as well as female speakers, but these differences fail to reach the level of significance.

As stated in our introductory section, a breathy spectrum contributes to perceptual limitations. Why, Henton and Bladon [2] ask themselves, should women adopt articulatory postures that render their own speech less efficient in communicative terms? The answer to this question lies, according to these authors, in the ethological-sociolinguistic domain: "...[women] imitate the voice quality associated with arousal. ...A breathy woman can be regarded as using her paralinguistic tools to maximize the chances of her achieving her goals, linguistic or otherwise"

With all due respect we would like to regard this explanation with some caution. First of all perceptibility of speech is affected only in extreme cases of breathiness. Secondly, not only was DH in the afore-mentioned experiment the only acoustic correlate considered to indicate breathiness, whereas other parameters probably deserve consideration as well, but breathiness in its turn is certainly not the only characteristic of a "sexy" voice.

As to voice source characteristics, it is generally assumed that female speakers have a greater open quotient which implies that they produce more breathiness for physiological reasons.

In connection with our tentative F0 - DH correlation data we suggest that more research should be addressed to the question of whether a systematic relationship can be found between pitch and breathy phonation on an interindividual level.

6. REFERENCES

[1] BICKLEY, C. (1982), "Acoustic analysis and perception of breathy vowels", MIT Working Papers in Speech Communication, vol.1:73-83.
[2] HENTON, C.G. & BLADON, R.A.W. (1985), "Breathiness in normal female speech: inefficiency versus desirability", Language and Communication, vol.5,3:221-227.
[3] PANDIT, P.B. (1957), "Nasalization, aspiration and murmur in Gujarati", Indian Linguistics, vol.17:165-172.

I WISH TO THANK MY STUDENTS ASTRID SMEETS AND SONJA SENGERS FOR THEIR CONTRIBUTION TO THIS PAPER.

ARTICULATORY GENERALIZATIONS IN ACOUSTIC PHONETIC RESEARCH: A COMPARISON OF DATA FROM FRENCH AND ENGLISH

Sarah N. Dart

Department of Linguistics, UCLA, Los Angeles, U.S.A.

ABSTRACT

Simultaneous articulatory and acoustic data were recorded for 21 French speakers and 20 English speakers uttering phrases containing the coronal consonants /t,d,n,l,s,z/. It was found that, in both languages, individual variation in articulation of these consonants makes it difficult to make precise language-specific generalizations in terms of both place of articulation and apicality. The formant patterns in the acoustic signal, however, are much more homogeneous and suggest that the difference in consonant production in these two languages lies more in the general shape of the tongue body behind the constriction than in the placement of the constriction itself.

1. INTRODUCTION

The coronal consonants of French and English have been claimed to be articulated differently in terms of place of articulation and apicality. For example, French coronal stops are regularly described as dental, either with the tip of the tongue on the upper incisors (apical), or with the tip down behind the lower incisors and the blade making contact (laminal). English coronals, on the other hand are usually said to have an apical alveolar constriction. Such information forms the basis not only of foreign language pronunciation instruction, but also of acoustic phonetic analysis, where data from acoustic recordings of speakers of the same language are assumed to originate from a homogeneous set of articulations. The present study seeks to discover to what extent the articulation of an individual can be predicted by language community affiliation. Is the precise point of articulation as given on a traditional consonant chart really crucial to the pronunciation of a given language or are there other factors which are more important?

The articulatory data presented here is in the form of palatograms and linguagrams taken by the direct method to ascertain the point of contact on the upper surface of the vocal tract, as well as the part of the tongue used to make the constriction. Audio recordings were also made synchronous with the palatograms and linguagrams, in order to be certain that each given acoustic signal corresponded to an articulation with known articulatory characteristics. Data from 21 French speakers (northern standard pronunciation) and 20 English speakers (west coast American) were recorded of the consonants /t,d,n,l,s,z/ in both word-initial and word-final position, in the environment of a low vowel ([æ] in English, [a] in French).

2. ARTICULATORY DATA

Place of articulation was determined from the palatograms in the manner described in detail in Dart [2], briefly as follows: if the vertical surface of the back of the upper central incisors was contacted, either completely or partially, the articulation was called *dental*; *alveolar* articulations were those where the most forward part of the contact was in an area extending from the base of the teeth to approximately 5 mm back; and

articulations made behind this area were called *postalveolar*. The linguagraphic categories into which the data were sorted are *apical*, where only the tip and rim of the tongue were contacted; *laminal*, where only the blade made contact; and *apicolaminal*, where both the tip and blade were contacted. The fricatives were classified as either *apical* or *laminal*, depending upon whether the tip or only the blade was contacted. Table 1 below gives the results of the articulatory study.

Table 1. Percent of the total number of tokens for each place of articulation and apicality classification. A= apical, L= laminal, AL= apicolaminal.

French

	/t,d,n/			/s,z/		/l/		
dental	6.3	12.7	39.7	15.8	26.3	2.4	---	2.4
alveolar	13.5	16.7	11.1	7.9	30.3	69	2.4	---
post-alveolar	---	---	---	7.9	11.8	23.8	---	---
	A	L	AL	A	L	A	L	AL

English

	/t,d,n/			/s,z/		/l/		
dental	6.7	6.7	4.2	20	2.5	34.2	2.6	13.2
alveolar	59.6	5	12.6	22.5	31.2	31.6	---	15.8
post-alveolar	5	---	---	---	23.8	2.6	---	---
	A	L	AL	A	L	A	L	AL

It is clear from the table that the greatest number of French speakers produced an apicolaminal dental articulation for /t,d,n/. This accords with the claims in the literature that these segments are apical dental, it being difficult for a speaker with normal dentition to produce a purely apical dental, without the blade of the tongue also contacting the alveolar ridge. Some authors have also claimed tip-down laminal dental articulation for these segments and 12.7% of the data support this. There remain, however, 41% of the data left unaccounted for, that is all those tokens which are not dental. Clearly, a number of French speakers articulate farther back than was previously supposed.

The point of view of the sources consulted on fricative articulation was more open to variation, with both dental and alveolar articulations mentioned (although only one source allowed for both possibilities). Most sources, however, stated quite firmly that French /s/ and /z/ were laminal. It is clear from the table that, although the majority of tokens were indeed laminal, still nearly a third were apical, and thus not accounted for by the descriptions.

In English, 59.6% of the data for /t,d,n/ are, indeed, apical alveolar as predicted. 11.7% of the tokens are also apical, but either dental or postalveolar, and 17.6% are also alveolar, but use a different part of the tongue. A total of 17.6% of the tokens are dental and 28.5% are either laminal or apicolaminal.

The fricatives /s/ and /z/ , usually said to be either apical or laminal alveolar in English, were indeed divided between

these two ways of articulating, the laminal predominating with over half (57.5%) of the tokens. Again, most of the tokens were alveolar or postalveolar (77.5%).

As it turns out, the English laterals are far more likely to be dental than their French counterparts, going against the neat organization of the consonant charts, which usually put /t,d,n,l,s,z/ in the same column. Exactly half of the /l/ tokens were dental in the English data (as compared to 4.8% of French tokens), in spite of the general acceptance in the literature that such English segments should be alveolar, just as the French are assumed to be dental. Even the apical articulation of the lateral, which was nearly universal for the French speakers (95%) was less strong in English (68%), the quintessential "apical" language. It seems, then, that /l/ need not necessarily share the articulatory characteristics of the other coronal consonants in any given language.

The articulatory data thus shows that, although the articulation of these consonants may be predicted in a general way for the majority of speakers, the variation is such that one cannot assume an articulation to be of a certain type only on the basis of the native language of the speaker.

3. ACOUSTIC DATA

Formant transition frequencies were measured from wide band spectrograms for all tokens: for the word-initial tokens immediately after the closure, and for word-final tokens immediately before the closure. To normalize for absolute frequency differences between speakers, the difference was calculated between the transition formant values and the average steady-state formant values of the adjacent vowel. The resulting number was used for comparison rather than the raw formant frequencies. The formant values of the steady-state vowel were comparable between the two languages except for the value of the second formant, which was higher in English.

Two general differences between French and English articulation were noted: the value of the F1 transition in French was always lower in relation to the steady-state vowel than the corresponding English value for all the coronal consonants, no matter what method of articulation was used. Similarly, the transition value of F4 was always higher in French than in English. These differences suggest different tongue shapes behind the constriction in the two languages. A lower F1 could indicate a wider pharyngeal cavity and a higher F4 a smaller sublingual cavity in front for French. In addition to these general characteristics, a specific tongue shape difference between apical alveolar articulations in the two languages was inferred from the formant data, particularly in fricatives. French apical alveolar fricatives have lower transitional F1 values and higher transitional F2 values than do apical dental fricatives, whereas the reverse is true for English. Similarly, French apical consonants have higher F2 values than laminals, whereas in English F2 is higher in laminals.

One interpretation of these facts would be to posit a differently shaped tongue behind the constriction in the apical and alveolar articulations in the two languages. The F1 and F2 evidence suggests that the body of the tongue in French is high and forward during these consonants, thereby diminishing the area of the cavity directly behind the constriction and enlarging the pharyngeal cavity. The English apicals, on the other hand, would come up to the constriction from a lower and more posterior position in the mouth, thus creating a larger cavity behind the constriction and a more constricted pharynx. Both kinds of apical alveolar articulations can be seen in the x-ray literature, as exemplified by the two tracings in Figure 1. The tracing on the top is of French /s/ (after Bothorel et al. [1]) and resembles an apical alveolar tongue position like that posited for the French speakers, and the tracing on the bottom is of English /s/ (after Subtelny et al. [3]), and has a descending tongue shape as posited for the English speakers in the present study.

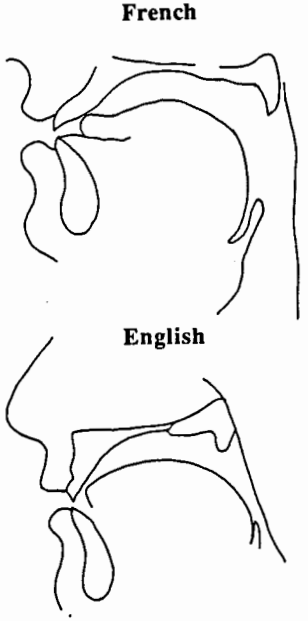


Figure 1. X-ray tracings of French (after Bothorel et al [1]) and English (after Subtelny et al [3]) showing two different tongue shapes in apical alveolar /s/.

In order to explore the possibility of such an articulatory difference as that suggested by the acoustic data, additional articulatory measurements were taken from the palatographic data in conjunction with palate casts from each speaker. It was presumed that a higher tongue position would show up on the palatograms as a wider contact area behind the constriction and, indeed, such a difference seemed to be evident from the palatograms. Accordingly, the contact area from each articulation was measured inwards from the base of the first molar and this measurement given as the ratio of the contact area on one side to half of the total distance following the curve of the palate from first molar to first molar. These measurements were shown to be significantly larger in French by one factor, repeated measures analyses of variance for all apical and alveolar stops, nasals and fricatives.

4. CONCLUSION

With the abovementioned facts taken together, there appear to be language-specific characteristics affecting the formant values, which are associated with vocal tract shapes that are not fully specified by simply characterizing the segments in terms of the articulatory contact involved. The difference between French and English coronal consonant production, rather than being one of place of articulation and apicality, would seem to be better described as a difference in the overall shape given to the tongue body in the two languages.

5. REFERENCES

[1] Bothorel, A., P. Simon, F. Wioland and J-P. Zerling (1986), "Ciné-radiographie: des voyelles et consonnes du Français", Strasbourg: Institut de Phonétique.

[2] Dart, S.N. (1991), "Articulatory and acoustic properties of apical and laminal articulations", UCLA PhD dissertation, *UCLA Working Papers in Phonetics* 79.

[3] Subtelny, J.D., N. Oya and J. Subtelny (1972), "Cineradiographic study of sibilants", *Folia Phoniatrica* 24, 30-50.

EFFECTS OF BITE-BLOCK AND LOUD SPEECH ON TONGUE HEIGHT IN THE PRODUCTION OF GERMAN VOWELS

O.-S. Bohn*, J. E. Flege, P. A. Dagenais*** and S. G. Fletcher****

*English Dept., Kiel University, Germany; ** Dept. Biocommunication, U. Alabama Birmingham, U.S.A; ***Dept. Speech Pathology and Audiology, U. South Alabama, Mobile, U. S. A.

ABSTRACT

Compensatory tongue positioning in vowel production was examined in two conditions of lower-than-normal jaw positions (bite-block speech and loud speech), and compared to a "normal" speech condition. Tongue-palate distances in multiple productions of the German vowels /i, I, u, U, y, Y/ were measured using glossometry. The tongue compensated for the lower jaw positions in both perturbation conditions. Jaw lowering in bite-block and in loud speech did not much affect the degree of precision in tongue positioning.

1. INTRODUCTION

Comparisons of normal and perturbed speech may help understand important aspects of speech motor control. Over the past twenty years, a research paradigm has become established which addresses issues such as invariance in the control of speech gestures, adaptive abilities of the speech motor system, and the role of feedback through experiments in which normal production patterns are disrupted. By examining the behavior of unperturbed articulators, the acoustic output, and/or the intelligibility of perturbed speech, studies employing this paradigm have aimed at determining if, how, and how successfully talkers reorganize articulatory gestures.

Probably the majority of perturbation studies examined the acoustic properties of vowels produced with and without the mandible being fixed in positions that required talkers to reorganize tongue gestures in order to produce intended vowel qualities. These studies have generally shown that adults [7, 8] and children [2] compensate remarkably well for a fixed jaw even before auditory feedback can occur. The small number of articulatory studies that examined tongue shapes for bite-block vowels [4, 6, 11] indicate that intended acoustic output in bite-block speech is achieved through selective compensation, i.e., by preserving "cavity configuration(s) at points of maximum constriction" [6].

Although previous research on bite-block vowels has contributed importantly to the construction and refinement of models of speech motor control, this line of research has not made it clear whether talkers aim to achieve invariance in the acoustic, perceptual, or articulatory domain.

The recently renewed interest in speech produced with loud vocal effort [9, 10] is to some extent motivated by a desire to determine the nature of talkers' "goals" or "targets". Loud speech is similar to bite-block speech in that the jaw assumes lower-than-normal positions which, however, are not artificially induced but "natural". In the only detailed study of articulatory consequences of loud speech, Schulman [9] found that the upper lip compensates for the lowered jaw in bilabial stop production, demonstrating that motor equivalence for bilabial closure occurs in both the "natural bite block condition" [9] and its artificial counterpart [1].

However, the acoustic properties of vowels produced with loud vocal effort, which have been examined in a number of studies (summarized in [10]), suggest that the analogy between loud and bite-block speech does not extend to vowel production, for the frequencies of F1 and F0 (but not usually the upper formants) are much higher in loud than in normal speech. The increase in F1 for shouted vowels led Traunmüller [10] to hypothesize that the tongue does not compensate for lower jaw positions in loud speech.

The present study, which compared tongue-palate distances for normal, bite-block, and loud vowels, was primarily motivated by the fact that only very few studies have presented direct evidence (as opposed to inferences from the acoustic output) concerning compensatory tongue positioning in bite-block vowels [4, 6, 11], and by the complete lack of published data on tongue shapes in loud vowels. Bite-block and loud vowels were compared to normal vowels to determine if and how the tongue would compensate for an artificially and a naturally lowered jaw. This study also examined variability in tongue positioning for normal, bite-block, and loud vowels. Because most earlier studies [6, 11] used x-ray techniques, which preclude detailed analyses of token-to-token variability, very little evidence exists concerning this aspect of motor control precision for the tongue in perturbed speech (but see [4]).

2. METHODS

2.1 Subject, Material, Procedure

A male native speaker of German (age: 35 years) produced 12 tokens each of the German vowels /i, I, u, U, y, Y/ in the carrier phrase ob er /bVp/ habe (blocked on vowel). The vowels were produced in three conditions. In the normal (NO) condition jaw movement was unperturbed and vocal effort was conversational (64 dB SPL). In the bite-block (BB) condition the talker's jaw was fixed in a lower-than-normal position for non-low vowels. An acrylic bite block, held between the right premolars, provided an interincisal distance of 21 mm. In the loud (LO) condition the talker produced the vowels with loud vocal effort (84 dB SPL).

Tongue-palate distances were measured using glossometry. This opto-electronic device for measuring and displaying tongue positions below the hard palate has been described previously (see [5] and references therein). Briefly, the glossometer makes use of four sensor assemblies mounted on a thin acrylic pseudopalate. Each assembly contains an LED and a phototransistor. The assemblies are positioned equidistantly along the palatal vault and are oriented perpendicularly to the occlusal plane. Sensor 1 is located just posterior of the alveolar ridge, and sensor 4 just anterior of the juncture of the hard and soft palates. Infrared light emitted from the LED is reflected from the tongue's surface, detected by the phototransistor and transduced to a voltage level. The detected voltage is approximately proportional to the inverse square of the distance of the tongue from the sensor assembly.

2.2 Data Analysis

Tongue-palate distances for tokens 2-11 for each vowel in the three conditions were measured at that point within the acoustic vowel interval that best represented the endpoint of tongue movement for each token. Endpoints were selected by visual inspection of the time-varying distance traces, which were displayed together with RMS intensity on a high-resolution graphics terminal.

Articulatory compensation with respect to tongue positioning below the hard palate was considered (by way of definition)

-*complete* if the average unsigned tongue-palate distance at the four sensor locations differed by less than 1.0 mm for NO vs. BB or LO productions of a given vowel;[1]

-*selective* if the mean tongue-palate distances in BB or LO productions at sensor locations that are near the acoustically critical maximum constriction for a given vowel were within the range of the standard deviation (SD) associated with the mean for the NO tokens at those sensor locations;

-*partial* if the tongue compensated for the lowered jaw, but did not compensate completely or selectively.

Overshoot and *undershoot* refer to partial compensation with higher-

1 Reasons for selecting this criterion to determine tongue shape overlap are given in [5].

than-normal and lower-than-normal tongue positions, respectively. Finally, in *zero compensation* the tongue does not compensate for the lowered jaw in BB and LO speech.
Variability in tongue positioning was assessed in terms of the SDs associated with the multiple productions of NO, BB, and LO vowels.

3. RESULTS

3.1 Tongue Positions

The most important result was that in the production of all six vowels, the tongue compensated for the lower-than-normal jaw position in both BB and LO speech. However, the tongue was lower in LO than BB speech at all four sensor locations for five vowels, suggesting that the tongue did not compensate as much for jaw lowering in the "natural" as in the artificial BB condition. The exception was /Y/ with overlapping tongue configurations in the BB and LO conditions.
Complete compensation by the tongue for jaw perturbation was observed in only two instances: For /i/ in the LO and for /U/ in the BB condition. Compensation was selective for /i/ in the BB condition, for /I/ in the BB and LO conditions, and for /y/ in the BB condition.
Partial compensation (undershoot) was observed for /y, Y, u, U/ in the LO and /Y/ in the BB condition. Undershoot relative to NO tongue positions, which increased monotonically from anterior to posterior sensor locations, was small for /Y/, medium for /y/ and /U/, and large for /u/. Surprisingly, undershoot for /u/ and /U/ in the LO condition was largest at sensor 4, which is located close to the acoustically critical maximum constriction for these back vowels at the velum. Results for perturbed /u/-productions differed from all other results in that undershoot in LO speech contrasted with overshoot (at the posterior sensors) in BB speech.

3.2 Variability of Tongue Positioning

The most important result concerning variability of tongue positioning in the three conditions was that perturbed vowels were not produced with uniformly more or uniformly less precise tongue gestures than NO vowels. The SDs associated with the multiple productions of the six vowels averaged 0.84 mm in the NO, 0.93 mm in the BB, and 0.77 mm in the LO condition. Tongue positioning for /i, I, y, Y/ was slightly more variable in the BB than the NO condition (SDs were 0.1 - 0.2 mm larger), but variability did not differ for /u, U/ across these conditions. Token-to-token variability was slightly larger in the LO than the NO condition for /i, I, U/ (SDs were 0.1 - 0.2 mm larger), did not differ for /Y/, and decreased for /u/ and /Y/ (by 0.3 mm and 0.6 mm, respectively).
The most conspicuous result was that for all vowels and all conditions, SDs increased monotonically from anterior to posterior sensor locations. This front-to-back increase in variability was observed irrespective of whether the acoustically critical maximum constriction was in the prepalatal (/i, I/), palatal (/y, Y/), or velar (/u, U/) region. It may be of some interest to note that tongue positioning for each of the nominally tense vowels /i, y, u/ was *more* variable than for its nominally lax counterpart (/I, Y, U/) in all three conditions.

4. DISCUSSION

The single-subject experiment reported here showed that the tongue compensated for a lowered jaw in both BB and LO speech, and that both conditions of jaw perturbation did not importantly affect the precision of motor control for the tongue.
Results of previous BB studies led to the expectation that articulatory compensation by the tongue in BB speech would be selective or complete. The present results for four (i, I, y, U/) of the six vowels examined conformed to this expectation. However, tongue positions for /Y, u/ in BB speech did not overlap with NO tongue positions or maintain NO tongue-palate distances near the acoustically critical maximum constriction. Preliminary acoustic analyses of the vowels examined in the present study indicated that partial compensation for /Y/ (undershoot) and /u/ (overshoot) did not result in changes in acoustic output that one might expect given the differences between NO and BB tongue positions below the hard palate. This suggests that compensation for the lowered jaw in the BB production of /Y, u/ may have occurred in an area of the vocal tract not registered by the glossometer.
The hypothesis being tested for LO vowels was that the tongue would not compensate for the "natural bite block". This hypothesis, which Traunmüller [10] based on the acoustic properties (increase in F1) of LO vowels, was not supported. The present experiment showed that compensation by the tongue for a lowered jaw in LO speech may be partial (y, Y, u, U/), selective (/I/), or even complete (/i/). This suggests that motor programming in *both* LO and BB speech involves reorganization of tongue positioning to achieve precisely defined articulatory goals that are not necessarily (as for /i/ in LO speech) the same as in NO speech. The lower tongue positions in LO than in NO speech for four of the six vowels examined may have been effected to increase F1, so that the perceptually important distance between F1 and the increased F0 in LO speech would be maintained for a given vowel irrespective of vocal effort (see [10]).
Degree of precision in tongue positioning did not differ much across the three conditions. The SDs associated with multiple productions of NO (0.84 mm), BB (0.93 mm), and LO (0.77 mm) vowels were of approximately the same magnitude as the mean SD for the complete set of NO German vowels (0.78 mm [3]), the complete set of NO English vowels (0.81 mm [5]), and five Spanish and English vowels spoken normally (0.76 mm) and with a BB (0.80) [4]. These earlier studies suggested that neither vowel inventory size [4] nor mechanisms used to differentiate large vowel inventories [3] affect variability of tongue positioning. The present results corroborate and extend Flege's [4] BB study by showing that both artificial and natural jaw perturbation need not importantly affect degree of precision in tongue positioning.
(Research supported by grant NS20963-04 from the U.S. National Institutes of Health to the second author)

5. REFERENCES

[1] ABBS, J., GRACCO, V. & COLE, K. (1984), "Sensorimotor contributions to the coordination of multicomponent behaviors: Evidence from recent studies of speech movement control", *J. Motor Behavior, 16*, 195-231.
[2] BAUM, S. & KATZ, W. (1988), "Acoustic analysis of compensatory articulation in children." *J. Acoustical Society of America, 84*, 1662-1668.
[3] BOHN, O., FLEGE, J., DAGENAIS, P. & FLETCHER, S. (1991), "Differenzierung und Variabilität der Zungenpositionen bei der Artikulation deutscher Vokale", *Forum Phoneticum* (in press).
[4] FLEGE, J. (1989), "Differences in inventory size affect the location but not the precision of tongue positioning in vowel production", *Language & Speech, 32*, 123-147.
[5] FLEGE, J., FLETCHER, S., McCUTCHEON, M. & SMITH, S. (1986), "The physiological specification of American English vowels", *Language & Speech, 29*, 361-388.
[6] GAY, T., LINDBLOM, B. & LUBKER, J. (1981), "Production of bite-block vowels: Acoustic equivalence by selective compensation", *J. Acoustical Society of America, 69*, 802-810.
[7] LINDBLOM, B., LUBKER, J. & GAY, T. (1979), "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", *J. Phonetics, 7*, 147-161.
[8] LINDBLOM, B. & SUNDBERG, J. (1971), "Acoustical consequences of lip, tongue, jaw, and larynx movement", *J. Acoustical Society of America, 50*, 1166-1179.
[9] SCHULMAN, R. (1989), "Articulatory dynamics of loud and normal speech", *J. Acoustical Society of America, 85*, 295-312.
[10] TRAUNMÜLLER, H. (1988), "Paralinguistic variation and invariance in the characteristic frequencies of vowels", *Phonetica, 45*, 1-29.
[11] TYE, N., ZIMMERMANN, G. & KELSO, J. (1983), "'Compensatory articulation' in hearing impaired speakers: A cinefluorographic study", *J. Phonetics, 11*, 101-115.

A MODEL FOR THE DISCRIMINATION OF PURE TONE PITCH.

Alain de Cheveigné

Laboratoire de Linguistique Formelle, CNRS - Université Paris 7, France.

ABSTRACT

This paper presents a model of auditory processing that can account for the very small frequency difference limens observed psychophysically for pure tones. In a first step, an autocoincidence histogram is calculated from nerve-fiber channels synchronized to the pure tone, according to a model similar to that of Licklider [3, 4, 5]. In a second step, this histogram is "folded", resulting in a "narrowed autocoincidence histogram". The peak of this narrowed histogram is sharper than that of the autocoincidence histogram, and its width depends on stimulus duration in a way similar to frequency difference limens.

1. INTRODUCTION

Listeners can discriminate differences in the frequency of pure tones as small as 0.2% [1]. Thresholds get larger as stimuli get shorter, but discrimination remains good even when the stimuli contain only a few cycles. Moore [1] argued that the thresholds are too low to be compatible with a place mechanism of frequency discrimination based on the differences in intensity that might arise when the excitation pattern for a tone is shifted along the basilar membrane. They would be compatible, on the other hand, with a time domain mechanism. Based on this assumption, Goldstein and Srulovicz [2] proposed a theory that predicts thresholds under the hypothesis of optimum processing of interspike intervals. Goldstein and Srulovicz noted that information from as few as *nine* fibers is sufficient to account for discrimination thresholds. Since many more fibers are available for processing, performance must have other limits, perhaps due to the actual neural processing mechanism. The question arises as to whether such processing has the same behavior as optimum processing. It is therefore of interest to examine candidate processing models with respect to pure tone frequency discrimination. One such model is that of Licklider [3, 4, 5], based on the autocoincidence of nerve fiber discharges (see also [6, 7, 8, 9, 10]). If we assume this particular model, can we still predict discrimination thresholds?

In this study it is found that a) the basic autocoincidence mechanism of Licklider's model does not adequately predict performance, but b) it can be followed by a second stage of processing, described by a "narrowed autocoincidence histogram" (NAC), to form a model that predicts thresholds similar to those observed psychophysically.

2. DISCRIMINATION THRESHOLDS FOR PURE TONE PITCH

Moore [1] measured frequency difference limens for pure tones as a function of frequency and stimulus duration. His data are plotted in Fig. 1. At all frequencies, thresholds tend to be smaller for longer stimuli. Discrimination gets better as frequency increases, up to 2 kHz. For the lowest three frequencies there is a zone of durations for which threshold varies approximately as the *inverse of stimulus duration*. These frequencies are in the region for which a time-domain frequency analysis mechanism such as Licklider's is in principle applicable.

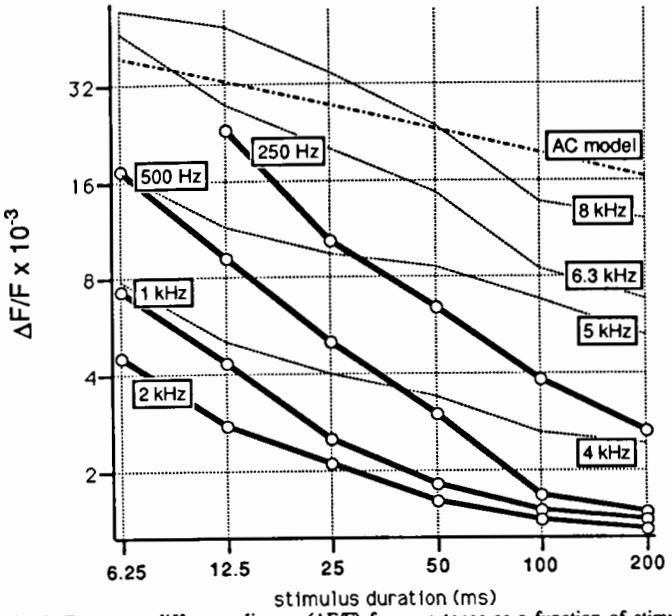


Fig. 1. Frequency difference limens (ΔF/F) for pure tones as a function of stimulus duration and frequency (replotted from Moore [1]). Frequencies up to 2 kHz are plotted with continuous lines, higher frequencies with dotted lines. Straight line: difference limen predicted by basic autocoincidence model.

3. AUTOCOINCIDENCE MODEL

In Licklider's model [3, 4, 5], patterns of discharge within auditory-nerve fibers are processed in the auditory nervous system by a neural network that calculates the equivalent of an autocoincidence (or autocorrelation) histogram [11, 12]. The result is a pattern of activity over the two dimensions of *frequency* (inherited from peripheral filtering) and *lag* (provided by nerve conduction or synaptic delays). In response to a periodic stimulus, this pattern shows a ridge at a lag equal to the period, thus providing a cue to the pitch. Licklider's model was designed to explain the pitch of complex stimuli, however it works as well for pure tones. In response to a pure tone of frequency f, nerve fibers with characteristic frequencies within a band surrounding f will respond with a periodicity of 1/f. The result is an autocoincidence pattern with a ridge at 1/f. Actually, the pattern also shows ridges at period multiples; the model supposes that the position of the *first* ridge is the cue to pitch. Because synchronization deteriorates above 2-5 kHz, the model can only apply to frequencies below that limit (this excludes the upper 2 or 3 octaves of the 10 that span the audible range).

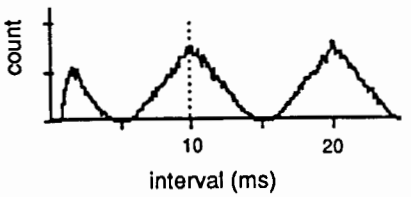


Fig. 2. Autocoincidence histogram in response to a pure tone of 100 Hz. The dotted line marks the period lag. The histogram was calculated using "spike" data produced by a model [13].

Let us define the pitch cue more precisely as the position of the *maximum* of a composite pattern obtained by taking the sum of histograms across frequency channels (alternative assumptions are possible but won't be discussed here). In response to a pure tone the histograms are all identical, so the effect of summing them is simply to reduce variability, as if a single

histogram were calculated with more spikes. How precise is this cue?
As evident in Fig. 2, the histogram is "noisy", which causes the position of the maximum to be uncertain. The standard deviation of this position can be estimated [14] as a function of discharge rate R, stimulus duration D, histogram bin width ε, and number N of histograms summed together:

$$\sigma_T \approx 0.12\ R^{-1/2}(D\varepsilon N)^{-1/4} \qquad (1)$$

It is evident from (1) that the standard deviation varies as the inverse of the *fourth root* of stimulus duration. This dependency can be understood as follows: due to the parabolic shape of the AC histogram near its peak, the incertitude of the position of the maximum varies with the *square root* of the standard deviation of the bin "noise", itself proportional to the square root of the counts in the histogram bins. If spikes are allowed to accumulate during the entire stimulus presentation, the count within each bin is proportional to duration, hence the $D^{-1/4}$ dependency.
To get a more quantitative estimate, let us make the assumption that 1250 fibers respond each at 100 s/s, that the spike trains are pooled before histogram calculation into 10 histograms that are then summed, and that histogram resolution is 1µs (R = 12500, N= 10, $\varepsilon = 10^{-6}$). Given these assumptions, the difference limen $\Delta F/F$ (supposed equal to σ_T) varies as plotted in Figure 1. We can draw the following conclusions:

a) The dependency of $\Delta F/F$ on duration, predicted by the model as $D^{-1/4}$, does not match that observed in Moore's data at low frequencies.

b) The $\Delta F/F$ predicted by the model is almost an order of magnitude larger than the best difference limens observed.

4. NARROWED AUTOCOINCIDENCE MODEL

In the AC model, the effect of making the stimulus longer is to make more spikes available, thus reducing statistical uncertainty. Clearly this is insufficient to account for the difference limens observed and their dependency on duration. There is however a source of information that the AC model neglects: that carried by the peaks of *higher rank* of the autocoincidence histogram.
Recently, a method has been proposed for sharpening the peaks of the autocorrelation *function* (for purposes of musical pitch estimation) [15]. This method incorporates information from higher-order peaks into a compact representation called "Narrowed autocorrelation function". A similar operation can be applied to the autocoincidence histogram (AC), resulting in a "Narrowed autocoincidence histogram" (NAC):

$$NAC(\tau) = \sum_{k=1}^{N-1} (N-k)\ AC(k\tau) \qquad (2)$$

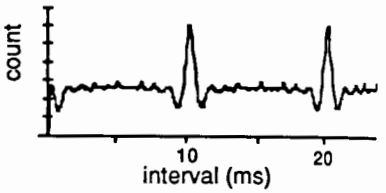


Fig. 3. Narrowed autocoincidence histogram in response to a pure tone of 100 Hz. Order of narrowing is 10.

It can be seen from Figure 3 that the period peak of the NAC is narrower than that of the AC histogram. Peak width is inversely proportional to the narrowing order N. The practical value of N is limited by the duration of the stimulus, since it is impossible to calculate an AC histogram for intervals greater than the stimulus duration. If this is the factor that limits frequency discrimination, then difference limens should vary as D^{-1}. (Whereas probabilistic factors determined the thresholds of the AC model, these factors are considered negligeable in the analysis of the NAC).

Fig. 4 displays $\Delta F/F$, under the further assumptions that the width of an AC peak before narrowing is about 10 %, and that the only effect of frequency is to vary the number of cycles within a stimulus. The effect of frequency is difficult to analyze in this model, because it affects the population of fibers that respond and their degree of synchronization, as well as the number of AC histograms peaks that can fit within a given duration.

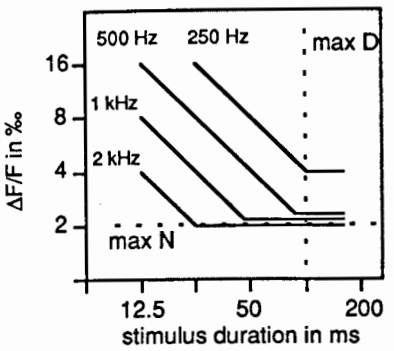


Fig. 4. Difference limens predicted by the NAC model.

The dotted lines labeled "max D" and "max N" in Fig. 4 represent additional hypothetical limits on discrimination due to two factors. The first factor limits length of AC histograms (it could be for example a limit on the allowable length of the neural delay lines assumed by Licklider's model). Making stimuli longer than this limit can bring no improvement. The second factor would limit the order of narrowing, due to the complexity of neural circuitry available for the calculation of the NAC.
The trend of the thresholds visible in Fig. 4 is similar to that of Moore's data for frequencies below 2 kHz (Fig. 1). The major differences are that the curves in Fig. 1 are somewhat shallower, and the spacing smaller than predicted by the model. There is also no evidence in Moore's data for the first of the hypothetical limits mentioned above ("max D"). Apart from these differences the agreement is quite good.

CONCLUSION

The basic autocorrelation model due to Licklider is not sufficient to account for frequency difference limens observed psychophysically. However, a modified model (the NAC model) can successfully account for these limens, and for the form of their dependency on duration. This result is of interest given the recent renewed concern for time-domain models of auditory processing.

ACKNOWLEDGEMENTS

Part of this work was carried out at ATR Interpreting Telephony Research Laboratories, under a fellowship awarded by the European Communities STP programme in Japan. The author wishes to thank ATR for its hospitality, and the CNRS for leave of absence.

REFERENCES

[1.] Moore, B. C. J. (1973), "Frequency difference limens for short-duration tones", JASA. *54*, 610-619.
[2.] Goldstein, J. L. and P. Srulovicz (1977), "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement", *Psychophysics and physiology of hearing*, Evans and Wilson ed. Academic Press: London, 337-345.
[3.] Licklider, J. C. R. (1956), "Auditory frequency analysis", *Information theory*, Cherry ed. Butterworth: London, 253-268.
[4.] Licklider, J. C. R. (1959), "Three auditory theories", *Psychology, a study of a science*, Koch ed. McGraw-Hill: 41-144.
[5.] Licklider, J. C. R. (1962), "Periodicity pitch and related auditory process models", International Audiology. *1*, 11-36.
[6.] de Cheveigné, A. (1986), "A pitch perception model", *Proc. IEEE ICASSP*, 897-900.
[7.] Lyon, R. (1984), "Computational models of neural auditory processing", *IEEE ICASSP*, 36.1.(1-4).
[8.] Meddis, R. and M. Hewitt (1988), "A computational model of low pitch judgement", *Basic issues in hearing*, Duifuis, Horst and Witt ed. Academic: London, 148-153.
[9.] Moore, B. C. J. (1982), *An introduction to the psychology of hearing*, Academic Press: London.
[10.] van Noorden, L. (1982), "Two channel pitch perception", *Music, mind, and brain*, Clynes ed. Plenum Press: London, 251-269.
[11.] Ruggero, M. A. (1973), "Response to noise of auditory nerve fibers in the squirrel monkey", J. Neurophysiol. *36*, 569-587.
[12.] Evans, E. F. (1983), "Pitch and cochlear nerve fiber discharge patterns", *Auditory frequency analysis*, Moore and Patterson ed. Plenum Press: 253-264.
[13.] de Cheveigné, A. (1990), "Auditory nerve fiber spike generation model.", ATR technical report TR-I-104, 15p.
[14.] de Cheveigné, A. (1989), "The narrowed autocoincidence histogram and pure tone pitch.", submitted for publication.
[15.] Brown, J. C. and M. S. Puckette (1989), "Calculation of a "narrowed" autocorrelation function", JASA. *85*, 1595-1601.

PERCPTUAL SENSE UNITS IN THE PROCESS OF LISTENING COMPREHENSION

Morio Kohno

Kobe City University of Foreign Studies
Kobe, Japan.

ABSTRACT

Research was added to the previous studies which made clear that the perception of sound sequence consists of two-fold processings--holistic and analytic, and that the former is applied to the fast tempo sequences whose intersound intervals are within 300ms, and the latter to the slow ones whose intervals are over 400ms. They are neuropsychologically different from each other (See another paper of the author in this conference named "Two Processing Mechanisms in Rhythm Perception.") Experiments I and II proved that in the process of listening comprehension, meaning units which are made up of 1 to 7±2 syllables closely combined each other with the intervals of less than 300ms are perceived holistically as definite units. If holisticality is deprived semantically and/or physically by pausing, the listenability is extraordinarily decreased. There is some evidence that this 'perceptual sense unit' was imbedded in human-beings' deep cognitive system.

The present paper is going to try to identify 'perceptual sense unit'(P-unit, henceforth), perceived holistically and stored in echoic memory in an unprocessed form in the process of listening comprehension.

1. PREVIOUS STUDIES

Kohno and Kashiwagi [3] made clear the processing mechanism in rhythm perception using as subjects a normal right-handed woman, seven children with age variety from one year and four months to nine years old and a patient with infarction involving the forebrain commissural fibers. The summary of the results got by the study is as follows.
1) The left hand of the patient and children under four years of age can not synchronize their tapping with the slow rhythms of 500 and 1000ms inter-beat intervals (IBI), but they can follow the rapid stimuli of 250 ms IBI. The right hand of the patient and children older than four years old, however, can fit their tapping both to the rapid and slow rhythms as well as the normal adult. 3) Negative autocorrelations were detected among adjacent IBIs in slow response beats by the normal adult, the children older than four and the right hand of the patient, but never found in the responses to the slow stimuli by the children younger than four and the left hand of the patient. 4) Negative autocorrelations were never detected in the rapid response movements (250ms) of all the subjects. 5) The above-mentioned facts suggest that the slow repetitive sound sequences are normally perceived by ongoing and analytic way of processng, but rapid ones by at-a-time and holistic way. Evidence was found that the children younger than four years old and the left hand of the patient always use only the holistic approach not only for the rapid rhythm but also for the slow rhythm and that it is the very reason why they cannot synchronize the slow tempos. These two kinds of processing, therefore, are neuropsychologically different from each other. 6) Other experiments using nonsense words on timing condition of syllable sequences and echoic memory were held with the following results. a) The sequences of closely connected nonsense syllables, each of whose inter-voice-onset-piont intervals(IVI, henceforth) are less than 300ms, bring forth longer retention than the syllable sequences whose IVIs are longer than 400ms, when they are recalled after some lapse of time for doing two digit number multiplication. As suggested in 5), the durational condition less than 300ms may be processed holistically, and that more than 400ms, analytically. As holistic processing is qualitatively different from analytic one, the former is never disturbed by the latter, and this may be the reason why the words whose syllables are closely connected have longer retention than the ones which consist of losely connected syllables after doing some cognitive work. b) There is some evidence that durational condition between 300 and 400ms IVIs is border area mixed with both holistic and analytic processings, different by individuals.

2. IDENTIFICATION OF PERCEPTUAL SENSE UNIT

2.1. Experiment I

[Subjects] Students of a high school in Japan, 108 in number, were divided into four homogeneous groups in their English ability based on their academic records .

[Materials] Two original stories, one in English (95 words), the other, in Japanese (133 words), were recorded by an American instructor and a Japanese one respectively. Then pauses were mechanically placed by the use of Pause Controller, SONY LLC 5000, at every end of word (the set pause length = 1 second), at every end of phrase, of clause and of sentence(the set pause length = 2 seconds). Phrase here means a meaning unit which consists of one content and no or some function words. A no-pause version was also prepared. An important thing here is that the number of syllables which made up a phrase was 1 to 7 in English version and 2 to 8 in Japanese. The unit of clause, however, consisted of 2 to 13 syllables in English, and 7 to 25 (8 to 26 morae) in Japanese, and the sentence unit, 12 to 22 in English and 12 to 48 (15 to 53 morae) in Japanese. According to Miller [4], maximum number of elements which can be perceived in a flash is 7±2. Only the unit of phrase out of the above units meets Miller's condition --in the case of English clause, for example, half of them consisted of more than 7 syllables. Words, of course, consist of less than 7 syllables, but the separation by this unit destroys the unit of meaning, because no function word has any independent meaning.

[Method] Each subject group was requested to listen to one version in Japanese and another in English, and were asked to write the content of the stories as precisely as possible in Japanese, immediately after they had finished hearing them.

Table 1 - A (Japanese) full marks = 29

pause	n.	$\bar{x}$	S.D.
no pause(A)	27	4.2	1.78
every clause(B)	26	9.7	5.27
every phrase(C)	24	15.5	5.65
every word(D)	31	11.8	5.82

B≒D ($t=1.39$, N.S.); C>D ($t=2.39$, $P<0.05$)
C>B ($t=3.79$, $P<0.01$) ; C>A ($t=9.75$, $P<0.001$)
B>A ($t=2.85$, $P<0.02$) ; D>A ($t=5.41$, $P<0.001$)
(Table 1 - B (English) omitted.)

[Results] Pauses which were placed at every end of clauses, and of phrases increased the scores in this order in both English and Japanese materials--the no-pause version produced the worst ones. As found in [2], the pause-at-every-phrase version brought about the highest scores. This 'the more pauses, the higher score' principle, however, did not go on in the pause-at-every-word versions, which remarkably reduced the scores($p<0.05$-0.001). This fact also coincides with the result of [3]. If pauses would play a crucial role for listening comprehension by giving chances to analyze and synthesize the stimuli and by giving clues to separatethe sound stream into proper units as pointed out in Pimsleur(1971), we may rightly say that these results suggest

the unit of phrase, more precisely, grammatical meaning unit which consists of 7±2 syllables, might render a most suitable chunk for listeners' cognitive processing of connected speech.

We should now notice that the P-unit, the unit of phrase, for example, generally consists of several syllables which are combined one to another with IVIs about 100 to 200ms in the case of Japanese or 100 to 250ms in English, all of which are so rapid as to be processed holistically. The last IVIs of each P-unit in English, however, are somewhat longer than the preceding ones. Here is an example of realities of IVIs together with syllable lengths, that is, computational observations (ILS, Micro PDP 11/73) on the syllable durations and IVIs between syllables in the stories read by native speakers.

a) Duration of syllables (syl. dur.) and IVIs in spoken Japanese sentences.

	Tsu	ki	ga /	no	bo	ri/
syl.dur.	118	104	96	108	111	115
IVI		※	146	143	121	116

	ha	ji	me	ma	shi	ta./
syl.dur.	139	139	163	165	192	158
IVI		169	111	143	(117)	(185)

b) Duration of syllables (syl. dur.) and IVIs in spoken English sentences

	Scott/	came	out	of	the
syl.dur.	331	225	242	115	196
IVI		386	225	221	165	148

	house/	and	locked	/	the
syl.dur.	311	122	345	(183)	168
IVI			232		290

	door /	behind	him.
syl.dur.	329	162	308 345
IVI	170	287	198 323

※ = boundary of p-unit)

In Japanse, all IVIs, as well as syllable durations, are all within 300 ms. This means that the syllable sequences so closely connected have to be processed holistically. In English, however, the states of IVIs and syllable durations vary very much, and generally they are longer than Japanses syllables. We can notice, however, that in English, a long IVI is put at each P-unit, especially at the end of it, that is, at each semantic unit which usually consists of 7 or less syllables. The fore parts of the unit are composed by a syllable succession with short IVIs which may be processed holistically. To put the long syllables at the ends of units is very effective to show the terminations of the units. This device seems to help listeners out of difficulty of holistic processing caused by the variety of syllable durations.

We can now more precisely say that a perceptual sense unit, is a semantic unit which is composed of one to several syllables which are so closely combined one to another in less than 300ms IVIs so that the unit can be processed holistically.

If the unity and, therefore, holisticness are lost by the long IVIs of more than 500ms, listenability of the utterance will be remarkably decreased. The analytic processing, on the other hand, may concern the processing of two and more perceptual sense units one by one to get the whole meaning of utterances. It might, as a matter of course, take a longer time.

2.3. Experiment Ⅲ

In order to verify the above hypothesis, the following experiment was carried out.

[Subjects and Method] An essay in Japanese (Material A) was read by a Japanese female instructor (age : twenties). It was then mechanically separated at every end of perceptual sense units by pauses whose durations were 3, 4 and 5 seconds. Another Japanese essay(Material B) was also read by the same instructer, but IVIs among syllables were spread to each of 200, 250, and 500ms. Subjects were Japanese high school students, 122 in number. Other procedures were the same as in Experiment Ⅰ.

[Results and Discussion] The results are shown in Tables 2-A and -B. When the IVI among syllables was 200 or 250ms which will be processed holistically, the scores remained almost the same, but when they were lengthened to 500ms, which may be processed analytically, the scores significantly decreased(Table 2-A).

Table 6 - A

	Intervals among syllables		
	200ms	250ms	500ms
n.	27	34	26
$\bar{x}$	14.52	13.03	11.27
S.D.	2.49	3.91	4.52

200ms ≒ 250ms N.S.; 200ms > 500ms $P < 0.01$

Table 6 - B

	Intervals among phrases			
	natural	3000ms	4000ms	5000ms
n.	25	26	34	27
$\bar{x}$	15.88	16.07	16.38	16.11
S.D.	2.09	2.00	1.67	1.95

NS among factors

In connection with this, [4] pointed out that extremely slow English pronunciation in which each syllable is drawled more than 500ms ($\bar{x}$=514 actually) also decreases listenability very much ($p<0.05$-0.01), even though pronunciation of segments was clear and precise. Table B, on the other hand, shows that long intervals among the perceptual sense units never bring about negative effects, perhaps because the work to be done here would be an analytic task to look for grammatical and semantic relations among several units, which by nature requires lots of time.

3. DISCUSSION

We can now exactly explain the results of Experiment Ⅰ: the no-pause version brought about the worst scores, because all the chances to perform the work of analytic processing were deprived. The pause-at-every-word version also produced very bad scores, because the unity of meaning is lost. The low listenability of drawled pronunciation [3] can be explained in the same way. Perceptual sense unit is a unit in which 'a unit of meaning' is closely attached to a closely connected syllable succession. Listeners, therefore, when they hear the unit, can reflectively recall the meaning to their mind. In the pause-at-every-word versions, however, function words which are put separately by pauses whose durations are beyond holistically perceived time intervals, prevent listeners from reflexive, at-a-time recalling of a unit of meaning. The relatively low scores produced by the pause-at-every-sentence version and by the pause-at-every-clause version can be explained by the fact that the number of syllables composing those units go beyond the suitable number that human beings can holistically process a ta time, that is, 7±2 .

4. CONCLUSION

We may conclude by saying that the processing of listening comprehension is a mix of both holistic and analytic works. The existence of the unit which should be holistically dealt with at a time has long been overlooked, but it is crucial element to make clear the processing of listening comprehension which listeners can do very efficiently and rapidly. We should notice that perceptual units whose syllables are closely connected may be preserved in an unprocessed form longer than the separately lined syllables after the execution of some cognitive works[3]. On account of this nature, listeners can do the works of semantic and sometimes grammatical analysis over several units, if necessary, referring back to some precedent unit which is still retained even after having processed some units.

REFERENCES

[1] KOHNO, M. (1981), "Effect of pau singin listening comprehension." T. Konishied. Studies in Grammar and Language, 392-405, Kenkyusha, Tokyo.

[2] KOHNO, M. Kashiwagi, A. and Kashiwagi, T. (1990), "Similarity of rhythms produced by young children and the patient with infarction involving the corpus callosum." AILA 1990.

[3] MILLER, G. (1956), "The magical num-ber seven, plus or minus two." Psychological Review, 63-2, 81-97.

METHODS FOR REDUCING CONTEXT EFFECTS IN THE SUBJECTIVE ASSESSMENT OF SYNTHETIC SPEECH

Chaslav V. Pavlovic, Mario Rossi, and Robert Espesser

Institut de Phonetique, LA 261 CNRS, Université de Provence,
Aix en Provence, FRANCE
& (1st. author only) University of Iowa, Iowa City, Iowa, USA.

ABSTRACT

The contextual invariance of categorical and magnitude estimates of speech quality could be improved by introducing a reference system (natural speech) and by appropriately normalizing the results with respect to it.

1. INTRODUCTION

A potential problem with subjective scaling of speech quality occurs when the rating of a certain system needs to be generalized outside the set of systems used in the experiment. Namely, the rating of a system may change depending on the selection of other systems evaluated at the same time ("context effect"). We evaluate here whether the context effects could be reduced by introducing a reference system (natural undistorted speech) common to all experiments, and by normalizing the rating of any given synthesizer in reference to the rating of the natural speech. Two subjective psychophysical techniques are evaluated: magnitude estimations (MEs) and categorical estimations (CEs).

The ratings of four systems labeled "A" and four systems labeled "B" were evaluated in two different types of context: "A and B" context and "A or B" context. Systems A were of superior quality to systems B. Both systems A and systems B were evaluated separately within their groups (A or B context), and together (A and B context). The research question is whether the ratings of the stimuli are invariant to these changes in context, both in the absolute and in the relative sense. These context effects were evaluated both with and without the reference condition. This particular design was selected because past research indicates that all scaling techniques may be particularly sensitive to it. It is hypothesized that subjects always use one restricted range of numbers regardless of the stimuli being evaluated. If this were indeed the case, there would be a strong tendency to use the same range of numbers for systems A only, systems B only, and systems A and B together. Given that systems A are superior in quality to systems B, the ratings of B will, therefore, be better when these systems are presented alone than together with A. The opposite would be true of systems A.

2. METHOD

The subjects were equally divided into 12 experimental groups. Six experimental groups gave ME and the other six CE judgments. The groups are identified by letters that correspond to the listening conditions they were exposed to. These six labels are ABR, AR, BR, AB, A, and B. Symbol A signifies that the group judged conditions A, symbol B that the group judged conditions B, and symbol R that the group judged the reference condition. The non-normalized group results for each condition were calculated as the means across subjects and condition repetitions. The arithmetic means were used for CEs, while the geometric means were used for MEs. Neither for the MEs nor for the CEs was the reference condition explicitly defined to the subject as such. Rather, it was treated as just another experimental condition. The subjects were required to judge how satisfied they were with the particular communication situation. For CEs the scale from 1 to 20 was used. Direct ME procedure and the sentence test material described in more detail in [1] were used.

3. RESULTS

In the tasks which did not incorporate the reference stimulus, relatively large AB context effects were seen (Fig. 1 for CEs; Fig. 2 for MEs). They seemed to be particularly severe in the case of CEs, where the mean rating of systems A and B were almost equal to each other in the "A or B" context, but quite different in the "A and B" context. When the reference condition was present, a large decrease in the AB context effect was seen in the CE (Fig. 3), while no improvement was demonstrated in the ME (Fig. 4). The introduction of the reference condition did not seem to have affected the relative ratings of the other systems neither for the MEs (Fig. 5), nor for the CEs. This indicates that some form of normalization may prove beneficial with regards to context effects.

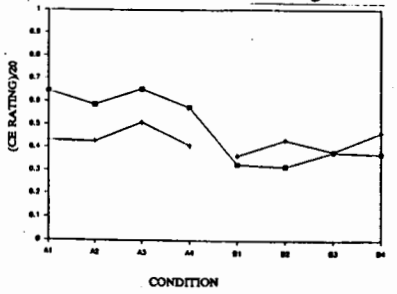


Fig. 1 CE ratings four groups AB (squares), A (pluses), and B (diamonds).

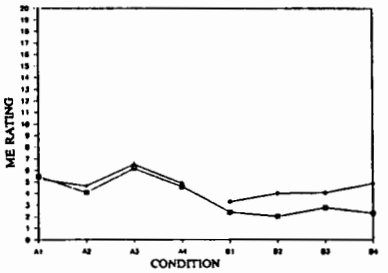


Fig. 2 ME ratings four groups AB (squares), A (pluses), and B (diamonds).

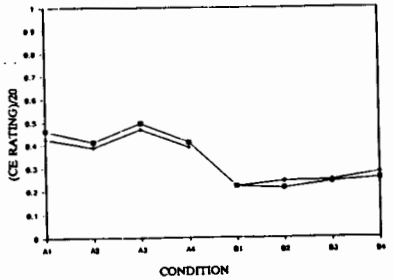


Fig. 3 CE ratings four groups ABR (squares), AR (pluses), and BR (diamonds).

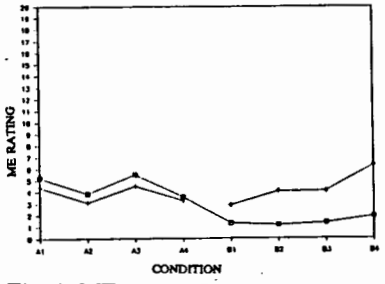


Fig. 4 ME ratings four groups ABR (squares), AR (pluses), and BR (diamonds).

4. NORMALIZATION

Two measures of the merits of normalization were used. These are the standard deviation (s), and the corre-

lation (r) between the ratings of the eight experimental systems (A and B) observed, on one hand, in the "A and B" context, and on the other hand, in the "A or B" context. Measure s expresses the absolute proximity of the measurements made in the two contexts. Measure r is sensitive to how well relative ratings of the systems agree in various contexts. The smaller the s and the larger the r the more context-free the procedure is.

The application of measure s presumes that all results are on the same scale. This is indeed the case for all normalized values. This is also the case for the non-normalized CEs that are divided by the maximum scale value. However, in the case of the non-normalized MEs the scales are arbitrary and cannot be transformed to a 0 to 1 range. In the latter case, instead of s, the measure labeled s' was used. It is defined as s divided by the mean rating of the stimuli in the "A and B" context.

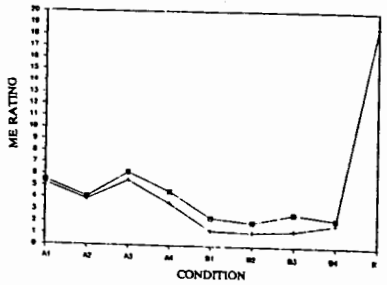


Fig. 5 ME ratings four groups AB (squares) and ABR (pluses).

The CE procedure typically results in an interval-type scale. Therefore, it is invariant to multiplication by a constant, or to addition of a constant. Thus, the results could be normalized by either of these operations. In addition, normalization could be performed on the group results, or on the results of individual subjects. In the case of normalization by multiplication, the rating of a stimulus is multiplied by the reciprocal of the rating of the reference stimulus. This operation applied to the mean group results is labeled "CE_MG," where M stands for "multiplication," and G for "group." Normalization by multiplication applied to the results of individual subjects is labeled "CE_MI," where I stands for "individual." In normalizing results by adding a constant, first the complement to 20 (maximum scale value) of the reference stimulus rating is added to the non-normalized value of the stimulus. Subsequently, these numbers are divided by 20. This procedure leads to the same results regardless of whether it is applied to the group or to the individual results. It will be labeled "CE_C," where C stands for "complement."

The measures of context effect s and r for CEs are given in Table I. All normalization procedures substantially reduced context effects with respect to the non-normalized results of the groups that did not judge the reference system. For example, in the case of method CE_MG correlation-type measure r increased from 0.48 (for the non-normalized results) to 0.98 (for the normalized results), while s decreased from 0.13 (for the non-normalized results) to 0.05 (for the normalized results). However, with respect to the non-normalized results obtained by the groups that judged the reference condition, the context effect was made somewhat worse with normalization.

The ME procedure results in a ratio-type scale, and is invariant to multiplication by a constant. Consequently the normalized results are obtained if the ratings of stimuli are multiplied by the reciprocal of the rating of the reference system. As was the case with CEs, this operation could be performed either on the group results, or on the individual subjects' results. In addition, the ME results could be calculated as "absolute" or "relative" [1]. The normalization procedures on the absolute group results is labeled "ME_AG" (symbols A and G represent "absolute" and "group," respectively), while the normalization procedure on the absolute individual results is labeled "ME_AI" (symbol I stands for "individual"). The normalization procedures either on the group or individual relative ratings yield the same values which are labeled "ME_R" (R stands for relative).

The measures of context effect r, s (if meaningful), and s' are given in Table II for these three normalization procedures, as well as for non-normalized results. Fig. 6 gives normalized MEs for the best of these procedures, i.e. ME_AI. All normalization procedures substantially reduce the context effects with respect to the non-normalized results of the groups that judged the reference system. However, the real benefit of normalization should be assessed against the non-normalized results obtained without the reference system. There, only the procedure ME_AI appears to reduce the context effects.

TABLE I. CONTEXT EFFECT MEASURES R AND S FOR CE.

PROCEDURE	R	S
CE (NON-NORMALIZED) (WITHOUT R)	0.48	0.13
CE (NON-NORMALIZED) (WITH R)	0.99	0.02
CE_MG	0.98	0.05
CE_MI	0.95	0.06
CE_C	0.88	0.08

TABLE II. CONTEXT EFFECT MEASURES R AND S, AND S' FOR ME

PROCEDURE	R	S	S'
ME (NON-NORMALIZED) (WITHOUT R)	0.84		0.35
ME (NON-NORMALIZED) (WITH R)	0.01		0.75
ME_AG	0.79	0.06	0.40
ME_AI	0.89	0.06	0.24
ME_R	0.66	0.08	0.47

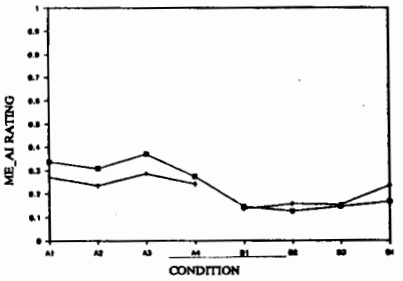


Fig. 6 Normalized ME ratings, method ME_AI, groups ABR, AR, BR

In the s value the best ME procedure (ME_AI) is practically equal (s = 0.6) to the best normalized CE procedures (CE_MG, CE_MI), but inferior to the non-normalized CE procedure when the reference stimulus is used (s = 0.2). In the r values the procedure is worse (r = 0.89) than both, better normalized CE procedures (r = 0.95 to 0.98), or the non-normalized CE procedure when the reference stimulus was presented (r = 0.99).

5. ACKNOWLEDGMENTS

This research was made possible by a grant from the EEC Esprit SAM project (Grant #2589).

6. REFERENCES

1] Pavlovic, C.V., Rossi, M., and Espesser, R. (1990). "Use of the magnitude estimation technique for assessing the performance of text-to-speech synthesis systems," J.Acoust.Soc.Am. 87, 373-382.

ORDER EFFECT AND THE ORDER OF ACCENTS

L. Schiefer and A. Batliner

Institut für Phonetik und sprachliche Kommunikation,
Universität München, München, F.R.G.

ABSTRACT

The order effect causes in a "same-different" task the one presentation order to be better discriminated than the reverse order. The effect was investigated in the domain of pitch perception. Phonetic/psychoacoustic explanations are given, and parallels between the order effect and the perception of accents are discussed.

1. INTRODUCTION

The order effect (OE) has been known for more than 100 years in the field of psychoacoustics [5]; it causes in a "same-different" discrimination task (AX-paradigm) the one presentation order AB to be better discriminated than the reverse order BA. We will call the order that is discriminated better the "prominent" order and the stimulus that comes second in this order the "prominent" stimulus. In phonetics, the OE has not been dealt with very often. This might be due to the experimental design mostly used in phonetics - the ABX-task. Originally, we came across the OE in pitch perception while investigating the categorical perception of intonation contours with the AX-paradigm [4,6]. The "potbelly"-phenomenon described in part 2 was point of departure for several experiments, where we addressed the following questions:

(i) Can the OE be influenced by the experimental design?

(ii) What causes a specific order to be a prominent one?

(iii) Can the OE be traced back to general psychophysical factors?

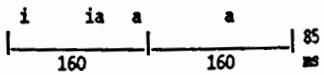
(iv) Is the OE an experimental artifact, or can it be found in real life as well?

In this paper, only a sketchy discussion of our research can be given. A thorough presentation of experiments and phonetic considerations (discussion of the state of the art) can be found in [4].

2. THE POTBELLY PHENOMENON

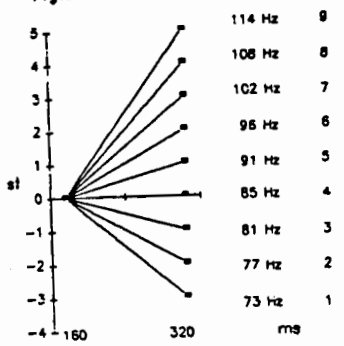
One of the authors (A.B.) produced the stimulus *ja* monotonously. The digitized stimulus (sample rate 20 kHz, cut off frequency 8 kHz) was segmented into single pitch periods. The intensity of the whole stimulus was left unchanged. The second part of the stimulus was subjected to different manipulations of the Fo contour (cf. fig.1).

Fig. 1: Segmental and durational structure



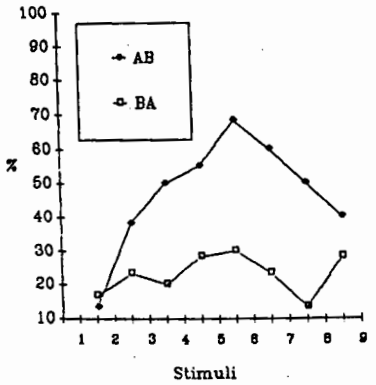
The continuum consisted of nine stimuli with a constant overall duration, three falls, one level and five rises. The duration of the manipulated part was kept constant, Fo offset and Fo slope differed. A logarithmic scale was used for the manipulation of the fundamental frequency (Fo): *semitone = 17.31*ln(Hz)*. The step from one offset height to the other was one semitone (cf. fig.2).

Fig.2: Continuum



Five repetitions of each pair (i.e. AB, BA, and the "same" order AA and BB, resp.) were presented in randomized order with an interstimulus intervall of 500 ms between the members of a pair. The pairs were separated by a pause of 3500 ms; after 10 pairs, a pause of 10 sec followed. The 12 subjects (students) were instructed to decide whether the two members of a pair were identical ("same") or different. The results are given in Fig.3. With this "potbelly shape"

Fig.3: Discrimination



function, a clear OE could be found; the order AB can be discriminated better than the order BA. The overall OE is consistent and significant in an analysis of variance, F = 60.67**. The prominent order shows a higher Fo offset in the second member of the pair.

In several other experiments, the factors duration of Fo contour, height of Fo offset, and slope were varied systematically, as well as the experimental design. The results of these experiments [4] lead to the following conclusions:

(i) The OE is no random effect, as it could be replicated in all experiments.

(ii) The OE is not an experimental artifact that can be traced back to a special design.

(iii) A stimulus is more prominent if it has a higher Fo offset and/or a longer Fo contour.

(iv) A stimulus pair is better discriminated if the prominent stimulus comes second.

3. A PHONETIC/PSYCHOACOUSTIC EXPLANATION

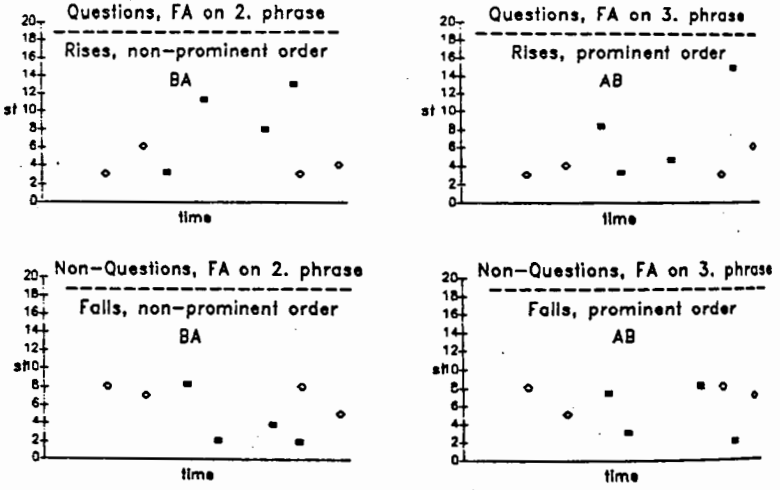
The prominence of a stimulus can be explained articulatorily and auditorily: We can assume that in production, greater pitch intervals are always connected with greater durations, and vice versa, greater durations of pitch elevations or pitch drops are related to a greater amount of pitch change. The perceptual effect of a higher Fo offset might be equal to that of a longer duration of a Fo contour, as both factors are normally interrelated. In our experiments, however, a longer lasting elevation of Fo (longer duration) does not lead to a higher Fo offset, as both factors were handled independently. At any rate, subjects seem to perceive a higher Fo offset, if the Fo contour is longer and, vice versa, a lower Fo offset if the Fo contour is shorter. The prominence of a stimulus might be caused by a greater effort in the production, i.e. a higher muscular tension needed to achieve a steeper rising or falling Fo contour and a higher or lower Fo offset as well. The prominence of a stimulus can thus be explained by articulatory and/or physiological mechanisms. But why

does the prominent stimulus come second in the prominent order? At evaluation time, the Fo information of stimulus A is still kept in memory, but it is influenced by the Fo information of stimulus B. If we substitute "weakened" for "influenced", then the prominent order can be explained: the auditory trace of stimulus A is weakened by stimulus B.

4. ORDER EFFECT AND PROMINENCE OF ACCENTS

There is at least one task for the "normal" native speaker/hearer that is comparable to the task of our subjects and that he/she has to accomplish in everyday conversation: to decide which of the phrases in an utterance carries the focal accent (FA) and thereby the "new" information. In [2,3], we investigated the acoustic structure of the FA in German. The material consisted of 360 utterances, spoken by six untrained speakers (3 male, 3 female). In these sentences, the last two phrases could be stressed depending on the surrounding context. The sentences formed minimal pairs that could only be differentiated by their intonational form: FA in final (3rd) vs. FA in prefinal (2nd) position, on the one hand, and questions (Qs) vs. non-questions (NQs), on the other hand [3:210]. In perception experiments the position of the FA was decided upon [3:211]. The task of the listeners is comparable to that in a "same-different"-task: No contextual information whatsoever is given; if we equate the two phrases that can carry the FA (2nd and 3rd phrase) with the two stimuli in the AX-task, then in both cases, the order can be "non-prominent followed by prominent stimulus", or the other way round.

Fig.4: Overlay plot



In fig.4, a sort of overlay plot is shown; the mean values of the Fo maxima and minima (full square) and their position on the time axis in the FA material (y-axis: semitones above speaker-specific lowest Fo value, x-axis: centiseconds) is compared with a schematic description of the order AB vs. the order BA (open circle). In some aspects, the OE material and the FA material cannot be compared in the strict sense. (The "turning point" in the OE material e.g. was fixed on 84 Hz, whereas in the FA material, it could be varied by the speakers.) A thorough discussion of differences and points of comparison is beyond the limits of this paper; we will therefore confine ourselves to one of the possible explanations (i.e. not the whole truth, but a substantial part of it). As for the Q/FA constellation and the OE rises in fig.4, the point of comparison is the more pronounced rise on the prominent stimulus/phrase. The prominent order AB, where the prominent stimulus comes second, corresponds to a FA on the third (last) phrase.

As for the falls, a discrepancy between the OE material and the FA material (NQ) can be observed. In the FA material, the more pronounced fall is on the phrase that carries the FA, but in the prominent order AB, the prominent stimulus B has a less pronounced fall than the non-prominent stimulus A. We believe that a solution can be found if we take the two stimuli that follow each other (*ja-ja*) not only as two acoustic or "purely" phonetic (i.e. auditory/articulatory) events but as some linguistic "gestalt" analogous to an utterance produced by a "normal" native speaker. If we imagine a (speech specific) declination line (for the sake of the argument, an all point regression line) then, in the case of the FA on the 2nd phrase and the order BA, the declination line is steeper than in the case of the FA on the 3rd phrase and the order AB. Ceteris paribus, a rather flat declination line indicates openess and/or prominence on the final part of the utterance. (Note that we do not necessarily plead in favor of a declination line as the decisive "underlying entity"; it merely seems to be the most convenient way to sum up the traits in common.)

5. FINAL DISCUSSION

We have found that one order can be better discriminated than the other one; this was called the "prominent order". Phonetic/psychoacoustic reasoning lead us to the conclusion that in the prominent order, the second stimulus is more prominent than the first one. The concept of "prominence" is the link to the marking of the FA in natural speech. The Fo contour of the prominent stimulus in the OE material can be compared with the Fo contour of the FA of the third phrase in the natural material. As for the rises, the interpretation is straightforward. Phonetic, linguistic, and psychoacoustic factors cannot be told apart. For the falls, some additional assumptions have to be made that can be summarized under the heading "perception of linguistic gestalt".

6. REFERENCES

[1] ALTMANN, H./BATLINER, A./OPPENRIEDER, W. (eds.) (1989): *"Zur Intonation von Modus und Fokus im Deutschen"*. Tübingen .

[2] BATLINER, A. (1989): "Fokus, Modus und die große Zahl. Zur intonatorischen Indizierung des Fokus im Deutschen." In: ALTMANN, H./BATLINER, A./OPPENRIEDER, W. (eds.): 21-70.

[3] BATLINER, A./NÖTH, E. (1989): "The prediction of focus." In: Tubach, J.P./Mariani, J.J. (eds.): *Eurospeech 89*. Paris: 210-213.

[4] BATLINER, A./SCHIEFER, L. (1990): "The order effect in pitch discrimination - a speech or a non-speech phenomenon?" To appear.

[5] FECHNER, G. TH. (1860): *"Elemente der Psychophysik"*. Leipzig. [1964 reprinted. Amsterdam].

[6] SCHIEFER, L./BATLINER, A. (1988): "Intonation, Ordnungseffekt und das Paradigma der Kategorialen Wahrnehmung." In: ALTMANN, H. (ed.): *Intonationsforschungen*. Tübingen: 273-291.

CATEGORICAL, PROTOTYPICAL AND GRADIENT THEORIES OF SPEECH: REACTION TIME DATA

D. H. Whalen

Haskins Laboratories

ABSTRACT

How do we make phonetic decisions? Categorical, prototypical, and gradient theories were tested using the times to identify a /sa/-/sta/ continuum created by inserting varying amounts of silence into a /sa/ syllable or deleting silence from a /sta/. The gradient model requires 6-8 times as many parameters as the others, and so is difficult to compare. Two variants of a prototypical model and a simple categorical one accounted for some of the variance in the reaction times, but a modified categorical model with the same number of parameters accounts for more. In identification, it seems that all unambiguous syllables elicit identical reaction times, but syllables farther from that range elicit increasingly longer times.

1. INTRODUCTION

When we listen to speech, we are exposed to a great deal of variation in the acoustic waveform, much of which we accept with ease. How is it that we can hear these unique acoustic events and yet extract a few categories from them? Early studies of categorical perception (e.g., [1]) proposed that acoustic variation was not even perceived. Reaction time data from Pisoni and Tash [3] seemed to confirm this notion for plain identification. For those stimuli within a phonetic category, identification times were the same. However, for same/different judgments, physically identical tokens were judged the same faster than ones that differed within the category. They interpreted this finding as evidence that different levels of processing are available to different tasks.

Another theory assumes that phonetic continua are evaluated in relation to phonetic prototypes [4]. In Samuel's account, phonetic decisions should be easiest when the prototypical value is used, and increasingly less easy as the acoustic distance between the stimulus and the prototype increases.

Other explanations of phonetic perception depend on the combination of gradient acoustic parameters. Massaro and Cohen [2], for example, compute phonetic decisions from interactions of two acoustic parameters. They have little to say about experiments with only one factor, however, so their theory will not be elaborated on here.

The present study will test the prototypical model against an extended categorical model in explaining identification times. The extension to the categorical model is that of an ambiguous region, rather than just a single boundary between categories. Such an extension is necessary to account for the fact that there are ambiguous stimuli that subjects can report as being ambiguous, rather than hearing the stimuli first as one category and then as another. Such a modification reduces but does not eliminate the differences between the models.

2. EXPERIMENTAL METHOD

2.1 Stimuli.

A male native speaker of American English recorded several tokens of the nonsense syllables /sa/ and /sta/. These were low-pass filtered at 10 kHz and digitized at 20 kHz on the Haskins PCM system [5]. One token of each syllable was selected, with each having the same duration in the fricative noise and in the vocalic segment. (160 and 240 ms respectively). A continuum of gap closures was made by inserting silence between the noise and the vocalic segment for /sa/. The original silence and the burst were removed from /sta/ and replaced as in /sa/. The values ranged from 0 to 78 in 3 ms steps, yielding 27 values; with two sources, there were 54 unique tokens.

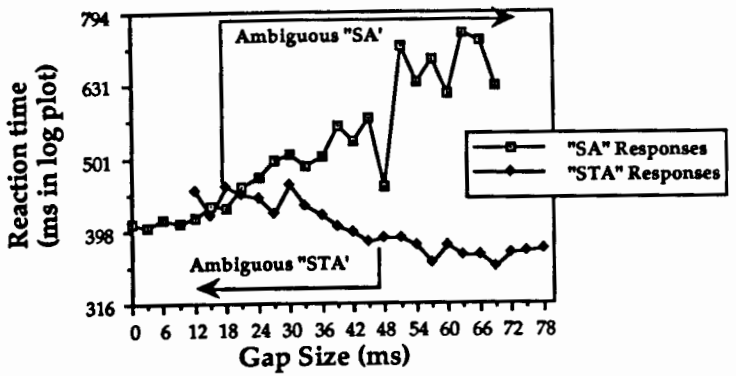


Figure 1: Identification times averaged across 10 subjects.

2.2 Subjects

The subjects were 10 Yale undergraduates who were paid for their participation.

2.3 Apparatus

The stimuli were recorded onto audio tape and played to the subjects over headphones. Their judgments as to whether the syllable was "SA" or "STA" were made by pressing a button, which generated a signal that stopped a clock on an Atari computer, giving the reaction time. Times were assessed from the onset of the vocalic segment, not from the onset of the syllable so that the times would not directly vary with stimulus duration.

2.4 Procedure.

A tape containing twenty exemplars of the stimuli was played to familiarize the subjects with the kinds of judgments they would have to make. Then four blocks, each containing five repetitions of each of the 54 stimuli, were presented. Each block, which had a different randomization of the stimuli, began with four "warm-up" stimuli which were not included in the analysis. A brief rest period was given between blocks.

3. RESULTS AND DISCUSSION

An analysis of the reaction times showed that the subject variances increased as the mean time increased, suggesting a log transform. All further times, though reported in ms, are means of the log values. An analysis that included block and source (original /sa/ or original /sta/) as factors revealed no effect of block, and an effect of source that was the same for both "s" and "st" judgments (the /sta/ source gave slightly faster times). Therefore, further analyses collapsed across these two factors.

It was desirable to eliminate mistaken responses, but the subjects had no way of indicating whether a response was the one intended or not. Instead, "isolates" were excluded. These were responses that were separated from a region of judgments by one or more gaps with no responses of that category. Thus one subject might have "s" responses at the 48 ms gap that would be included in the analysis (since gaps 45 and lower also had "s" responses), while another might have such a response excluded (since at least the 45 ms gap received no "s" responses). Isolates accounted for 1.1% of the data. Figure 1 shows the reaction times averaged across the 10 subjects.

The models were tested by examining how much of the possible variance they could account for. The variance of the individual times in relation to the overall mean established the minimal level for a

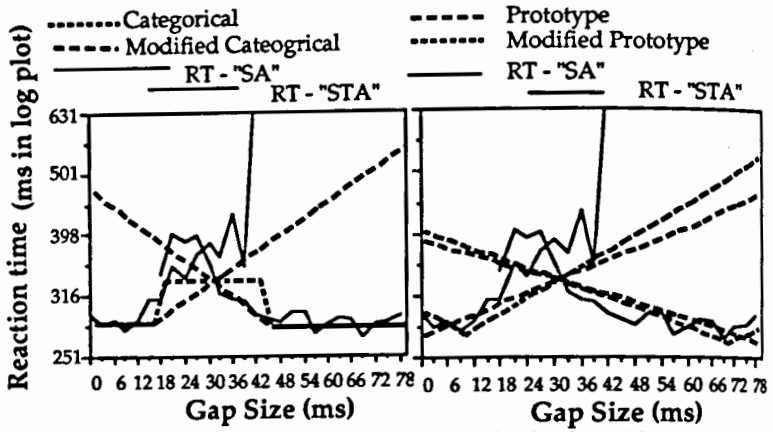


Figure 2: Models generated for the first subject's data (also presented).

model to attain, while using the mean for each judgment for each gap duration established the maximal level. (Since only one acoustic parameter was varied, this maximal description is essentially what would be proposed by a gradient theory, such as Massaro's fuzzy logic model.) The minimal model thus had two parameters (the overall mean for each response), while the maximal model would have between 27 and 54 (since the ambiguous regions could overlap), though the average was 39.9.

Figure 2 shows the four models that were generated for the first subject. (All the modelling was done for each subject individually.) The categorical model was generated with the five parameters: s-boundary (that is, the upper limit of gap values at which 95% of the responses were "s"), the st-boundary (the lower limit of gap values at which 95% of the responses were "st"), the mean times for the "s" region and for the "st" region, and the mean time for the ambiguous region (which included non-isolate "s" responses in the "st" region and "st" responses in the "s" region). In the modified categorical model, the time for the ambiguous responses was calculated from two parameters, a linear interpolation from the edge of the unambiguous region through the mean time for the ambiguous stimuli, temporally located in the center of the ambiguous region.

The prototypical model was generated by taking the fastest time for stimuli of any gap duration as the interpolation value for the continuum endpoints, with the other value being the mean of the responses to ambiguous stimuli (temporally located in the middle of the ambiguous region as in the modified categorical model). The modified prototypical model (which more closely resembles Samuel's) used the location of the fastest time as the definition of the subjects prototype. Values were then interpolated through the ambiguous region as before, and values toward the endpoints were interpolated with a mirror image of the pattern.

Figure 3 shows the percentage of possible variance accounted for by the four models. Since the range of "possible variance" was defined by two more models, it was possible to do worse than the minimum. For two subjects, this was in fact the case for all four models. Subject 5 had very little variation across the gap durations, and had very long times in general (about two standard deviations above the group mean). Subject 7 had a very small "s" range (i.e., only the 0 ms gap), and actually had faster times for ambiguous "st" judgments than for unambiguous ones. Still, for 9 of the 10 subjects, the modified categorical model performed better than the modified prototypical one. An analysis of variance was run on the percentages shown in the figure, with the factors of type (categorical or prototypical) and modification (modified or not). While type was not

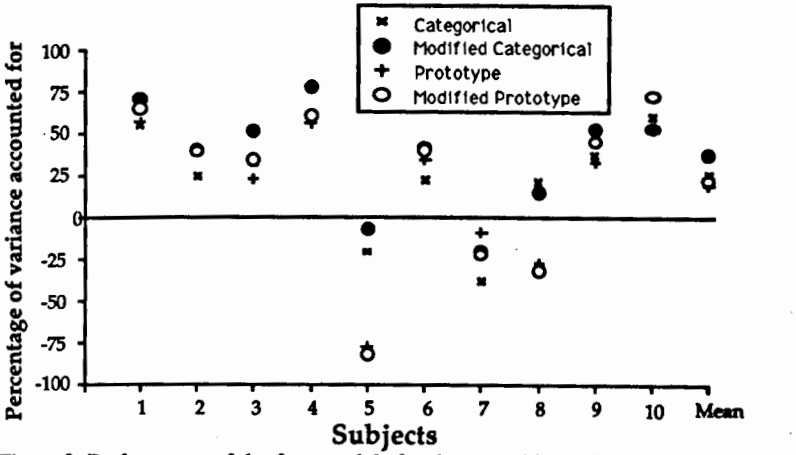


Figure 3: Performance of the four models for the ten subjects. Some of the symbols for the unmodified models are hidden by symbols for the modified ones.

significant as a main effect ($F(1,9)$ 1.49, n.s.), modification was ($F(1,9)$ 9.38, $p < .05$), as was the interaction ($F(1,9)$ 8.86, $p < .05$). As is apparent, only the modified categorical stands out from the others (by a Newman-Keuls post-hoc test).

Since the simple categorical model had one more parameter than the simple prototypical one, comparisons between those two models are somewhat problematic. Both modified models, however, required six para-meters, putting them on an equal footing.

This does not exhaust the possibilities for modelling the data, of course. One further modification of the prototypical models would be to allow the interpolation to be parabolic rather than linear. Though initially appealing, such a modification would make it very difficult to tell the prototypical model from the categorical--perhaps giving us a benign ambiguity. It may also be that there is a floor effect on the reaction times. Perhaps the times in the unambiguous regions were subject to, say, a mechanical limitation, so we might have found a more prototypical pattern if the limitation were circumvented. It is possible that a fast repetition (shadowing) paradigm might be useful here.

For the present results, however, it appears that the best model is the one that assumes that all unambiguous judgments are equally easy, while more difficult (due to ambiguity) ones become increasingly so the greater the distance from the category region.

Acknowledgments:

Work supported by NICHD grant HD-01994. Thanks to A. G. Samuel and A. G. Levitt for helpful comments.

References:

[1] Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957), "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology, 54,* 358-368.

[2] Massaro, D. W., and Cohen, M. M. 1983 Phonological context in speech perception. *Perception and Psychophysics, 34,*338-348.

[3] Pisoni, D.B., and Tash, J. (1974), "Reaction times to comparisons within and across phonetic categories", *Perc. and Psychophysics, 15,* 285-290.

[4] Samuel, A. G. (1982), "Phonetic prototypes", *Perc. and Psychophysics, 31,* 307-314.

[5] Whalen, D. H., Wiley, E. R., Rubin, P. E. and Cooper, F. S. (1990), "The Haskins Laboratories pulse code modulation (PCM) system", *Behavior Research Methods, Instruments and Computers, 22,* 550-559.

INFLUENCE OF NEGATIVE INTENSITY GLIDES ON THE DISCRIMINATION OF SPEECH SEGMENT DURATION

Y. Nishinuma & S. Santi

Institut de Phonétique, Université de Provence
CNRS R.U.A. 261, 13621 Aix-en-Provence, France

ABSTRACT

The discrimination of duration was investigated using synthetic vowels containing negative intensity glides (0 dB, -6 dB, -12 dB, and -18 dB). Test stimulus durations ranged from 100 to 300 ms in steps of 20 ms. The standard stimulus was 200 ms in duration and had a stable intensity. Stimulus pairs were presented to 20 subjects (constant method) and their task was to state which vowel in the pair sounded longer (forced choice). Results indicate that a drop in intensity of more than 12 dB has a significant effect on the perception of duration, and thus on its discrimination.

1. INTRODUCTION

The prosodic analysis of speech, which consists of interpreting acoustic parameters such as duration, fundamental frequency, and intensity, is not an easy task. Two reasons for this are that (1) these factors are not independent in human perception and (2) they vary as the speech signal evolves in time (within a syllable, a word, a clause, etc.). It is known that the perception of pitch variations depends upon their duration [11]. Furthermore, the melodic contour of segments with negative and positive intensity glides are perceived differently [13].

However, we know little about the interaction between duration and intensity in speech. In particular, the influence of intensity variations on the ability to discriminate the duration of speech sounds has not been experimentally documented. This problem came up in our previous study, which investigated the differential threshold of syllable duration in a sentence context [8]. Duration discrimination was found to be significantly less accurate on the final syllable than on preceding syllables. The same tendency was observed in Klatt and Cooper's data [7], which show a higher threshold for fricatives at the end of sentences than in other locations. This led us to raise the question of whether a drop in intensity (-16 dB in our case) on the final syllable of a sentence would make it difficult to correctly perceive that syllable's duration. An experiment carried out to verify this hypothesis is reported below.

2. EXPERIMENT

Klatt's formant synthesizer was used to generate stimuli for the perception test [6]. The goal was to obtain speech-like stimuli which varied in both duration and intensity. Negative intensity glides were used to approximate the final syllable of declarative sentences.

The material was designed to be used in a psycho-acoustic test based on the constant method. The standard stimulus was the vowel /a/ with a duration of 200 ms (an average syllable length) and a stable intensity of 80 dB. The test stimuli were synthesized with durations ranging from 100 ms to 300 ms in 20 ms steps (for a total of 11 different durations). Four linear intensity glides were utilized: 0 dB, -6 dB, -12 dB, and -18 dB. The fundamental frequency contour was the same for all stimuli. A slight lowering of pitch from 140 Hz to 130 Hz made the stimuli sound natural. The standard stimulus was paired with each of the test stimuli. The two vowels in each pair were separated by a silent pause lasting 600 ms. The interval separating one pair from the next was three seconds. Both within-pair orders were used (standard-test, test-standard). Each pair occurred four times. Thus, the total number of pairs was 352 (11 durations x 4 intensity glides x 2 stimulus orders x 4 repeats). Stimuli were generated in random order by a computer and recorded on a digital audio-tape. A trial series of 22 pairs was added to the beginning of the test sequence. A short beep followed by a five second silence was inserted every 22 pairs.

The perception tests were carried out in a soundproof room. Twenty subjects were tested individually, each in a single trial lasting 20 minutes. The listening level of the standard stimulus leaving the headphones was set at approximately 70 dB SPL. The written instructions to the subjects were as follows: "You are going to listen to many pairs of vowels /a, a/. For each pair you have to ask yourself the following question: Which of the two vowels is longer, the first or the second? Even if the intensity changes, please judge only the duration." The subjects responded by checking the appropriate answers on a forced-choice answer sheet.

3. RESULTS AND DISCUSSION

An analysis of variance on the data yielded a significant difference between subjects ($F_{(19,351)} = 36.84$, $p < 0.001$). This means that each subject had his or her own strategy to carry out the task. In addition, as expected in this kind of psycho-acoustic test, the order in which the stimuli were presented affected the subjects' behavior ($F_{(1,351)} = 55.6$, $p < 0.001$). Variations in the intensity factor also produced significantly different scores ($F_{(3,351)} = 168.97$, $p < 0.001$). Note however that neither the repeat factor nor the response

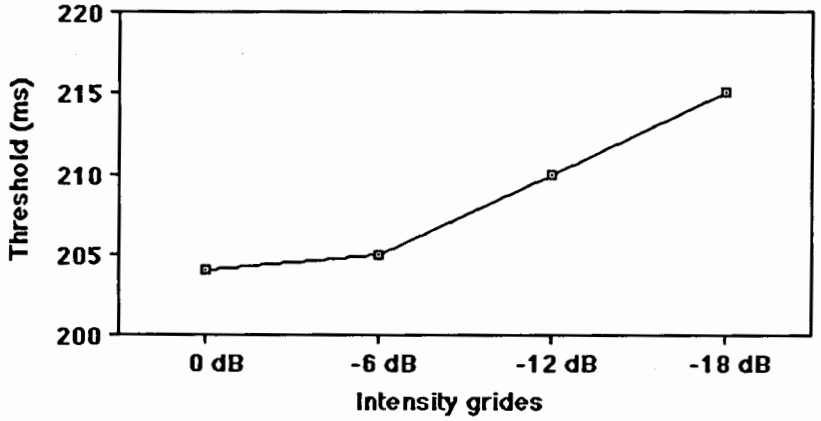


Figure 1. **Duration threshold as a function of intensity glide**

category (which member of the paircategory (which member of the pair was perceived as longer) had a statistically significant effect.

In order to compute the mean threshold (for the 20 subjects pooled) we interpolated the value for the duration at the 75% correct answer level by summing the four repeats per subject. This mean was calculated for the two stimulus presentation orders, two response categories, and four intensity glides. The averages of these values are shown in **Figure 1** and **Table 1.**

The threshold turned out to be proportional to the magnitude of the intensity glide. In other words, as the intensity glide became steeper, the detection of duration became less and less accurate. This effect on the discrimination of duration is clearly shown by the progressive increase in the means and standard deviations shown in Table 1. The difference between the stimulus with the steepest drop (-18 dB glide) and the stable intensity stimulus (no glide) exceeds 7%. Interestingly, the T-test on the data for the first two intensity glides (0 dB and -6 dB) did not yield any significant differences. Apparently, a 6 dB drop in intensity does not lead to difficulty in detecting the correct duration. In contrast, the stable vs. -12 dB difference ($t_{(19)} = 4.14$, $p < 0.001$) and the stable vs. -18 dB difference ($t_{(19)} = 5.36$, $p < 0.001$) were both highly significant. It is noteworthy that our results indirectly support those obtained by Rossi [13], who estimated the intensity glide threshold to be approximately 11 dB for a vowel lasting 200 ms. The observed change in the way intensity information is processed seems to depend on whether or not the intensity decreases beyond that critical value, although we do not know precisely where in our auditory system that change occurs.

This tendency is even more apparent if we consider stimulus presentation order. For the standard-test order, it can be hypothesized that subjects pay attention to the duration of the final syllable, which has a negative intensity glide in this experiment. The computed threshold values were 220 ms, 230 ms, 233 ms, and 235 ms, for 0 dB, -6 dB, -12 dB, and -18 dB, respectively. This indicates that when the penultimate syllable measures 200 ms and has a stable intensity, final syllables with intensity glides of 18 dB may have to be longer than 235 ms.

However, this hard and fast interpretation may need some qualification due to one peculiarity of this experiment. In comparison to the results published in psycho-acoustic studies using speech sounds, our threshold value at 200 ms is remarkably (even excessively) precise (1%, or 2 ms; cf. Figure 1 and Table 1). For a standard stimulus duration of about 200 ms, the reported threshold values fall between 8% and 30%. These experiments used several standard stimulus durations ranging from some ten milliseconds to several hundred milliseconds [1, 2, 3, 5, 7, 9, & 12]. In our experiment, all 352 stimulus pairs had a 200 ms vowel (the only standard duration used), with a level intensity in first or second position. This may have overexposed subjects to that particular duration, causing better performance. Therefore, the duration threshold defined here, (i.e. as a function of intensity glide) should be used in conjuction with those obtained under normal, stable intensity conditions.

Table 1. Duration threshold as a function of intensity glide

Intensity glides	Mean threshold	Standard deviation	N
0 dB	202	10	80
-6 dB	206	14	80
-12 dB	216	16	80
-18 dB	217	22	80

4. CONCLUSIONS

This study has provided some experimental evidence of how well we hear at the end of declarative sentences. The results of our perception tests demonstrated that the discrimination of duration may be significantly deteriorated by a progressive decrease in intensity of more than ten decibels. Our results may have some implications for the interpretation of prosodic data at the sentence level.

5. REFERENCES

[1] Bochner, J.H., Snell, K.B., & MacKenzie, D.J. (1988) "Duration discrimination of speech and tonal complex stimuli by normal hearing and hearing-impaired listeners", *JASA*, 84(2), 493-500.

[2] Bovet, P, & Rossi, M. (1977) "Etude comparée de la sensibilité différentielle à la durée avec un son pur et avec une voyelle", *Du temps biologique au temps psychologique*, PUF, Paris, pp.289-306.

[3] Eilers, R.E., Bull, D.H., Oller, D.K., & Lewis, D.C. (1984) "The discrimination of vowel duration by infants", *JASA*, 75(4), 1213-1218.

[4] Huggins, A.W.F. (1971) "Just noticeable differences for segment duration in natural speech", *JASA*, 51(4), 1270-1278.

[5] Huggins, A.W.F. (1971) "On the perception of temporal phenomena in speech", *JASA*, 51(4), 1279-1290.

[6] Klatt, D.H. (1980) "Software for a cascade/pararel formant synthesizer", JASA, 67(3), 971-995.

[7] Klatt, D.H., & Cooper, W.E. (1975) "Perception of segment duration in sentence contexts", in *Structure and Process in Speech Perception* (eds. A. Cohen & S.G. Nooteboom), Springer, New York, 69-89.

[8] Nishinuma, Y., (1990) "Discrimination of syllable duration in English and French short sentences", *JASA*, 87(Suppl 1), s72.

[9] Nooteboom, S.G. (1973) "The perceptual reality of some prosodic durations", *Journal of Phonetics*, 1(1), 25-46.

[10] Rochester, S. (1971) "Detection and duration discrimination of noise increments", *JASA*, 49(2), 1783-1794.

[11] Rossi, M. (1971) "Le seuil de glissando ou seuil de perception des variation tonales pour les sons de la parole", *Phonetica*, 23(1), 1-33.

[12] Rossi, M. (1972) "Le seuil différentiel de durée", *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, (éd. A. Valdman), Mouton, The Hague, 435-450.

[13] Rossi, M. (1978) "The perception of non-repetitive intensity glides on vowels", *Journal of Phonetics*, 6, 9-18.

AMPLITUDE AS A CUE TO WORD-INITIAL CONSONANT LENGTH: PATTANI MALAY

Arthur S. Abramson

The University of Connecticut and Haskins Laboratories

ABSTRACT

Word-initial Pattani Malay consonants are short or long. The closures of the "long" consonants are longer than those of the "short" ones; this is a sufficient cue for perception, but in voiceless plosives the duration of the silent closure is audible only after a vowel, yet listeners label such isolated words well and so must use other cues. The peak amplitudes for the first syllables of disyllabic words are greater for initial long plosives. In this study, increments of closure duration and amplitude were pitted against each other for original short plosives and decrements for original long plosives. In tests, duration was by far the more powerful cue, although amplitude did affect the category boundary. By itself, however, amplitude is a weak cue. Further work is planned on the possible role of the shaping of the amplitude contour.

1. INTRODUCTION

Many languages are described as having a phonological distinction of length in vowels or consonants, or even both. If the term is taken literally, we would expect to find that the underlying mechanism is control of the relative timing of the articulators. Even so, a single mechanism might have a number of acoustic consequences, each of which could help in perception.

Pattani Malay, spoken by about a million ethnic Malays in southern Thailand, is unusual not only in having a length distinction for consonants in word-initial position but also in having one that is relevant for all phonetic classes of consonants in that position [3]. Here are some minimal pairs of words showing the contrast:

/labɔ/	'to profit'	/ɭabɔ/	'spider'
/make/	'to eat'	/mːake/	'eaten'
/bule/	'moon'	/bːule/	'months'
/katoʔ/	'to strike'	/kːatoʔ/	'frog'

If, indeed, the crucial aspect of the articulatory gesture is the duration of the closure or constriction, for pairs like the first two it would not surprise us to find that the length distinction is quite discernible whether in utterance-initial or intervocalic position. But what about the stop consonants, especially the voiceless unaspirated stops of the language? The voiced stops do have voicing lead, so if you are close enough, you can hear short or longer stretches of glottal pulsing during the occlusion. The occlusions of the voiceless stops, however, are silent.

In earlier work [2], I presented acoustic measurements of closure durations for the language, showing that the putative length categories are well separated by duration. Of course, the voiceless stops could not be measured in utterance-initial position. In another study [1], I demonstrated, by systematically increasing the durations of short closures and decreasing the durations of long closures, that this feature is a sufficient and powerful acoustic cue for the perception of the distinction.

As for the voiceless stops, it was conceivable that the two categories were auditorily distinguishable in medial position only. This turned out not to be so in my control tests with unaltered words. Doing only slightly worse than with the other classes of consonants, native speakers rather accurately identified short and long voiceless stops in isolated words.

Among the various plausible acoustic effects of the mechanism, the most likely for the largely disyllabic words involved, was the peak amplitude of the first syllable relative to the second. Indeed, measurements [2] revealed that this ratio is greater for long plosives, that is, both stops and affricates. Presumably, greater air pressure accumulated behind the occlusion before release accounts for the differences. Although both voiced and voiceless plosives showed a significant difference, the level of significance was higher for the latter. No doubt, this is to be explained by differences in glottal impedance of the airflow. The difference is not significant for the continuants.

2. PROCEDURE

This paper is a progress report of my test of the hypothesis that the peak amplitude of the first syllable relative to the second in disyllabic words is a sufficient cue for the perception of the distinction between short and long voiceless stops in Pattani Malay. For my major experiments, as part of an interest in combinations of phonetic features underlying the same phonemic distinction, I have pitted variants in duration and amplitude against each other to determine their relative power.

2.1. Control tests

Although the identifiability of initial short and long consonants had been demonstrated [1], it seemed desirable also to do control tests for the recordings of my new speaker for this study. For each of seven minimal pairs of words I prepared a test containing 20 tokens of each of the two words, yielding 40 randomized stimuli. There were two such randomizations for each word pair. The nasal, lateral, fricative, and plosive categories were represented. The plosives included voiced and voiceless stops and voiceless affricates. (Unfortunately, my only pair of voiced affricates included a word, as I learned later, that would have embarrassed the women among the subjects, so I could not use that test.) The subjects were 30 undergraduate students, all native speakers of Pattani Malay, at the Prince of Songkhla University, Pattani, Thailand.

2.2. Amplitude vs. duration

To test for the relative power of amplitude and duration, three pairs of words with velar, dental, and labial short and long stops respectively were used. All of them were recorded at the end of the carrier sentence /dio katɔ/ 'he said.' By means of the Haskins Laboratories Waveform Editing and Display System (WENDY), the stop closure of the short member of each pair was lengthened in 20-ms steps until it reached or exceeded the duration of its long counterpart. The closure of the long member was shortened in the same way. The first syllable of each variant of the original short stop was increased in amplitude in five 2-dB steps. Likewise, the first syllable of each variant of the original long stop was *decreased* in amplitude in five 2-Db steps. Two test orders were recorded from randomizations of two tokens each of all the resulting stimuli and played to 30 native speakers for identification of the key words.

2.3. Amplitude in isolated words

The perceptual efficacy of amplitude without help from closure duration was tested by taking all the amplitude variants from the original short and long forms of one of the word pairs in section 2.2. Two test orders were recorded from randomizations of four tokens of each stimulus and played to 30 native speakers.

3. RESULTS

3.1. Control tests

The previously demonstrated identifiability of the utterance-initial consonants [1] was reaffirmed. The major difference is that the voiceless long affricates in this sample were labeled correctly 96% of the time, whereas in the last study it was just above chance at 55%.

3.2. Amplitude vs. duration

Because of the limitation on space, the results of only two of the experiments are given here. Figure 1 gives the responses of 30 native speakers to nine durations in 20-msec steps of the [k]-closure in /kameŋ/ 'goat' combined with six amplitude levels in 2-dB steps. The vertical axis

shows the percentage identification as short /k/. The earlier crossover of the higher-amplitude curves at the 50% point to the long-/kː/ category, giving judgments of /kaməŋ/ 'goatlike,' is highly significant [F (40, 1160)=9.0, p < .001]; nevertheless, the values of duration at the short end are very little affected. The opposite procedure, shortening original long /kː/ and lowering the amplitude, yielded similar results, as shown in Figure 2. The results are essentially the same for the other two places of articulation.

3.3. Amplitude in isolated words

In Figure 3 both the short and long responses are plotted for increments of amplitude on original /pagi/ 'morning.' While the two curves converge, they never cross each other. Figure 4 shows rather similar effects for decrements of amplitude combined with isolated tokens of /pːagi/ 'early morning.'

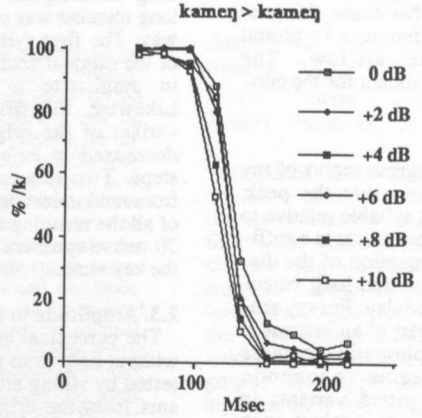


Fig. 1. Responses to /kamen/ 'goat' and its variants with increased closure duration and first-syllable amplitude.

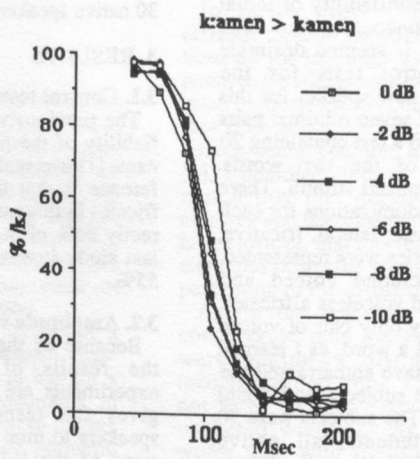


Fig. 2. Responses to /kːamen/ 'goatlike' and its variants with decreased closure duration and first-syllable amplitude.

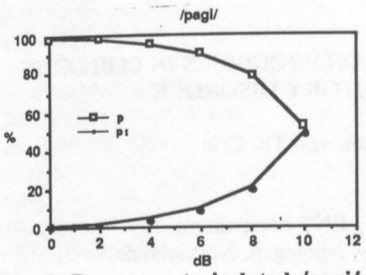


Fig. 3. Responses to isolated /pagi/ 'morning' and its variants with increased first-syllable amplitude.

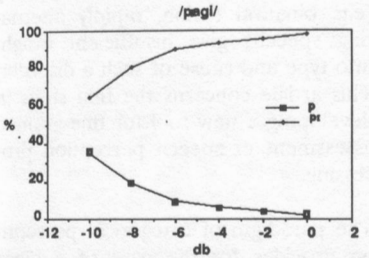


Fig. 4. Responses to isolated /pːagi/ 'early morning' and its variants with decreased first-syllable amplitude.

4. CONCLUSION

It is clear that when both features are present, duration is dominant; nevertheless, the boundary between the two perceptual categories is significantly affected by relative amplitude. In utterance-initial position, however, relative amplitude is only a weak cue, apparently secondary to something else.

To understand how the length distinction is perceived in uitterance-initial voiceless plosives, perhaps further work should be done on the possible role of the shaping of the amplitude contour. That is, maybe a finer analysis of utterances and a more complicated making of stimuli will show, for example, that the rise-time of the amplitude carries more weight than the peak value, or that the two work together. Indeed, a very preliminary look at this time suggests that the rise time is shorter in the production of the long stops. Also, it is possible that the major amplitude difference is confined to the region of the release burst. Other features that have not seemed promising so far, such as fundamental frequency and rate of formant transitions, may have to be examined more closely too.

5. ACKNOWLEDGMENTS

The work was supported by NICHD Grant HD-01994 to Haskins Laboratories. The fieldwork in Thailand was made possible by a sabbatical leave from The University of Connecticut in 1988. I am grateful to the National Research Council of Thailand, the Department of Islamic Studies of The Prince of Songkhla University, Pattani, and the Department of Linguistics of Chulalongkorn University, Bangkok for their warm hospitality and help.

6. REFERENCES

[1] ABRAMSON, A.S. (1986), "The perception of word-initial consonant length: Pattani Malay", *Journal of the International Phonetic Association, 16*, 8–16.

[2] ABRAMSON, A. S. (1987), "Word-initial consonant length in Pattani Malay", In *Proceedings of the XIth International Congress of Phonetic Sciences., 6* (pp. 66–70), Tallinn: Academy of Sciences of the Estonian S.S.R.

[3] CHAIYANARA, P.M. (1983), *"Dialek Melayu Patani den bahasa Malaysia"*, Kuala Lumpur: Master's thesis, University of Malaya.

TESTING SUBPHONEMIC PERCEPTION PROCESSES IN CHILDREN SUSPECT FOR AN AUDITORY DISORDER

P. Groenen, B. Maassen and Th. Crul

Child Neurology Centre / ENT-department
Institute of Medical Psychology, Nijmegen, Netherlands

ABSTRACT

Currently available speech perception tests provide insufficient insight into type and cause of an auditory processing disorder. The paradigm of categorical perception which discriminates an auditory- and a phonetic stage, combined with the idea of reducing redundancy of speech stimuli in order to have access to speech processing, enables us to develop a new tool for fine-grained assessment of central auditory pathology. By now, the first steps have been taken in the development of the test consisting of a series of experiments on verbal dyspraxic children, reading and spelling disordered children and children with a severe history of otitis media.

1. INTRODUCTION

Several categorical perception studies on auditory perceptual behaviour in groups of children with a specific speech- or language disorder have shown a deviant response pattern as compared to a normal control group. Auditory processing disorders, without there being a hearing-loss according to tone- and speech-audiogram, have manifested themselves in a delayed speech- and language development. Currently available speech perception tests, generally containing tasks based on monotic measures (e.g. filtered speech, competing signals), dichotic measures (e.g. syllables, words, sentences) or tests of binaural functionality (e.g. binaural fusion, rapidly alternating speech) give insufficient insight into type and cause of such a disorder. This article concerns the first steps in developing a new tool for fine-grained assessment of speech perception processing.

The paradigm of categorical perception provides for the base of a more sensitive and analytical test. According to this paradigm, a continuüm of speech stimuli covering a phonological contrast is constructed by digitally manipulating a single acoustic cue. The central idea is that perceptual boundaries arise along the physical continuüm, with qualitive resemblances within each category and qualitive differences between them or in a more modern psychophysic sense [2], there is a quantitative discontinuity in discrimination at the category boundary, as measured by a peak in discriminative acuity at the transition region of adjacent categories.

Two starting-points for research are of interest. Firstly, the speech perception model of Pisoni and Sawusch [3], which differentiates between a pre-categorical or auditory stage and a categorical or phonetic stage, plays an important role. During the first, auditory stage, listeners take in short stretches of the raw acoustic signal and make a preliminary auditory analysis. No speech segments have yet been identified. In the second, phonetic stage listeners examine their auditory memory and combine the acoustic cues to form a phonemic representation. This stage preserves the nature of the identification but does not preserve the acoustic cues upon which it was based.

By comparing identification and discrimination performance (labeling words and telling them apart) we can derive the level of the auditory disorder. Discrimination scores can be predicted on the base of identification scores [4]. Assuming that the listener bases discrimination judgements only on phonetic information, the observed discrimination scores should correspond to the predicted ones. If the subject has access to auditory precategorical information, the discrimination scores should be higher than the predicted scores.

Secondly, we assume that a speech perception problem may be caused by a neurological reduction of the redundancy of the auditory processing system [1] which we call the internal redundancy of the perceptual system. During speech perception there has to be a considerable reduction of the internal redundancy before stimuli with their normally high external redundancy (which is implicit in the normal structure of the acoustic speech signal) cannot be identified anymore. The conclusion is that the external redundancy must be reduced such that the speech perception test becomes more sensitive to small reductions of the internal redundancy.
Exactly this occurs in constructing a speech-continuüm, a word (one end of a phonological contrast) is transformed to another (other end) bij systematically decreasing and increasing the salience of the acoustic cue.

A speech continuüm will contain stimuli which do not discriminate between normal and pathological groups. In order to maximize the efficiency of the testprocedure and to minimize effects of response bias these stimuli can be eliminated out of the test. Only the critical stimuli, the stimuli which are as sensitive to account for differences between normal and deviant children will be included in the final test. Sensitized speech stimuli can be singled out of stimuli near the phonemeboundary.

2. PROCEDURE

By now the first steps in developing such a sensitized perception test are being carried out. We examine scores on tests where two acoustic cues are systematically varied: VOT (/bak-pak/, i.e. BOX-PACKAGE) and place of articulation, second and third formant transition (/bak-dak/, i.e. BOX-ROOF).

Three experimental groups are tested:
- verbal dyspraxic children aged between 6 and 11 years. These children show dysfunctions in the programming of their articulomotoric organs. We suppose that (a part of) this group is marked by a central auditory dysfunction.
- children with reading- and spelling problems, 2 groups; one in the age of 6-7 years, the other in the age of 8-10 years. Disordered auditive functions could be one of the main causes of the problems.
- children with a severe history of otitis media with effusion (OME) in the early childhood (aged round 2 years). At the time of testing they are 6-8 years of age. Due to temporary conductive hearing-loss these children show disorders in their auditory deve-

lopment. We take interest in the extent in which the central auditory functions are affected. Our experimental design is based on a division into 4 groups: a) Children with an OME-history and with a delayed speech- and language development, b) Children with an OME-history and without a delayed speech- and language development, c) Children without an OME-history and with a delayed speech- and language development and d) a control group children without an OME-history and without a delayed speech- and language development.

3. RESULTS

At the time of writing this paper only the data of the verbal dyspraxic group have been fully analyzed. For, at this time, lack of data of a control group we shall present in this paper some examples of identificationcurves of the place-of-articulation continuüm /b-d/ of a male adult without hearing problems or a delayed speech- or language development with a normal response pattern and a verbal dyspraxic child with a abnormal response pattern (figure 1).

Remarkable is the less steepness of the slope of the curve of the verbal dyspraxic child compared to the adult. We interpret this as a less consistent labeling ability, pointing towards a processing dysfunction at the phonetic level. In figure 2 the corresponding discriminationcurves are presented. Again there is a difference between the verbal dyspraxic child and the adult; the ability to discriminate is less in the verbal dyspraxic child. Furthermore the overall form of the discriminationcurve doesn't agree with the foneemboundary found in the identificationcurve.

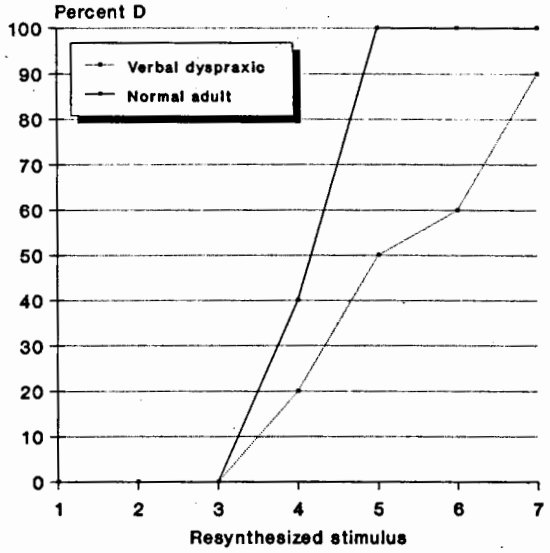


Figure 1. Identificationcurves of a normal hearing adult and a verbal dyspraxic child (dotted line).

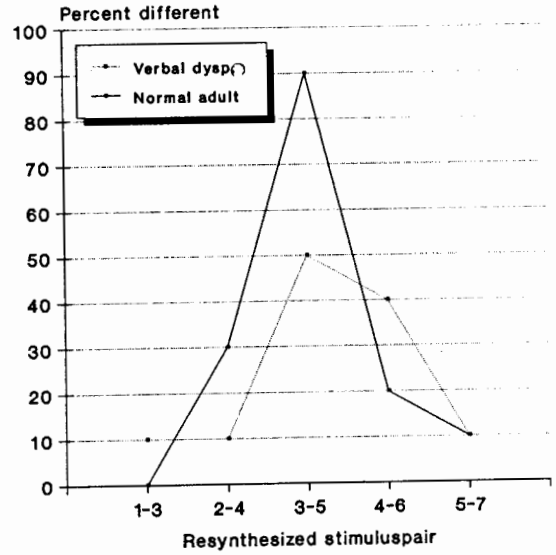


Figure 2. Discriminationcurves of a normal hearing adult and a verbal dyspraxic child (dotted line).

Taken together, this verbal dyspraxic child shows a disordered auditory stage (less discriminative power), followed by an inadequate functioning of the phonetic stage (less steep slope, shifted phoneemboundary), partially an effect of the malfunction in the auditory stage.

If we take the group verbal dyspraxic children as a whole, there's a great variability in scores suggesting the possibility of dividing the group in a number of verbal dyspraxic children having speech processing problems and a number of verbal dyspraxic children showing no central auditory dysfunction.

More elaborated analysis of the experimentai groups, yet unavailable, will be given at the presentation.

REFERENCES

[1] Bocca, E., & Calearo, C. (1963), "Central hearing processes", in: J. Jerger (Ed.) *Modern developments in audiology*, 337-370.

[2] Harnad, S.R. (1987), "*Categorical perception*", Cambridge University Press.

[3] Pisoni, D., & Sawusch J. (1975), "Some stages of processing in speech perception", in: A. Cohen & S.G. Nooteboom (Eds.) *Structure and process in speech processing*. New York: Springer Verlag.

[4] Pollack, I. & Pisoni, D. (1971), "On the comparison between identification and discrimination tests in speech perception", *Psychonomic science, 24*, 299-300.

THE INFLUENCE OF SPEAKERS' OWN SPEECH TEMPO ON THEIR TEMPO PERCEPTION

M. Gósy

Research Institute for Linguistics, Budapest, Hungary.

ABSTRACT

The point of departure for the present paper is the assumption that the speaker's own speech tempo determines his judgements concerning that of other people. Experimental results supported a significant concurrence in tempo perception of 'extreme' speakers as opposed to 'moderate' speakers. A significant correlation was found between the speakers' comprehension and their own speech tempo. It can also be claimed that speakers/listeners judge speech tempo on the basis of the active levels of their speech perception mechanism.

1. INTRODUCTION

Authors of a number of studies agree that tempo perception is basically determined by three factors: articulation rate, the number of pauses, and the duration of pauses [4, 5]. Tempo perception studies are also made difficult by problems concerning the recognition, demonstration, and order of importance of a number of other factors including changes in fundamental frequency (pitch), average intensity, word frequency (of occurrence), syllable structure, rhythmic structure, syntactic properties, etc. [3]. However, very few attempts have been found in the literature to deal with the connection of *production and perception in relation to speech tempo*. It is usually postulated that there should be a very close connection between the speaker's own rate of speaking and his perception and comprehension with respect to speech tempo [1, 2]. Can we then claim that there is a linear connection between tempo production and perception, namely: the faster the speaker's usual speech tempo the faster his/her comprehension as well? Does this apply to tempo perception, too? What are the criteria of applicability of this rule?

In order to answer these questions, an experiment containing 4 subtests has been performed with Hungarian-speaking native speakers/listeners. The aim of the experiment was to describe the effect of the subjects' own speech tempo on (i) their tempo perception and (ii) their speech comprehension.

2. PROCEDURE

Various methods were used for the subtests. (i) For the first experiment nine speech samples were recorded in random order from Hungarian-speaking native subjects (ages ranged from 25 up to 80). Subjects were selected so that all categories be represented from very slow (articulation rate /AR/: 8.85 sounds/s, overall speech rate /OSR/: 7.25 sounds/s) up to very fast (AR: 18.2 sounds/s, OSR: 14.3 sounds/s). Each speech sample was taken out of a longer monologue and took 1.5 minutes on average. The listeners' task was to judge the speech tempo of each speaker's sample by means of a questionnaire. The categories of the questionnaire were 'very slow', 'slow', 'normal', 'accelerated', 'fast', and 'very fast'.

(ii) The material of the second test consisted of 12 artificial, synthesized sentences (the synthesis was made by a PCF speech synthesizer controlled by an IBM PC). The same sentence had been altered in relation to its overall speech tempo in two ways: by changing the "articulation rate" of the sentence and by adding one or two pauses at the appropriate grammatical boundari(es) of the sentence. The subjects' task was the same as in the first subtest.

(iii) 8 sentences announced by a trained male speaker were chosen for the third test, and a verification method was used. The sentences were speeded up, and their articulation rate ranged from 20.2 sounds/s up to 24.4 sounds/s. The subjects' task was to decide whether the sentences they heard were true or false. The reaction times (RT) of each subject were measured by means of a fundamental frequency and intensity meter with the accuracy of 10 ms.

(iv) The subjects' spontaneous speech was tape recorded in the final experiment. From their recorded 8-10-minute speech 2-minute samples were picked out for further analysis concerning AR & OSR. The duration and types of pauses were also examined. Counting the speech sounds of the speech sample, the rate was expressed in terms of sound/s.

After finishing the experiments, each subject was asked to judge his/her own average speech tempo according to the formerly used tempo categories. The subjects' sex and age were also recorded on the same answer sheet.

37 subjects were selected from all candidates for further examinations. Three tempo groups were defined: a group of 'slow' speakers, a group of 'moderate' speakers and a group of 'fast' speakers. Examining the data, significant correlation was found between the AR and OSR values of our subjects ($p<0.05$). 6 subjects were found to be 'fast' speakers in terms of AR and 'moderate' speakers in terms of OSR. So, a fourth tempo category had to be established consisting of subjects having 'fast' AR and 'moderate' OSR and this was labelled the group of 'rapid' speakers.

3. RESULTS

Figure 1 shows the responses of various groups of subjects for all synthesized sentences according to the possible tempo categories. The listeners do perceive the physical changes of sentences. In the case of sentences containing 1 or 2 pauses, however, the judgements spread along the various tempo categories. The question is whether the distribution of tempo perception is based on the subject's own tempo production. Analyzing the average values for each sentence of each group, it can be stated that there are no important differences among the subjects' judgements. However, the data of the three groups are significantly different at the level of 0.05. This means that there is a slight but definitive difference of tempo perception among subjects with diverse speech tempo production. The mean values of the judgements show very constant changes across the tempo categories. These changes reveal more similarity for the 'slow' and 'fast' speakers than for the 'moderate' and 'rapid' speakers. There is a significant difference in the judgements of the 'slow' speakers concerning the category of 'accelerated' tempo as opposed to the judgements of the other two groups. 'Rapid' speakers' performance shows a relatively different distribution in relation to that of the other two groups. On the basis of these data, a *hypothesis* has been developed on the interrelatedness of the speakers' own tempo production and their tempo perception: *'slow' and 'fast' speakers tend to perceive tempo similarly to one another* while 'moderate' speakers do not. 'Rapid' speakers seem to behave perceptually in a way different from the other three groups. We also found that the extreme speakers tend to perceive tempo more on the basis of AR than on the basis of OSR, so the pauses might not influence their tempo perception.

Figure 2 shows the responses of various groups of speakers for the speech samples used across the possible tempo categories. Subjects appear to judge the tempo of the speech samples according to AR rather than on the basis of OSR. The data show: (i) There are larger differences among the tempo categories in each test group than in the case of isolated sentences, and (ii) the distribution of the judgements does not show a constant trend.

The number of responses referring to the 'moderate' tempo category is significantly different in the case of the perception of the synthesized sentences and the speech samples ($p<0.05$). This means that people's perception mechanism has grown accustomed to the tempo changes of human speech and they are more flexible when judging it than in the case of one sentence where the upper levels of the decoding mechanism should not work, so they can judge the tempo of each sentence more accurate to the actual physical values. The data show again a very similar concurrence of judgements made by the 'slow' and 'fast' speakers. Similar judgements of the 'slow' and 'fast' speakers were found in all tempo categories with the exceptions of the 'fast' and 'very fast' categories. In the case of these two tempo categories the 'slow', the 'moderate' and the 'rapid' speakers judged similarly while the 'fast' speakers differed from all the others. The 'rapid' speakers show a significant difference in their judgements from the other groups of speakers. However, in some cases their judgements fall close to the judgements of one of the goups of speakers. This co-occurrence does not show any systematic character.

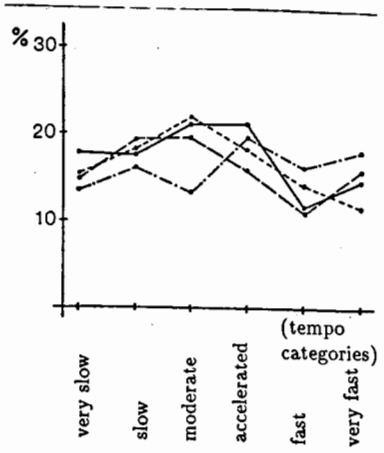


Figure 1.
Tempo perception of sentences by 'slow' (•——•), 'moderate' (•– – –•), 'fast' (•–—•), 'rapid' speakers (•–•–•–•).

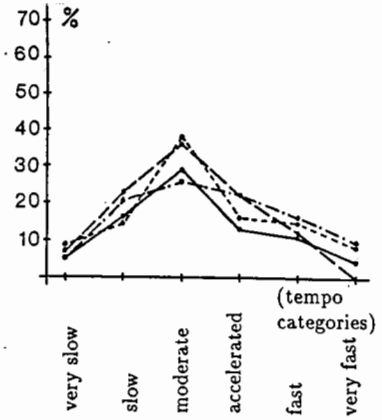


Figure 2.
Tempo perception of texts
see Fig.1. for the key

The question concerning the analysis of the reaction times was whether the subjects' own speech tempo influences their comprehension rate. The following RT values were obtained: in the case of 'slow' speakers 0.46–1.83 s, in the case of 'moderate' speakers 0.62–1.13 s, in the case of 'fast' speakers 0.3–1.15 s, and in the case of 'rapid' speakers 0.55–0.7 s. A significant difference has been found in reaction times between the various goups of speakers ($p<0.001$). There is a strong correlation between the subjects' articulation rate and their reaction times which shows that if the tempo of speech production increases the reaction time of the subject decreases. However, there are important differences among the subjects' reaction times within one group. The 'extreme' speakers' reaction times are extreme while the 'moderate' speakers' reaction times are not. There are subjects with fast AR and both with short and long reaction times; and – similarly – there are other subjects with slow AR and with both short and long reaction times. This part of our analysis *supports again the similar perceptual behaviour of the 'slow' and 'fast' speakers.* The largest RT values were found with the 'slow' speakers and the 'fast' speakers. The 'moderate' and the 'rapid' speakers' RT values were similar

to one another.

There is a significant difference between the RT values of affirmative and negative sentences with true contents; however, there was no significant difference between the same structures with false contents.

Finally, the subjects' age, sex, and their opinion about their own speech tempo were taken into consideration. There was a very strong correlation between the subjects' objectively measured speech tempo and their subjective judgements ($p<0.001$). We found that most of our extreme speakers were male while the 'moderate' speakers were mainly female subjects ($p<0.05$). There was no significant correlation between the subjects' age and their speech tempo categories.

4. CONCLUSIONS

– It had been assumed that the faster the speaker's own speech the less fast he perceives that of others. From this hypothesis only the basic point of departure was supported by the results that the speakers' own speech tempo really influences their tempo perception. However, the direction of this influence shows an interesting pattern involving significantly different behaviour for the various groups of speakers. The 'slow' and 'fast' speakers tend to behave perceptually similarly while 'moderate' and 'rapid' speakers tend to differ from the previous two groups. The standard deviation of the reaction time values show the same concurrence for the 'slow' and 'fast' speakers and for the 'moderate' and 'rapid' speakers.

– It has been supported that tempo perception depends primarily on articulation tempo. However, according to our findings, speakers/listeners perceive tempo significantly depending on *the activated levels of their whole perception mechanism.* If the upper levels of the speech perception mechanism do not play any role in the actual perception process, the tempo judgements (a) are closer to the actual physical parameters of the speech sample and (b) do not show big differences among the speakers having various own speech tempi. If the higher levels also participate in the decisions then other factors (contents of the speech samples, articulation of the speaker, lexicon of the speech sample, timbre, types of hesitations etc.) also play an important role.

– On the basis of the significant differences in perception and comprehension of various groups of speakers, we assume that various ways and storage systems should exist for the *intersections between the higher and lower levels of the speech perception mechanism* determined by the temporal organization of the speakers' speech production.

5. REFERENCES

[1] BEASLEY, D.S.; MAKI, H.E. (1976), "Time- and frequency-altered speech." In: *"Contemporary issues in experimental phonetics"*. Ed. LASS, N.J. New York, San Francisco, London. 419-458.

[2] BOVES, L. (1984), *"Perceptual ratings of speech tempo"*, Dordrecht: Foris.

[3] CUTTING, J.E.; PISONI, D.B. (1978), "An information-processing approach to speech perception", In *"Speech and language in the laboratory, school and clinic"*. Eds. KAVANAGH, J.F.; STRANGE, W. Cambridge, Massachusetts, London, 38-73.

[4] FELDSTEIN, S.F.; BOND, R.N. (1981), "Perception of speech rate as a function of vocal intensity and frequency", *Language and Speech, 24*, 387-395.

[5] den OS, E. (1988), *"Rhythm and tempo"*, Utrecht: Elinkwijk BV.

ACCOUNTING FOR THE REFLEXES OF LABIAL-VELAR STOPS

Bruce Connell

Phonetics Laboratory, University of Oxford

ABSTRACT

This paper presents a phonetic description, summarizing evidence drawn from different instrumental techniques, of the voiceless labial-velar stop as it occurs in Ibibio, one of the Lower Cross languages of SE Nigeria[1]. The description is then drawn on to offer an account of the variety of reflexes attested for labial-velars, both within Lower Cross and elsewhere. Important characteristics are that a) that the timing of the two articulatory gestures involved is asynchronous, and b) that the degree of asynchrony, as well as other aspects of their articulation, is variable, both across and within speakers. Recognition of this variation is the key to understanding the associated diachronic developments.

1. INTRODUCTION

1.1. Descriptions of Labial-velars

Labial-velar stops are relatively rare in languages of the world (cf. Maddieson [9]) and have received scant attention in the phonetic literature. Instrumental phonetic analyses have been presented by Ladefoged [8], Garnes [7] for Ibibio, and by Dogil [5] for Baule. Painter [11] also gives some discussion of labial-velars in an article dealing primarily with laryngeal mechanisms, Ward [14] for Efik presents kymograph tracings of [kp], but no systematic analysis, and finally, Ohala and Lorentz [10] present a general discussion of phonetic characteristics of labial-velar articulations, though without focussing on stops. In this paper, I summarize the results of a set of instrumental investigations that have been conducted on Ibibio [kp], and then use these results to attempt to account for the variation in reflexes found for labial-velar stops where diachronic change has occurred.

Apart from instrumental work, impressionistic descriptions of articulatory and auditory characteristics labial-velar stops can often be found in the Africanist linguistic literature. Generally, the labial and velar articulations are said to be simultaneous (e.g., Westermann and Ward [15]. Other than this, Ladefoged's [8] remarks (p. 12) in comparing labial-velar stops to velarized labials [p^w, b^w], and that they have a tendency to impart a labialized quality to following vowels, emphasize the possibility of perceptual confusion with labials, and Ohala and Lorentz [10] have provided acoustically based explanations for these tendencies. Comparisons have also been made to labial implosives by Ladefoged [8], Painter [11], and by Elugbe [6], who sees this as a general characteristic of labial-velars in the Edoid languages. Bearth and Zemp [1] describe the labial-velar stops of Dan as having "strong bilabial implosion", and Puesch [12] reports a voiced implosive labial-velar for Bekwil.

1.2. Diachronic Developments

The earliest account of diachronic correspondences of labial-velar stops in the literature is found in Westermann and Ward [15], who cite evidence for sound change that, "where kp or gb are weakened, it is the labial element which disappears and the velar element remains, sometimes reduced to x or ɣ" (p. 58). Elsewhere in the text (p. 108), correspondences are presented from the Nupoid languages Gbari and Nupe, and also from Bari and Kakwa (k^w ~ kp, g^w ~ gb in both cases) which to some extent confirm their conclusions. However, it is no difficult task to find instances of sound change involving labial-velars where it is the labial element which survives. It is probable that the velarized voiced labial implosive of some dialects of Igbo is a reflex of Proto-Igboid *gb. In the Lower Cross languages, PLC *kp has evolved into a variety of reflexes, most commonly [p], but also [b], [k^w], and possibly [gb] (and [kp] is retained in many instances).

[1] Ibibio is the largest of the Lower Cross group, Niger-Congo languages which are spoken mainly in Akwa Ibom and Cross River States of SE Nigeria. Connell [3] provides an overview of the group.

2. INSTRUMENTAL ANALYSES OF IBIBIO [kp]

2.1. Methodology

A variety of instrumental techniques were used to investigate the characteristics of the Ibibio labial-velar, including spectrography, laryngography, aerometry, and electropalatography. These were done during a period of approxi-mately three years, and used different speakers for the different investigations. Material for the spectrographic study was recorded by eight native speakers of Ibibio in Calabar, Nigeria, and analysed in the Phonetics Laboratories at the Universities of Ottawa and Edinburgh; further investigations were done in Edinburgh using primarily one speaker of Ibibio who was resident there (the aerometry was done with two speakers). Methodology and results are reported in greater detail in Connell [2, 4]

Spectrographic measurements were done to examine total duration of closure (TD), voice termination time (VTT) and voice onset time (VOT), as well as formant transitions and burst spectra, and compared to similar measurements for labials and velars. Laryngography (Lx) was done with both aerometry and electropalatography (EPG) to determine VTTs and VOTs and other information about phonation. The aerometry and EPG provided further details concerning the articulatory nature of these stops.

2.2. Spectrographic Analysis

In broad terms, the results of the spectrographic investigation confirmed those of Garnes [7]. This was true with regard to formant transitions, especially for CV transitions, where there was similarity to those of simple labials except for being steeper, having a lower locus, and apparently being more intense or stronger. This latter observation also corresponds with findings of Dogil [5] for Baule. On the other hand, VC transitions were variable, most often tending to resemble those of simple velars, but occasionally resembling labial transitions. Regarding the timing of the two gestures involved, evidence from transitions suggests a consistently later labial release, but variability as to which closure occurs first. Fig. 1 presents a spectrogram of the word [ékpê] 'leopard' demonstrating the asymmetry of formant transitions.

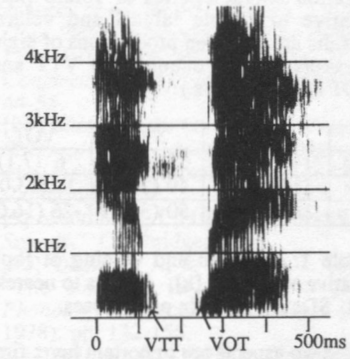


Fig. 1: Spectrogram of [ékpê] illustrating asymmetrical formant transitions of Ibibio [kp]. TD=156ms, VTT=38ms, VOT=-36ms. (Speaker E.E. Akpan.)

Noise in the signal (also evident to some extent in Fig. 1) made it difficult to determine burst spectra in my own data. However, indications are that the present work does not confirm Garnes' findings, which suggested that [kp] has a high frequency component (6 - 7 kHz) and weak energy spread throughout the higher frequencies (i.e., above 3.6 kHz), and, that energy in the lower frequency range was absent. The energy present in the spectra at release appears primarily in two areas - in the lower frequency range, i.e., below 1.2 kHz, and in the mid range, from 2 - 4 kHz. The lower concentration could indeed be a reflection of a labial

release, as would be expected, given the evidence from F2 transitions discussed above. The energy found in the mid-range could possibly be associated with a velar aspect of the release, but it is in this range that the noise band mentioned normally occurs. Finally, on occasion there is energy present throughout the spectrum, extending quite high in the frequency range. In this connection it is worth noting that Traill [13] reports burst spectra for labial clicks extending throughout the frequency range and being particularly strong in the 4 - 14 kHz range.

Table 1 summarizes voicing and duration characteristics of Ibibio [kp] relative to simple labials and velars. Results are based on productions of eight speakers (Fig. 1 exemplifies VTT and VOT measurements.)

	TD	VTT	VOT
p	147 (29.0)	65 (35.7)	6 (7.1)
k	113 (28.5)	49 (29.4)	21 (13.0)
kp	162 (28.6)	50 (35.9)	-26 (16.0)

Table 1: Duration and voicing of [kp] relative to [p] and [k]. Values to nearest ms. SDs are given in parentheses.

Two aspects are important here; first that Ibibio [kp] is prevoiced, and second, that the there is a relatively high amount of variation (as indicated by the standard deviations) in the voicing characteristics.

2.3. Aerometry/Lx

The aerometric work revealed a substantial pressure drop during closure and ingressive airflow, indicating use of either or both of velaric and glottalic ingressive airstream mechanisms. A relatively consistent variation in the pressure drop was taken as indicative of an earlier velar release (Connell [2, 4]). The associated Lx analysis confirmed and clarified voicing characteristics revealed by the spectrographic investigation.

2.4. EPG/Lx

Voicing characteristics described above were confirmed, and made more precise when considered against the EPG evidence of closure and release (Connell [4]). Also interesting was evidence from the EPG investigation confirming the spectrographic evidence of an earlier velar release. This was revealed through comparison of the EPG record with the accompanying audio signal. Since the audio signal in the set-up used was only a gross representation of intensity of the signal, its onset could represent either the onset of the following vowel or the onset of voicing, recalling that the release is prevoiced. Either way, given that the consonant is prevoiced, release of the velar closure prior to the onset of the audio signal would be a clear indication of the velar release preceding the labial one. This happened in all tokens, and on average 38ms, but ranging from 10ms to 80ms, prior to the onset of the audio signal. (SD = 15; calculations are based on 4 repetitions of 18 words containing [kp] in controlled environments.) Further research is planned, to monitor lip closure in conjunction with EPG, permitting a more accurate assessment of relative timing of both closure and release.

2.5. Summary

The various instrumental techniques revealed, among other characteristics, that: a) the two articulatory gestures are not totally simultaneous, nor completely synchronized: the velar release almost invariably precedes the labial one. There is more variability as to which closure occurs first, though this is most often the velar one; b) there is a considerable amount of voicing in this nominally voiceless stop, manifested in both a voicing tail, and a pre-release voicebar; and c) there is a high degree of variability in the timing of the various components of the articulation, both individually and relative to each other. This variation was manifested both within and across speakers. Finally, although evidence has not been presented here, there was some indication that the cross speaker variation observed correlated with dialect.

3. EVOLUTION OF PLC *kp

Proto-Lower Cross *kp has, in addition to [kp], the following reflexes across the group (Connell [3, 4]): [p, b, gb, k^w]. The phonetic characteristics of Ibibio [kp] allow us some insight into why such a range of reflexes should be manifested and, by extension, why yet others, such as [x, ɣ, $ɓ^w$] are also understandable.

The fact of the later labial release gives clear cause to expect a labial, or predominantly labial, reflex should the sound undergo change, as it would be the most salient. However, since the degree of asynchrony between the two releases demonstrated considerable variation, it is plausible to assume that a dialect of a language might exist where the two were much more closely simultaneous, or even with a later velar release; in these cases reflexes more predominantly velar might arise.

The variability in the duration of prevoicing in Ibibio also gives a clue as to why we find both voiceless and voiced reflexes; presumably those LC languages exhibiting PLC *kp > [p], originated in dialectal variation favouring a shorter voicebar, whereas those demonstrating PLC *kp > [b] would have emanated from ones with a longer voicebar. It is also possible that the existence of a relatively long voicing tail might have played a role in the development of voiced reflexes, particularly where PLC *kp > [gb] has been found.

An account of this nature fits the diachronic developments for Lower Cross based on the phonetic data for Ibibio. This implies that PLC *kp, at some stage in the history of the language was similar in its phonetic characteristics to that of Ibibio today. We might also expect that reflexes of labial-velars which are more predominantly velar (e.g., in the Nupoid languages cited above), or that are implosive (e.g., Igbo) came from parent languages whose labial-velars demonstrated characteristics conducive to those particular develop-ments. This is an empirical question which can, and hopefully will, be tested through a detailed phonetic analysis of language groups having the appropriate sets of reflexes.

4. REFERENCES

[1] Bearth, T., & Zemp, H. (1967) "The phonology of Dan (Santa)", *Journal of African Languages, 6*: 9-29.

[2] Connell, B. (1987) "Temporal aspects of labiovelar stops." in *Work in Progress, 20*: 53-60.

[3] Connell, B. (1990) *"Sound Correspondences, Lexicostatistics, and Lexical Innovation in the Lower Cross Languages."* Paper presented to the 20th CALL, Leiden, The Netherlands, Sept. 1990.

[4] Connell, B. (1991) *"Phonetic Aspects of Consonantal Sound Change in Lower Cross."* PhD, Edinburgh (in prep).

[5] Dogil, G. (1988) "On the acoustic structure of multiply articulated stop consonants (labio-velars)." *Wiener Linguistische Gazette, 42-43*: 3-55.

[6] Elugbe, B. O. (1989) *"Comparative Phonology of the Edoid Languages."* Port Harcourt: University of Port Harcourt Press.

[7] Garnes, S. (1975) "An acoustic analysis of double articulations in Ibibio." *Ohio State University Working Paper in Linguistics, 20 (Proceedings of the Conference on African Linguistics)*, pp. 44-55.

[8] Ladefoged, P. (1964) *"A Phonetic Study of West African Languages: an auditory - instrumental survey."* Cambridge: CUP.

[9] Maddieson, I. (1984) *"Patterns of Sounds."* Cambridge: CUP.

[10] Ohala, J. J., & Lorentz, J. (1978) "The story of [w]." *Report of the Phonology Laboratory, Berkeley, 2* (May 1978), pp. 132-155.

[11] Painter, C. (1978) "Implosives, inherent pitch, tonogenesis, and laryngeal mechanisms." *Journal of Phonetics, 6*: 249-274.

[12] Puesch, G. (1989) *"L'Opposition Implosive/Mi-voisées en Bekwil."* Paper presented to the 19th CALL, Leiden, The Netherlands, Sept. 1989.

[13] Traill, A. (1985) *"Phonetic and Phonological Studies of !Xóõ Bushman."* Hamburg: Helmut Buske.

[14] Ward, I. C. (1933) *"The Phonetic and Tonal Structure of Efik."* Cambridge: W. Heffer & Sons Ltd.

[15] Westermann, D., & Ward, I. (1933) *"Practical Phonetics for Students of African Languages."* London: OUP (for the International African Institute).

LASER-BEAM TECHNOLOGY IN DIACHRONIC PHONETIC RESEARCH AND ETHNOLINGUISTIC FIELD WORK

Tjeerd de Graaf

Department of Linguistics
Groningen University, The Netherlands

ABSTRACT

In recent years, laser beam technology has been used to reproduce the sound from wax phonograph cylinders and other old recordings. These methods have been used for the reconstruction of speech of the aboriginal population on Sakhalin which has been recorded in the beginning of this century. In this way, very useful data on earlier stages of languages and dialects have become available for linguists and anthropologists.

1. USE OF THE PHONOGRAPH

The principle of the original phonograph is simple: a metal horn focuses the energy of the sound waves onto a thin diaphragm, which supports a small needle in its centre. When the diaphragm vibrates in response to the energy of the focused sound waves, the needle, too, vibrates as it is drawn across the revolving surface of a wax cylinder. The needle cuts a groove consisting of microscopic gouges in the soft cylinder surface. In this way, a recording of the pattern of sound waves is made. To play back the recording, the needle is replaced over the gouges made during the registration. The attached diaphragm vibrates and creates sound waves duplicating those which had originally been recorded.

From the early use of the phonograph until the coming of portable disc-recording equipment, the phonograph was the only means of recording phonetic data. In the late 1880s, ethnographers were intrigued by the possibilities of applying the new cylinder phonograph for field work. It was used for the first time around 1890 for the study of American Indian speech and in the beginning of this century Ainu data were recorded on the island of Sakhalin, north of Japan.

2. WAX CYLINDERS AND THE PROJECTS FOR THEIR RESTORATION

Old recordings on wax cylinders are still being kept in many places. It would be of great interest to regain the sound material they contain, and to improve its quality by using modern digital techniques of registration and signal enhancement.

The events which have led to the project on the retrieval of sounds from wax cylinders started in Poland with the discovery of a number of phonographic wax cylinders. They contain linguistic, musical and ethnographic material, primarily on the Ainu people of Sakhalin. The recordings were made at the beginning of this century by the Polish anthropologist Piłsudski [1].

The Institute of Linguistics of Poznań University (Poland), Hokkaido University (Japan) and the Institute of Linguistics of Groningen University started a common project to analyse material obtained from the phonograms and other old recordings and to set up ethnolinguistic field work. The goals of this collaboration are the following:

a. Application of acoustic, electronic and optical engineering techniques to the retrieval of information on phonographic wax cylinders and other old recordings;
b. The interpretation of the phonetic and linguistic contents of the recordings, and the study of the languages of Sakhalin at the beginning of this century;
c. Phonetics and ethno-musicological analysis of the recorded speech and songs; comparison with present day material.
d. The organisation of field work expeditions to the Minority Peoples of the North in the USSR.

3. THE REPRODUCTION SYSTEM

Using the original Edison-type phonograph for the reconstruction of the sound material involves a risk of damaging the wax cylinders. This method cannot be applied to broken cylinders which have been repaired. Thus a non-destructive, non-contacting method has been developed on the basis of laser-optics technology.

A Gaussian beam emerging from the single-mode He-Ne laser with a wavelengthe of 0,633 μm is focused by an objective lense. The wax cylinder, which is translated during rotation, is illuminated by a diverging Gaussian beam of which the spot diameter on the cylinder can be adjusted to the width of the grooves. The detecting plane for the reflected beam is set perpendicular to the optical axis. The wax cylinder is rotated, whereas the intersection position of the reflected ray on the detecting plane moves in time on this plane. The time variation of the position is detected by a position-sensitive device and it corresponds to the acoustic signal. The signal stored can be deduced from the detected variation of the reflected beam.

The properties of the sounds reproduced in this way depend on the width of the illuminating laser beam, since its finite size breaks down the principles of geometric optics on which the method is based. Further, there is the obstructive noise in the sound, caused by using a coherent laser beam for illumination, and there is the tracking error resulting from improper contact with the grooves. These problems have been investigated experimentally by Asakura et al.[2], who found that the most suitable beam width for the laser-beam should have a spot diameter between 80 and 100 μm. In this way, the sounds reproduced can be heard naturally and without obstacles. Since the laser-beam reflection method is non-contacting and non-destructive, it is a powerful tool for retrieving sounds from old wax cylinders without damaging them.

4. SIGNAL ENHANCEMENT

The data are stored on new optical/digital sound carriers and in order to improve the quality of the sound obtained, special techniques have been developed. The sound reproduced from old recordings is usually of poor quality. This may be caused by the original recording

techniques (e.g. resonances in the horn), by the damage of the cylinders which has occurred over the years (clicks at burst positions) and by the techniques of reconstruction. In order to improve the sound quality, several methods have been developed which can also be applied to speech enhancement in general. In Japan, various programs have been developed and applied for this purpose [3].
In the case of the Ainu tapes, the result of the processed sounds was not always satisfactory: in several cases, the listeners preferred the original unprocessed sounds, even if there was noise on the tape. This was due to the fact that after processing the noise level is reduced, but certain bad-quality-features are still there and become more prominent. The recorded and processed Ainu data are stored at the Research Institute of Applied Electricity, Hokkaido University (Japan).

5. RESULTS OF THE WAX CYLINDER RESTORATION PROJECTS

The Japanese project has provided the possibility to study the Ainu language from Sakhalin as it was spoken at the beginning of this century. Originally, the Ainu people lived in the Northern part of Japan, on Sakhalin and the Kurile Islands; at present their language is only spoken on Hokkaido. Old Ainu people were consulted when the material from the Pilsudski wax roles was played to them. In some cases, they recognized their Ainu dialect and the voices from the past.
In this way, the last stages of a dying language have been safely recorded. The material can be studied by linguists and ethnologists in order to obtain information on the Ainu people. The wax cylinders and their contents can thus be considered to be part of a very important cultural heritage, because they contain valuable sound data of speech and songs of the Ainu people that were lost long ago.

6. THE ETHNOLINGUISTIC FIELD WORK ON SAKHALIN.

In July and August 1990, the University of Hokkaido has organized an international fieldwork expedition to Sakhalin in order to study the language situation of the original population on that island and the way this has been influenced by Japanese and Russian.
The idea was to look for the Ainu population and to investigate the status of the other small minority groups, in particular Nivkh (Gilyak), Uilta (Orok) and related Tungusic peoples, who were the first inhabitants of Sakhalin. Unfortunately, during our expedition no more Ainu people could be found, and the only person who represented the Ainu language and culture from Sakhalin was probably the old informant we met on Hokkaido.
The original population of Sakhalin consisted of some Paleo-Siberian and Tungusic peoples, in particular the Nivkh and Orok in the North and Centre, and the Ainu in the South. Their numbers were rather small, and during the colonisation process by the Russians from the North and by the Japanese from the South, they were soon numerically dominated by these stronger nationalities. Due to their isolated life as hunters and fishermen, they were able to keep their native language and culture for a long time, but since the beginning of this century the assimilation process has gradually become stronger.

The dramatic events of 1945, culminating in the Soviet occupation of the whole island, had enormous consequences for the ethnographic and linguistic situation on the island: practically all Japanese inhabitants and together with them many of the aboriginals, left Sakhalin for Japan. New imigrants came from all parts of the Soviet Union and at present, more than 100 nationalities are living on the island. Several of them still cultivate their own language.
During the expedition to Sakhalin, a great deal of material on the Minority Peoples of the North was collected: about 80 hours of audio, 30 hours of video and numerous photographs and written documents. Part of the recordings consists of interviews with representatives of the minorities of Sakhalin. These interviews can be considered as 'case studies' of the language situation for particular minorities. Most of the material is related to the Nivkh population.
The life of the Nivkh and other Minority Peoples of the North has changed considerably under the influence of Russification. They have become a small minority on Sakhalin, scattered over the island and surrounded by Russians and other immigrants who take part in Russian culture.
During our field work expedition on Sakhalin, most of the subjects for our project were elderly people with a strong motivation to use their language. Practically all young people we met no longer had an active knowledge of the language, and they only communicated in Russian.
It can be concluded that on Sakhalin a process of assimilation is taking place, which may result in the complete disappearance of these small languages and cultures. This process of "language death" may, however, slow down, if these minority cultures are receiving more attention. Further field work should be conducted in order to facilitate the conservation of data on the languages and cultures of these Peoples of the North and other minority groups.
The data collected about these aboriginals (Nivkh, Orok and others) are now analysed and a description is given. The availibility of these data will enable a comparison with the historical recordings.

6. REFERENCES

[1] ASAKURA, T. et al. (ed.) (1986), *Proceedings of the International Symposium on B.Pitsudski's Phonographic Records and the Ainu Culture*. Hokkaido University, Sapporo, 1985.
[2] ASAKURA, T. et al. (1986), "Reproduction of the Sound from old Wax Phonographic Cylinders using the Laser-Beam Reflection Method". *Applied Optics, 25*, 597.
[3] DE GRAAF, T. (1989), "Reconstruction, Signal Enhancement and Storage of Sound Material in Japan". *Proceedings of the 2nd Int. Conference on Japanese Information in Science, Technology and Commerce*. Berlin, pp. 367-374.

INTER-SPEAKER VARIABILITY IN SIBILANT PRODUCTION AND SOUND CHANGE INVOLVING SIBILANTS*

Alice Faber

Haskins Laboratories, New Haven, CT 06511 USA

ABSTRACT

The role of contextual and inherent lip rounding in mediating synchronic and diachronic interchanges between [s] and [ʃ] is described in the context of a model of sound change involving inter-speaker variability in speech production and perception strategies.

1. INTRODUCTION

In a variety of the world's languages, interchanges of [s] and [ʃ] in labial environments, that is, near sounds like [r m u ʕ (ɬ)], are attested. Such changes are seen in some ancient Semitic languages [3], Tigriña (a modern northern Ethiopian Semitic language) [12], early Indo-European [7], and southern American English [9]. Based on published studies showing that [ʃ] is rounded in a variety of languages [1,2,5], that [s] is lower frequency adjacent to [u] than to other vowels [4], and that the frequency boundary between [s] and [ʃ] is higher before [u] than before [a] [6], I attributed the changes in both directions to the phonologically ambiguous status of a phonetically rounded sibilant in a rounded context: if the lowered frequency is attributed to the rounded context, the sibilant is interpreted as [s]; but, if, instead, the lowered frequency is interpreted as inherent to the sibilant, the sibilant is interpreted as [ʃ] [3]. Sound change, in this model, is a result of individual differences in speech production and perception: Speakers will differ in how much rounding, inherent or contextual, they produce in a given instance, and listeners will differ in their interpretation of rounding in a specific instance as inherent or contextual.[1]

This paper is a tentative report of a preliminary series of experiments aimed at testing the hypothesis regarding [s]-[ʃ] interchanges as well as the general model of sound change in which that hypothesis is embedded. The present hypothesis is that [s] and [ʃ] will be less distinct acoustically in some labial contexts than in non-labial contexts for at least some speakers of some languages. Furthermore, this decreased distinctiveness should result in part from increased rounding of [s] in these labial contexts.

2. METHODS

There were 7 subjects for these experiments, including one Polish-English bilingual. All speakers produced utterances of the form *VSV*, with the flanking vowel {a i u}, and S one of {s ʃ} (vowel tokens), and the bilingual speaker in her Polish mode produced utterances with [ʂ ɕ]; the utterances with [ʂ] will not be discussed here, and [ɕ] will be treated as equivalent to [ʃ]. Two speakers also produced *aC(a)Sa* and *aS(a)Ca*, utterances with S again one of {s ʃ}, and C one of {k l m r} (consonant tokens). All tokens had penultimate stress. Subjects, their language backgrounds, and their data are summarized in Table 1.

Lip position was monitored with a modified Selspot opto-electronic tracking system. For some subjects, linguo-palatal constriction location was monitored with a RION artificial electropalate. Movement signals were digitized at 200 samples/sec., and the EPG signal at 64.1 frames/sec. The speech signal, recorded on a Telex unidirectional head-set microphone, was digitized at 20,000 samples/sec. (12 bits), without preemphasis. Automatic peak picking algorithms were used to identify upper lip protrusion (ULP) maxima and minima; for segments without clear extremes, an arbitrary point was measured. For present purposes, the measure of sibilant frequency used was the Centroid. Centroids were computed for all sibilants over the range c. 1,000–10,000 Hz.[2] In vowel tokens, one centroid was computed, at the approximate midpoint of S. For consonant tokens, three centroids were calculated, one in the middle, and one at each edge.

Table 1: Summary of subjects and data. Under Data Set, *vowel* means *VSV* utterances, and *cons aC(a)Sa* and *aS(a)Ca* utterances.

Subject	Language	Data Set	# of tokens
EF	Italian	vowel	50
JM	German	vowel	50
DR	Catalan/Spanish	vowel	50
ED	Polish	vowel	10
	English	vowel	10
EVB	English	vowel	40
KSH	English	vowel + cons	10
FBB	English	vowel + cons	10

For each speaker, the extent of variation in sibilant frequency and in ULP was assessed via Analyses of Variance. For vowel conditions, the analysis was Vowel X Sibilant, and for consonant conditions, Consonant X Sibilant X Order X Adjacency, with Measurement Point as a repeated factor for the acoustic ANOVA. Order refers to whether C preceded or followed S, and Adjacency to whether S and C were abutting or separated by a stressed [a].

3. RESULTS AND DISCUSSION

3.1 Acoustic Factors

All vowel condition speakers had significant main effects of Sibilant and Vowel, as well as significant interactions. Centroids were lower for [ʃ] than for [s], and were lower when the flanking vowel was [u] than in the other contexts. Fig. 1 shows the extent to which [s] and [ʃ] were distinct in the three vowel contexts.[3] The higher the percentage, the larger the frequency difference between [s] and [ʃ]. There are clear differences among subjects in the extent to which they distinguish the two sibilants, as well as in the nature of the vocalic effects on the distinction. The primary commonality is that the two sibilants are less distinct in the [u] context than in the other vocalic contexts. Examination of the raw centroid values reveals, further, that the decrease in sibilant distinctiveness in the [u] context results primarily from a decrease in frequency for [s]. Decreases in distinctiveness in the [i] context result, in contrast, from an increase in frequency for [ʃ].

Analyses for the consonant subjects are, not surprisingly, more complicated.

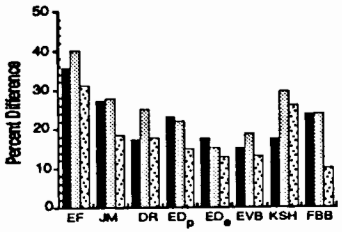


Figure 1: The percent difference between [s] and [ʃ] in mean centroid frequency for eight subjects. Dark bars are values for the [i] context, light bars for the [a] context, and speckled bars for the [u] context.

Both had significant main effects for Sibilant, Consonant, Order, Adjacency, and Measurement Point, as well as many interactions. The latter result is not surprising, since differential effects of context consonants on sibilants should be strongest at the sibilant edge adjacent to the context. In any case, [s] was higher frequency than [ʃ]. For KSH, sibilants were lower cooccurring with [r m] than with [k l], while for FBB, sibilants in tokens with [r l] were lower than those with [k m]. For KSH, Consonant effects on sibilants were stronger for [s] than for [ʃ], while for FBB the reverse was true. For both subjects, the sibilant was lower frequency at the third measurement location than at the earlier two.

The frequency differences between [s] and [ʃ] in the different consonantal contexts are illustrated in Figs. 2 and 3 In brief, these figures show that [s] and [ʃ] are slightly less distinct in the bisyllabic forms (in which S and C abut) than they are in the trisyllabic forms. For both subjects, one of the largest distinctions between the sibilants occurs, surprisingly, *after* [r]. For FBB, but not for KSH, there is a marked decrease in sibilant

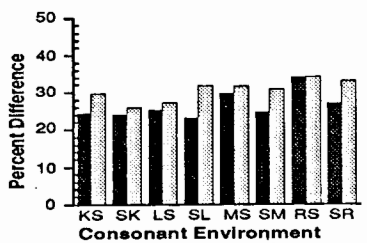


Figure 2: The percent difference in mean centroid frequency between [s] and [ʃ] for subject KSH. The *x*-axis legend shows the relative order of context consonant and sibilant. Dark bars represent sibilants adjacent to the consonant, and light bars sibilants separated from it by a stressed vowel.

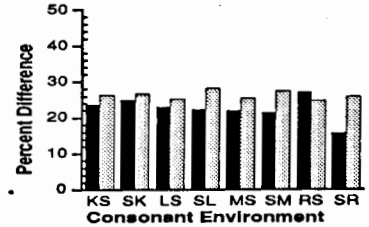


Figure 3: The percent difference in mean centroid frequency between [s] and [ʃ] for subject FBB. See Figure 2 caption for further details.

distinctiveness before [r]. As Fig. 1 shows, she also has a clear decrease in sibilant distinctiveness in the [u] context.

3.2 Articulatory Factors

While it is customary to attribute differences in sibilant acoustics to differences in linguo-palatal constriction (a more anterior constriction producing a higher frequency sibilant [10, 11]), we will first consider the effects of differential ULP on the acoustic differences observed so far. For all subjects, there were significant main effects of Sibilant and Vowel. There was consistently more ULP for [ʃ] than for [s], and more for sibilants next to [u] than for those next to [i] or [a]. As is clear from Fig. 4, which shows the percentage difference between normalized[4] ULP for [s] and [ʃ], subjects differ markedly in the extent to which they have a ULP difference between [s] and [ʃ] and in the extent to which the difference is sensitive to vocalic context. Subjects who, like JM, have a ULP contrast between [s] and [ʃ] in a rounded context have more protrusion for [ʃ] in such a context than in an unrounded context. Other subjects, like EF, neutralize the protrusion contrast between [s] and [ʃ] in a rounded context.

The two consonant subjects had somewhat different patterns of ULP. Subject KSH, as shown in Fig. 5, has less contrast in ULP in the bisyllabic tokens than in the trisyllabic ones. She further has less contrast in syllable initial sibilants than in syllable final ones, especially before [m] and [r], the consonants most likely to induce ULP in a preceding [s]. In contrast, FBB has more ULP contrast in trisyllables than in bisyllables, as shown in Fig. 6. This pattern makes it all the more striking that she, too, has much less contrast between the sibilants immediately preceding [m] or [r] than in the other contexts.

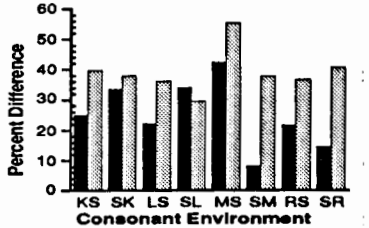


Figure 5: The percent difference in peak lip protrusion between [s] and [ʃ] for subject KSH. Conventions are as in Figure 2.

In some instances, there is a clear relationship between the acoustic and labial patterns of variation. These are, for the most part, instances in which the context segment is labial. For 6 of the 8 vowel condition subjects, the decrease in magnitude of the acoustic difference between [s] and [ʃ] is commensurate with the decreased distinction in ULP. However, two subjects, EF and KSH, have nearly as much acoustic difference between the two sibilants in the [u] context as in the other contexts, but virtually no difference in ULP. Since these are two subjects for whom EPG data are available, the linguo-palatal basis for the distinction can be identified. For KSH, both [s] and [ʃ] are retracted in the [u] context, and both are, therefore, lower frequency; the distinction is thus preserved.

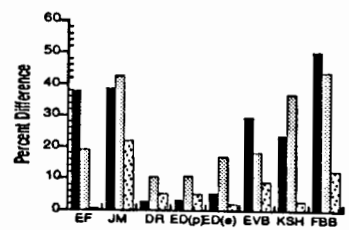


Figure 4: The percent difference in mean upper lip protrusion between [s] and [ʃ] for 8 subjects. See Figure 1 caption for further details.

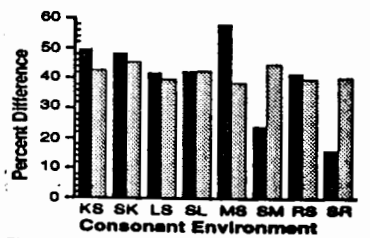


Figure 6: The percent difference in peak lip protrusion between [s] and [ʃ] for subject FBB. Conventions are as in Figure 2.

EF, in contrast, has a more *anterior* [s] in the [u] context than elsewhere, thus counteracting the frequency-lowering effects of the increased ULP.

The two consonant subjects have congruent acoustic and labial variation patterns in utterances containing [m r], but less congruent ones in utterances with [k l]. This is sensible, given that the former contribute active ULP gestures while the latter do not. Unfortunately, no EPG data are available for FBB. However, interpretation of KSH's EPG data is straightforward here too. First, [s] is retracted following [r], while [ʃ] is not affected by which side of [r] it is on; this decreased distinctiveness reinforces the effects of the decrease in ULP. Fig. 5 shows an additional pattern of differential effects of preceding and following [k] and [l] on ULP distinctiveness, a pattern not reflected in acoustic differences: [s] and [ʃ] are less distinct in ULP following [k] or [l] than preceding them. However, the EPG data show the reverse: [s] and [ʃ] are more distinct in linguo-palatal constriction location following [k] or [l] than preceding them. The effects of this linguo-palatal difference apparently cancel out those of the ULP difference.

4. CONCLUSION

The experimental results described above show that the decreased acoustic distinctiveness of sibilants in rounded contexts is the result of a complex interplay of labial and linguo-palatal factors. Speakers vary in the relative contributions of the two sorts of factors to their acoustic patterning. In order to preserve the acoustic invariance of [s] across a range of contexts, speakers must vary their linguo-palatal targets for [s], in an attempt to compensate for context-based variation in lip position. In contrast, preservation (to the extent possible) of articulatory invariance, at least as regards linguo-palatal constriction, leads to an increase in acoustic variability. Concentration on articulatory invariance can lead to instances of [s] that might be perceived as [ʃ]. But, concentration on acoustic invariance can lead to a proliferation of articulatory targets, the number of which may subsequently be reduced in novel ways, especially the reinterpretation of some instances of [ʃ] as [s]. Either way, sound change occurs.

5. REFERENCES

[1] ABRY, C. & J.-L. BOE, 1986, "Laws for Lips", *Speech Communication*, *5*, 97–104.

[2] BROWN, G., 1981, "Consonant Rounding in British English", *Towards a History of Phonetics*, R. E. Asher & E. J. A. Henderson, eds. Edinburgh: University of Edinburgh Press, pp. 67–76.

[3] FABER, A., 1986, "On the Actuation of Sound Change", *Diachronica*, *3*, 163–184.

[4] HEINZ, J. M. & K. N. STEVENS, 1961, "On the Properties of Voiceless Fricative Consonants", *JASA*, *33*, 589–596.

[5] MALMBERG, B., 1963, *Phonetics*, NY: Dover Press.

[6] MANN, V. A. & B. H. REPP, 1980, "Influence of Vocalic Context on Perception of the [s]-[ʃ] Distinction", *P&P*, *28*, 213–228.

[7] MARTINET, A., 1955, *Économie des changements phonétiques*, Berne: A. Francke.

[8] OHALA, J. J., 1989, "Sound Change is Drawn from a Pool of Synchronic Variation", *Language Change*, L. E. Breivik & E. H. Jahr, eds. Berlin: Mouton de Gruyter, 173–198.

[9] PENNINGTON, M. C., 1982. *The Story of S*, Philadelphia: University of Pennsylvania Ph.D. Dissertation.

[10] STEVENS, K. N., 1972, "The Quantal Nature of Speech", *Human Communication*, E. David & P. Denes, eds. NY: McGraw Hill, 51–66.

[11] STEVENS, K. N., 1989, "On the Quantal Nature of Speech", *J. Phon.*, *17*, 3–45.

[12] VOIGHT, R. M., 1988, "Labialization and the So-Called Sibilant Anomaly in Tigrinya", *BSOAS*, *60*, 525–536.

6. NOTES

* Research support from NIH grant DC–00016 is gratefully acknowledged.

[1] This view of sound change is like that of Ohala [e.g., 8]. The primary difference is that Ohala sees listeners' varying interpretations as *errors*, while I attribute them to inherent indeterminacies in the process of speech perception.

[2] This range was selected to minimize the influence of partial voicing of some tokens. In addition, there was a small amount of relatively low frequency interference as a result of the recording set-up, and it was necessary to reduce the impact of this interference on the centroids.

[3] Numbers for this and following figures were derived by the formula $n = 100 * \frac{x_1 - x_2}{x_1}$ with x_1 the larger of the two inputs.

[4] Normalization took place in two steps. During the analysis for each subject, raw ULP values were normalized with reference to a fixed rest position. Later, the lowest mean value for each subject was set to 5 mm, so that the variability apparent in Figure 4 reflects differences in subjects' articulatory patterns rather than scaling.

VOICE QUALITY AND MOULDING OF PHONOLOGIES : A SUBSTANTIAL EVIDENCE

Bharati Modi

Dept. of Linguistics, M.S.University,
Baroda, India

ABSTRACT

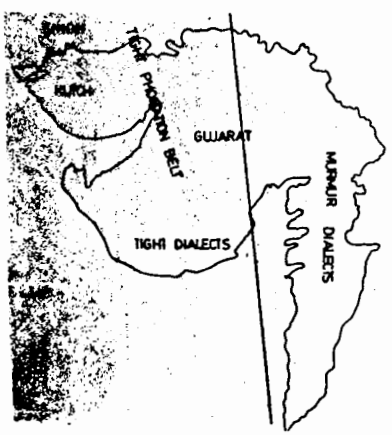
The issue of voice quality/ phonation types are not matters solely of individual peculiarity or emotional expression and thus safely to be ignored (Henderson)[2]. This paper while studying this issues with the help of tomograms proposes their relationship with phonology as manifested particularly in Gujarati and in some of the Western Indo Aryan languages such as Sindhi, Kutchi and Marathi.

1. INTRODUCTION

Though it is accepted that the study of phonological systems have to presuppose the study of phonetic substance we still have to know which phonetic aspects should be considered relevant to phonology. This paper shows how voice quality and phonation types are essential phonic material to understand the phonology of language and extends some substantial support to Henderson's [2] views. The observation regarding the relationship between voice quality/phonation types makes it imperative for us to revise our descriptive apparatus. It is suggested that voice quality/phonation types can control and mould the phonology of language. This is once again like the age old relationship between nature and culture and description which deprives the problem of one of the relata is doomed (Mol and Uhlenbeck)[10].

2. PHONATION TYPES IN GUJARATI

Gujarati very interestingly employs two distinct phonations: Murmur and Tight. Murmur has been taken for granted as feature associated with Gujarati since Pandit [11] and Jørgensen [4]. But the fact is that 40% of Gujarati speakers speak with tight phonation. These phonation based distinction of dialects coincides with the geographical divisions of Gujarat (See Map 1).



MAP 1

Dialects of Gujarati based on phonation types and tight phonation belt

These phonation types are like physiological habits of the speakers who with the vertical movements of larynx and muscular tension modulate the airflow. Murmur (breathy voice) occurs due to [ɦ] which can be independent phoneme or can be voiced aspiration of [bɦ, dɦ, ɖɦ, dʒɦ, gɦ].Murmur has constant airflow due to cartilaginous gaps (Fujimura)[1] and strong activity of posterior cricoarytenoid along with maintained vibration of vocal folds (Sawashima and Hirose)[2], (Hirose and Gay)[3]. Tomograms taken of Gujarati speakers show the lowered position of larynx for murmur and raised position for tight. Raised larynx increases the tension of vocal fold surface (Stevens)[14]. Tight phonation is a physiological adjustment maintained through out the speech and has a high pitched quality. Having the reverse physiology from murmur this phonation automatically inhibits murmur. Recently conducted preliminary acoustic study of these phonations (Schiefer et al)[13] has been able to distinguish these two phonations with the help of two efficient parameters; amplitude of first and second harmonics and band widths of F1 and F2.

By general phonological criteria these laryngeal dimensions can be discarded as irrelevant to phonology. However their behaviour in Gujarati language opens up a new direction in studies of Gujarati phonology. Modi's studies [8],[9] were done with the intention of showing the non-segmental character of murmur. Murmur was considered a prosodic phenomenon interacting with the surrounding sonorant sounds such as :

1) ɦ + V 2) V + ɦ 3) V + ɦ + V

Here 'ɦ' gets partly deoralized (See Fig.1)

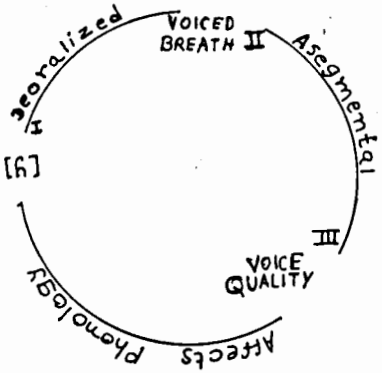


Fig. 1
Behaviour of 'h'

Tight phonation dialects (TD) due to musculature tension show the tendency towards fortition process as opposed to murmur dialects (MD), where there is conspicuous laxing having lenition effects. It is suggested that constant pulling of these opposite tendencies in the language act like controlling factor on the 'normalization' of phonology. MD and TD are further subdivided depending upon their having six vowels [i, e, ə, a, o, u] or eight vowels [i,e,ɛ, ə, a,ɔ, o, u]. But Modi [8] has considered only six vowel-norms for Gujarati.

Perceptually different

/e/ /o/

Articulatory optimization

e o jaw lowering

ɛ ɔ

æ ɒ

After Lindblom [7]

In TD the mid-vowels are higher than in MD. The fortition tendency is considered responsible for this (Modi)[8]. (It is worthwhile noting here that the speakers of TD with six vowels face great difficulty in pronouncing English [ɛ, æ] and [ɔ]. They

are a laughing stock of all Indians for having [e] vowel for 'rape' and 'wrap'). It is suggested that a process might have begun when the distance between [ɐ-ɛ] and [o-ɔ] could have become phonemic but tight phonation might have counter balanced such a shift.

The next issue is that of nasalization, which has resulted from diachronic N-loss, nasalization (except for some onomatopoetic forms). In MD denasalization is under progress. But TD hold the fort of fortition. The tense musculature of tight phonation once again is favourable to nasalization and hence denasalization remains only as a subdialectal phenomenon.

One more of such phenomena is of voiced stops spirantizing in MD intervocalically or when in cluster with liquids: e.g. [aβɾu] 'prestige',[aɣ̞o] 'latch',[saɦai] 'simplicity'. TD with inherent fortition does not allow such weakening of stops.

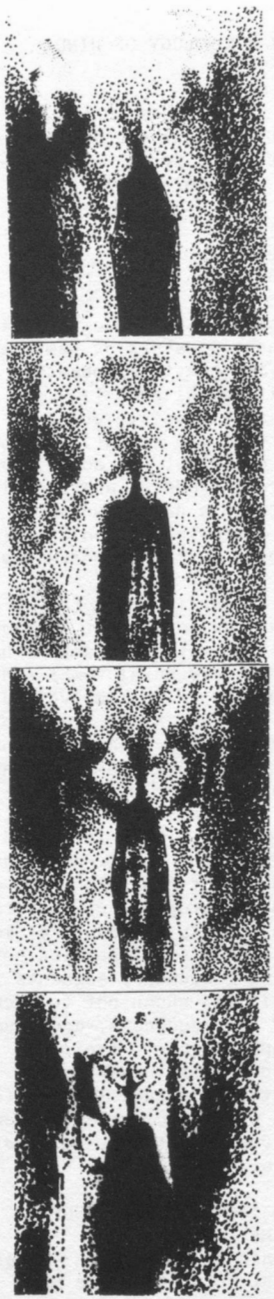
In short, denasalization and spirantization are prohibited from pervading the complete language. The tenseness and fortition of TD act as a preserving factor while as laxness and lenition of MD act as a weakening factor. Both phonation types work hand in hand : retaining-substituting preserving-effacing; thus balancing the phonemic inventory of language. They are relevant linguistic features as if purpose-built.

3. TIGHT PHONATION IN SINDHI AND KUTCHI

A little more support is extended to the hypothesis from two other Indo-Aryan languages: Sindhi and Kutchi. It is proposed that there is a tight phonation belt starting from Sindh (now in Pakistan) and spreading upto Northwest Gujarat (See the Map). Both these languages have implosives for which Ladefoged [5] notes a slight downward movement of larynx after the closure is formed. Lodefoged [6] has seen the possibility of different phonations in languages having implosives. The conflicting gestures of raising and lowering of larynx can be possible due to the musculature tension of tight phonation. It is interesting to note that this phonation has added totally non Indo-Aryan sounds-the implosives-to the phonemic inventory of these languages (See tomograms of Sindhi and Kutchi speakers). The suggestion is that due to these sounds a distance has been created between the phonemic systems of these languages and other Indo-Aryan languages.

4. VOICE QUALITY AND DIACHRONICALLY ACQUIRED PRECISION OF SOUNDS

Finally it is shown how the maintenance of the precision of sounds is attained by the tension of tongue musculature in one of the standard dialects of Marathi spoken by Poona Brahmins with exceptional scholarship in Sanskrit. They attained perfection in uttering Sanskrit sounds by following the ancient phonetic treatises. They formed a speech habit where fronting and raising of tongue with tongue musculature tension was sustained throughout the speech. The oral cavity gets reduced and sounds are marked by 'fortisness'. The habit was so much entrenched into the system of the community that it got transferred into their Marathi. It is proposed that this voice quality has played a very important role in moulding of Marathi phonology; compared to any other Indo-Aryan languages Marathi has retained maximum sanskritic sound sequences. The normal diachronic tendencies of weakening such as, deletions of final vowels, cluster simplification, media vowel reduction found in all other Indo Aryan languages are totally absent in Marathi.



Tomograms:Gujarati:1=MD, 2=TD; 3=Sindhi, 4=Kutchi.

5. REFERENCES

[1]FUJIMURA, O.(1977), Control of larynx in speech, Phonetica, 34, 280-288.

[2]HENDERSON, E.J.A.(1977),The larynx and language:A missing dimension, Phonetica, 34, 256-263.

[3]HIROSE, H. and T. GAY(1972), The activity of intrinsic laryngeal muscles in voicing control, Phonetica, 25, 140-164.

[4]JØRGENSON, F.E.(1967),Phonetic analysis of breathy vowels in Gujarati, Indian Linguistics, 28 70-138.

[5]LADEFOGED, P.(1975), A course in phonetics, New York, Harcourt Brace, Jovanovich, Inc.

[6]----(1981),Preliminaries to linguistics. The midway reprint,Chicago, The University of Chicago Press.

[7]LINDBLOM, B.(1971),Phonetics & description of language.Proceedings of VIIth ICPhS, The Haugue, Mouton.

[8]MODI, B.(1983),Some issues in the phonology of Gujarati, Ph.D. dissertation,Baroda,M.S.Univ.

[9]----(1984),Laryngeal dimensions in Gujarati phonology, In Vth Int.Phonol.meet Eisenstadt,Vienna Ling. Gaz. Supp. 3.

[10]MOL, H. and E. ULHENBECK(1959),Hearing and the concept of phoneme,Lingua, 4, 161-185.

[11]PANDIT, P.(1957),Nasalization aspiration and murmur in Gujarati Indian Linguistics, 17, 165-172.

[12]SAWASHIMA, N. and HIROSE, H.(1968),New laryngoscopic technique by use of fiber optics, JASA 43(1), 168-169.

[13]SCHIEFER, L.,CH.LANGMEIER,U. LUDERS and B.MODI(1987),An acoustic study on murmured and tight phonation in Gujarati dialects : A Preliminary report.Proc.of XIth ICPhS,Tallin,Estonia,USSR,Aug 1-7.

[14] STEVENS, K.(1977), Physics of laryngeal behaviour and larynx modes, Phonetica, 34, 264-279.

EPENTHETIC NASALS IN THE HISTORICAL PHONOLOGY OF HINDI

Manjari Ohala
Linguistics Program
San Jose State University
San Jose, CA 95192

John J. Ohala
Department of Linguistics
University of California
Berkeley, CA 94720

and:
Department of Linguistics
University of Alberta
Edmonton, Alberta T6G 2E7

ABSTRACT

We demonstrate the plausibility of a posited historical process whereby an epenthetic nasal consonant appeared between a sequence of nasal vowel + voiced stop (but not if the stop was voiceless) by showing that the same process occurs phonetically in present-day Hindi and French pronunciation.

1. INTRODUCTION

Modern Hindi (MH) words such as [dãṭ] "tooth vs [tʃãnd] "moon" present an interesting asymmetry in their phonological history: in their development from Middle Indo-Aryan (MIA) to Old Hindi (OH) and then to New IA both were subject to cluster simplification with compensatory lengthening and nasalization of the preceding vowel [1, 4]. Thus: Skt danta > MIA danta > OH dãta > MH [dãṭ]; Skt čandra > MIA čanda > OH čãda > MH [tʃãnd].[6] (Historical forms are given in conventional transliteration; modern forms in IPA where [ɑ] is inherently long.) In the latter example the nasal consonant, present in MIA but then subsequently lost, re-appears in MH. Is it plausible that a nasal be re-introduced only before a voiced stop or should we rethink the historical derivation of such words? The primary evidence that the nasal was indeed lost by the time of OH is the fact of compensatory lengthening of the vowel which in numerous other instances correlates with simplification of medial consonant clusters or geminates, e.g., Skt hasti "elephant" > Prakrit hatthi > MH [hɑtʰi]. We present phonetic evidence in support of the scenario that a nasal consonant (N) could have been re-introduced preferentially between a nasalized vowel (Ṽ) and a following voiced stop (D) but not a following voiceless stop (T).

In an earlier exploratory study of Hindi we found that in the transition between a word final distinctively nasal vowel and a following word initial voiced stop, the initial part of the voiced stop became a nasal consonant. For example, the Hindi utterance /ek mẽ do / (literally) "one 'I' give" was phonetically [ek mẽ ⁿdo]. Here it seemed clear that the nasal consonant formed out of the first part of the voiced stop was not lexical and was purely a product of low-level phonetic interaction between cross-word boundary segments. If verified, shown not to occur with Ṽ + T sequences, and found in other languages too, then this epenthetic nasal would constitute a plausible parallel to the posited diachronic scenario which requires the creation of a N out of a sequence of Ṽ + D.

2. AN INSTRUMENTAL STUDY

2.1. Methods

To obtain an indication of velic movement in speech in a non-invasive way we used a nasal olive [10] which gives a rough measure of nasal air flow, itself an approximate measure of velic opening. The nasal olive records air pressure behind one blocked nostril, the other nostril remaining open. This technique also permits a high-quality audio recording of the speech to be made simultaneously. Our subjects were two native speakers each, of Hindi and French; for both languages there was one male and one female speaker. The first author was the female Hindi speaker. The subjects read a list of sentences in their respective languages which included sequences of word-final Ṽ followed immediately by word-initial D or T, as well as control utterances.

2.2. Results

The nasal olive was quite sensitive and picked up nasal microphonics in addition to the DC pressure variations that would be more directly indicative of velic opening. Some nasal microphonics may be present even when the velic valve is closed; the acoustic transparency of the velum to low frequencies is well-known [2, 3]. This happens particularly with high vowels (which have low F1) and voiced obstruents. Such microphonics are less evident in sounds with the higher F1 characteristic of low vowels. Thus the evidence of velic opening is to be seen in a DC pressure change or a disproportionate increase in nasal microphonics, vis-a-vis other comparable oral utterances.

Fig. 1. presents records of simultaneous audio (bottom) and the output of the nasal olive (top) for portions of two utterances spoken by the female French speaker. Fig. 1a gives a portion of the utterance dit 'saint' pour moi ("say 'saint' for me") /di sɑ̃ puʁ mwɑ / and Fig. 1b, a portion of the utterance dit 'saint' bel enfant ("say 'saint' beautiful baby") /di sɑ̃ bɛl ɑ̃fɑ̃/. Here the initial parts of the word-initial stops (following the nasalized vowel [ɑ̃]) are nasalized through perseveratory assimilation, i.e., they are prenasalized stops. However, in the case of the word initial voiceless stop [p] the nasalization is very brief, on the order of 30 msec, whereas in the case of the voiced stop it is much longer, about 70 msec.

Fig. 2. presents two similar records spoken by the male Hindi speaker: Fig. 2a, a portion of the utterance / ɑp jəhɑ̃ tɑko/ "you glance here" and Fig. 2b, a portion of the utterance / ɑp jəhɑ̃ dekʰo / "you see here". In Fig. 2a the signal from the nasal olive shows the word-initial /t/ to have about 30 msec of pre-nasalization. In comparison, Fig. 2b shows about 60 msec of pre-nasalization.

Comparable results were obtained for the other speakers and other places of articulation.

3. DISCUSSION

It is essential for our argument that none of the words which provided the cross-word boundary sequences of V + D would actually exhibit a N when these words are spoken in isolation. This is certainly true of the Hindi examples. In the case of French one might recall that the liaison form of words with final nasal vowels would have a supposedly "underlying" nasal consonant appear, e.g., bon ,"good", [bɔ̃] but bon ami, "good friend", [bɔn ami]. Could the nasal element found in the French examples be this underlying nasal? We think not: such liaison consonants appear when the next word starts with a vowel, not a consonant. Second, the fact that the appearance of the intrusive nasal is influenced by the voicing of the stop suggests that it is a purely transitional phonetic event created by the nasalization of the vowel invading the initial portions of the following stops.

The nasal epenthesis parallels and thus supports the historical scenario posited above for words like MH [tʃãnd]. The

"phonetic" nasal can become phonologized (to use Jakobson's term) if listeners reinterpret this as an intended part of the pronunciation and not a predictable and thus discountable feature [5].

The phonetic and phonological literature on other languages reveals that voiced stops (but not voiceless ones) may tolerate nasal onsets when in contact with a preceding nasal segment (or occasionally even when there is no preceding nasal environment)[7, 8, 9, 11, 12]. We speculate that a possible phonetic basis for this phenomenon comes from perceptual evidence that some of the essential perceptual cues for voiced stop include an amplitude and spectral discontinuity with respect to adjacent sonorants, presence of voicing during the closure, and the stop burst at the release. It seems, then, that a perceptually adequate fully voiced stop may be made by allowing the initial portion to be nasal as long as the final portion has velic closure and concomitant oral pressure impulse in order to create the requisite stop burst on release. There is no motivation to the speaker to time velic closure precisely with the onset of the stop closure. On the other hand, in the case of voiceless stops there is motivation to achieve velic closure near the onset of the stop closure: to maintain voicelessness for a substantial portion of the stop closure to avoid the frication at the nostrils, i.e., a voiceless nasal, that would occur if velic closure were delayed.

It should be mentioned that the reason for selecting Hindi and French for this study is simply the fact that both their phonologies permit $\tilde{V}$ + D sequences spanning a word boundary. It is just a coincidence that it is also the history of Hindi which exemplifies the puzzle we were trying solve. If one accepts that there are universal and timeless phonetic factors which cause variation and change in pronunciation (which may lead to sound change through phonologization), then the parallels to phonetically-based sound changes should be evident, potentially, in any spoken language which exhibits the appropriate conditions.

4. REFERENCES

[1] BEAMES, J. (1872), A comparative grammar of the Modern Aryan languages of India. Vol 1, London: Trübner & Co.

[2] CLARKE, W. M. (1978), "The relationship between subjective measures of nasality and measures of the oral and nasal sound pressure ratio", Language & Speech 21,69-75.

[3] HIRANO, M., Y. TAKEUCHI, & I. HIROTO. (1966), "Intranasal sound pressure during utterance of speech sounds", Folia Phoniatrica 18,369-381.

[4] MISRA, B. G. (1967), Historical phonology of Modern Standard Hindi: Proto-Indo-European to the present, Doctoral Dissertation, Cornell University.

[5] OHALA, J. J. (1989), "Sound change is drawn from a pool of synchronic variation", L. E. Breivik & E. H. Jahr (eds.),, Language Change: Contributions to the study of its causes. Berlin: Mouton de Gruyter. 173-198.

[6] OHALA, M. (1983), Aspects of Hindi Phonology, Delhi: Motilal Banarsidass.

[7] OHALA, M. & J. J. OHALA. (Submitted), "Nasal epenthesis in Hindi", Phonetica.

[8] PARADIS, C. (1988-1989), "On constraints and repair strategies", The Linguistic Review 6,71-97.

[9] ROBERTS, E. W. & W. R. BABCOCK. (1975), "Parametric relationships in English CVNC and CVC monosyllables", Language & Speech 18,83-95

[10] SCRIPTURE, E. W. (1902), Elements of experimental phonetics, New York: Charles Scribner's Sons.

[11] SUEN, C-Y. AND M. P. BEDDOES. (1974), "The silent interval of stop consonants", Language & Speech 17,126-134.

[12] YANAGIHARA, N. & C. HYDE. (1966), "An aerodynamic study of the articulatory mechanism in the production of bilabial stop consonants", Studia Phonologica 4,70-80.

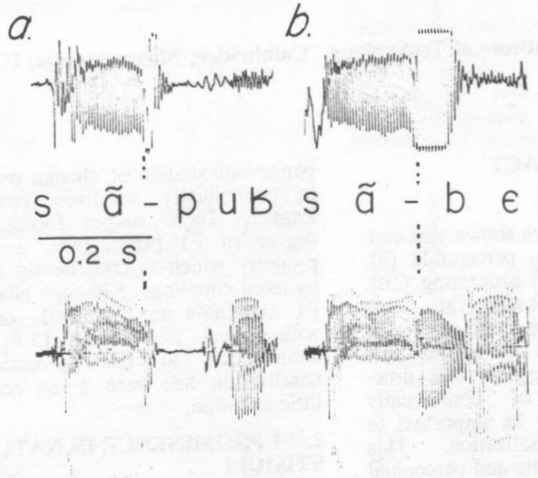


FIGURE 1. Records from the female French speaker: top: nasal olive; bottom: audio signal; a: portion of "dit 'saint' pour moi"; b: portion of "dit 'saint' bel enfant". Hyphen marks word boundary in phonetic transcription; vertical dotted line marks it in the graphic signals.

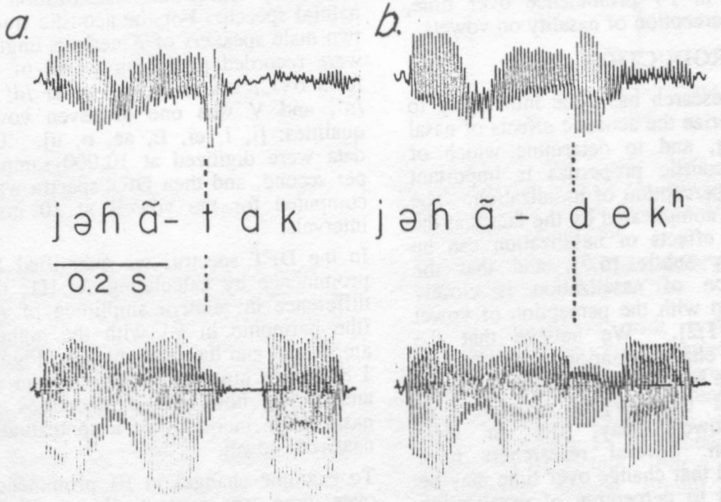


FIGURE 2. Records from the male Hindi speaker: top: nasal olive; bottom: audio signal; a: portion of /ap jəhã tako/; b: portion of /ap jəhã dekho/. Hyphen marks word boundary in the phonetic transcription; vertical dotted line marks it in the graphic signals.

TIME-VARYING PROPERTIES OF CONTEXTUALLY NASALIZED VOWELS: ACOUSTICS AND PERCEPTION

Marie K. Huffman

Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

ABSTRACT

Studies of English have shown that contextual nasalization is perceptible [8], and is used in speech processing [10]. However, when measured at single points in time, the spectral effects of contextual nasalization can appear quite subtle [7]. This suggests that time-varying properties of contextually nasalized vowels may be important to the perception of nasalization. This paper reports on acoustic and perceptual studies of changes in F1 prominence over time in contextually nasalized vowels of English. The results indicate that degree of F1 prominence, and change in F1 prominence over time, affect perception of nasality on vowels.

1. INTRODUCTION

Much research has gone into trying to characterize the acoustic effects of nasal coupling, and to determine which of these acoustic properties is important for the perception of nasalization. The effort is complicated by the fact that the spectral effects of nasalization can be relatively subtle [6,7], and that the perception of nasalization is closely bound up with the perception of vowel height [1,2]. We believe that the phonetic characterization of nasalization must take into consideration the possible role that time-varying properties of these vowels may play in their perception. Several researchers have suggested that change over time may be important to perception of nasalization on vowels [9,4], but little systematic investigation has been done in this area. This paper reports on acoustic and perceptual studies of change over time in contextually nasalized vowels in English. These studies focus on the degree of F1 prominence, a spectral property which is consistently affected by nasal coupling. Although changes in F1 amplitude are frequently observed with vowel nasalization [5,6], their contribution to the perception of nasalization has been given relatively little attention.

2. F1 PROMINENCE IN NATURAL STIMULI

Before running perceptual experiments, we conducted an acoustic study to determine how F1 prominence was affected by contextual nasalization in natural speech. For the acoustic study, two male speakers of American English were recorded producing words of the form bVC, where C was either /d/ or /n/, and V was one of seven vowel qualities: [i, I, ei, E, ae, o, u]. The data were digitized at 10,000 samples per second, and then DFT spectra were computed for the vowels at 10 msec intervals.

In the DFT spectra, we quantified F1 prominence by calculating A1-H1, the difference in relative amplitude of A1 (the harmonic in F1 with the highest amplitude) and the fundamental. Figure 1 illustrates how A1-H1 was measured, and shows how A1-H1 decreases as nasalization increases on a contextually nasalized vowel.

To examine changes in F1 prominence over time on contextually nasalized vowels, we plotted A1-H1 for each spectral frame during the vowel. To better judge which effects on F1 were attributable to nasalization as opposed to other factors, we plotted the data for the nasalized vowels along with similar data for comparable oral vowels. Figure 2 shows some examples of these combined plots, for one speaker. The second speaker showed a similar pattern.

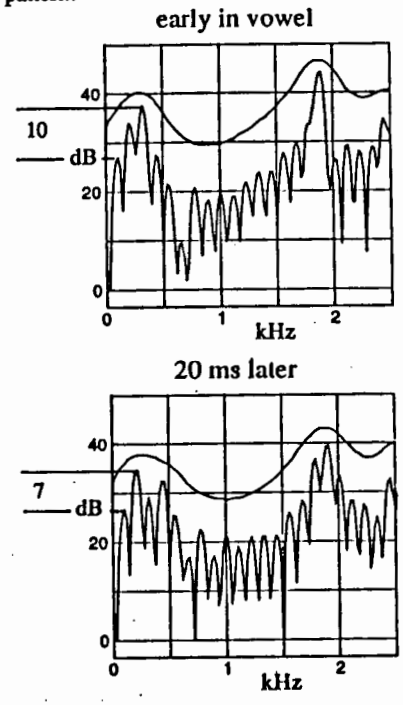


Figure 1. DFT spectra for two points in the vowel of "bin". An increase in nasalization over 20 msec results in a 3 dB decrease in A1-H1.

The data in Figure 2 show that, as we would expect, the contextually nasalized vowels had a smaller A1-H1 (less prominent F1), overall than the oral vowels. It should also be pointed out that A1-H1 values change over the course of the oral vowels as well as the nasalized vowels, suggesting that the prominence of F1 is affected by articulatory factors other than just nasal coupling. This underscores the importance of making relative, rather than absolute, measures of the spectral effects of nasalization [3]. Finally, it should be noted that A1-H1 tends to decrease over the vowel, though this is true only for the latter half of [i]. On average, A1-H1 decreases by at least 4 or 5 db over the course of a contextually nasalized vowel. Given the patterns in the natural stimuli, we used synthetic speech to investigate the role of average F1 prominence, and change in F1 prominence over time, in the perception of nasalization on vowels.

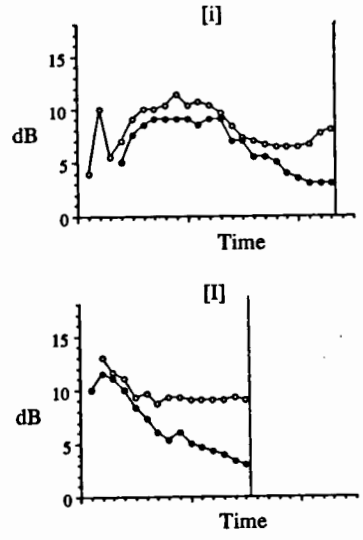


Figure 2. A1-H1 over time (10 msec intervals) for oral (open circles) and contextually nasalized vowels (filled circles), for one speaker of English.

3. PERCEPTION EXPERIMENT I: STATIC F1 PROMINENCE

The first experiment with synthetic speech focussed on how changes in average F1 prominence affect perceived nasalization. Stimuli were produced by starting with a synthesized oral vowel, and then decreasing F1 prominence by increasing F1 bandwidth. For a given vowel quality, several stimuli were produced, with different degrees of F1

prominence. For each item, F1 prominence was essentially constant over the duration of the vowel. The synthetic stimuli were chosen to have F1 prominence values which covered the range observed in natural speech items with the same vowel quality. Stimuli were constructed using 2 vowel qualities: [i], and [I]. Listeners heard a synthesized /bV/ syllable, followed by a vowel, and were instructed to choose which of two full words they felt the syllable could be an excerpt from. So, for example, on hearing [bi], the listener would circle either "bead" or "bean". 17 listeners participated in this experiment. The results of the tests are given in Figure 3, which shows the percentage of nasal responses for synthetic syllables having differing degrees of F1 prominence (as measured by A1-H1).

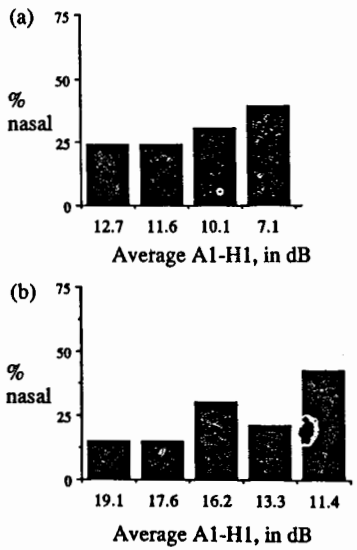


Figure 3. Percent nasal responses for synthetic /bi/ syllables (a) and /bI/ syllables (b), with static F1 prominence, at various values of A1-H1.

The majority of the time, the stimuli were heard as oral. However, the results suggest that decreased F1 prominence can contribute to production of a nasal percept. For stimuli made with the vowel [i], the percentage of times stimuli were heard as nasal increased with decreasing A1-H1 values, from about 25% nasal responses for a vowel with an A1-H1 typical of an oral vowel (12.7 dB), to about 40% nasal responses, for a vowel with an F1 that is about 5.5 dB lower. There is a similar pattern for [I], though the data appear to be a bit noisier.

4. PERCEPTION EXPERIMENT II: VARYING F1 PROMINENCE

We tested the effect of change in F1 prominence over time on perception of nasalization by comparing listener judgements of the nasality of stimulus pairs which were matched for vowel quality and overall average A1-H1, but which differed in having either an unchanging F1 prominence, as in the previous experiment, or a time-varying, decreasing, F1 prominence.

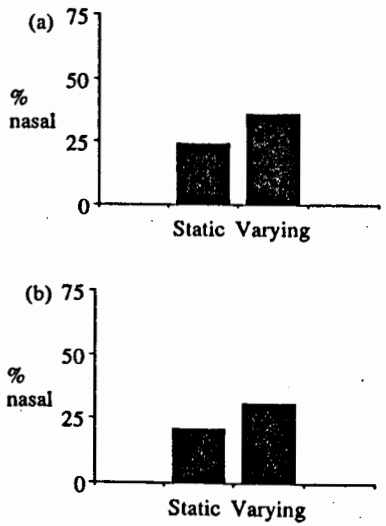


Figure 4. Percent nasal responses for stimuli with time-varying F1 prominence versus stimuli with static F1 prominence, matched for average A1-H1, for /i/ (a) and /I/ (b).

The time-varying vowel stimuli were synthesized with F1 prominence decreasing throughout the vowel. (The drop from vowel beginning to end was about 4 dB). Listeners heard synthetic /bV/ syllables containing these vowels within the same paradigm used in the previous experiment.

Figure 4 presents comparisons of the percentage of nasal responses for the stimuli with time-varying A1-H1 on the vowel, and their counterparts with static A1-H1 on the vowel. The time-varying stimuli show a higher percentage of nasal responses than the static stimuli, indicating that a decrease in F1 prominence over the course of the vowel results in more nasal responses than a simple, static reduction of F1 prominence.

5. SUMMARY

To conclude, we have seen that change in F1 prominence over time influences the perception of nasalization. By comparing perception of stimuli with static and time-varying F1 prominence, we determined that change over time is important, and that it is not just average F1 prominence which determines perceived nasalization. This is evidence for the importance of dynamic information in the perception of vowels. It also may have implications for predicting the likelihood of sound changes in which a contextually nasalized vowel becomes a contrastively nasalized vowel. Since physiological adjustments other than nasal coupling can affect F1 prominence, it is possible that changes in F1 prominence over time which come with diphthongization or laryngeal adjustments for voicing could contribute to a percept of nasalization. In combination with contextual nasalization, such effects could result in a stronger percept of nasalization, such that the language learner will be more inclined to posit a nasal vowel in that position, providing that other grammatical considerations do not prevent such an analysis. These questions await future research.

ACKNOWLEDGEMENTS

This research was supported by NIH grant #F32-NS08509. I am grateful to researchers at Haskins Laboratories and the MIT Speech Group for helpful discussions of this work.

REFERENCES

[1] BEDDOR, P., KRAKOW, R. & GOLDSTEIN, L. (1986). "Perceptual constraints and phonological change: a study of nasal vowel height", *Phonology Yearbook, 3*, 197-218.

[2] BEDDOR, P. & HAWKINS, S. (1990). "The influence of spectral prominence on perceived vowel quality", *J. Acoust. Soc. Am., 87*, 2684-2704.

[3] BELL-BERTI, F. (1980). "Velopharyngeal function: A spatial-temporal model", in *Speech and Language: Advances in Basic Research and Practice, Vol. 4*, 291-316.

[4] BLADON, A. (1986). "Phonetics for Hearers", in G. McGregor, (ed.) *Language for Hearers*, Pergamon, Oxford.

[5] DELATTRE, P. (1968). "Divergences entre nasalites vocalique et consonantique en francais", *Word, 24*, 64-72.

[6] HOUSE, A. & STEVENS, K. (1956). "Analog studies of the nasalization of vowels", *J. Speech Hear. Dis., 21*, 218-232.

[7] HUFFMAN, M. (1990). "An acoustic study of the timing of contextual nasalization in English", *J. Acoust. Soc. Am. Vol 87, Suppl. 1*, S67.

[8] HUFFMAN, M. (1990). "The role of F1 amplitude in producing nasal percepts", *J. Acoust. Soc. Am. Vol. 88, Suppl. 1*, S54.

[9] REENEN, P. T. VAN (1982). *Phonetic feature definitions.* Foris Publications, Holland.

[10] WARREN, P. & MARSLEN-WILSON, W. D. (1987). "Continuous uptake of acoustic cues in spoken word recognition", *Perception & Psychophysics, 41*, 262-275.

PHONETIC STRUCTURE OF WORD AND PECULIARITIES OF ITS DEVELOPMENT
(based on Germanic and Slavonic languages)

V. Taranets

Odessa State University, Odessa, USSR

ABSTRACT

Ancient rise-fall alteration of the articulation tension with a displaced apex (positive asymmetry) is reconstructed in CV syllable and in words with initial stress. In the cour-se of the development of language intensification of tension is observed at the beginning of a word and relaxation in the end of it, i.e. a redistribution of energy takes place. Presumably, the overall utterance energy remains, in principle, constant.

1. INTRODUCTION

In the process of the development of language its sound aspect undergoes the greatest alteration. Changes occur in sounds, syllables and whole words. The object of study is root-stressed words. The study of the peculiarities of such words implies in the first place a synchronic and diachronic investigation of its CV correlative.

2. PROCEDURE

An electro-acoustic study of the CV word (syllable) in the German and Ukrainian languages was made (the experiment was carried out in the Berlin University under the supervision of Prof. G.Lindner) as well as that of CV articulation tension by using pletysmographic method (Odessa University, Prof.V.Taranenko). The pletysmographic method made it possible to determine the platisma and suprahyoidei muscular tension while uttering a stressed CV-syllable.

In Germanic languages Runic (Old Futhark), Gothic, Old High German, modern German and English texts, in Slavonic languages Old Slavic, Old Russian, modern Russian and Ukrainian texts have been studied. The dynamics of the alteration of the initial stressed syllable as well as of the final unstressed one made it possible to determine the peculiarities of the alteration of the word as a whole in the course of language development.

3. RESULTS AND DISCUSSION

3.1. Phonetic characteristics of the CV-syllable

An analysis of the CV-syllable used in isolation and in words of the CṼCV(C) pattern revealed the following.

In Ukrainian the length of a consonant and of a vowel had an average value of, respectively, 0.380 and 0.620, in German 0.355 and 0.645 (the length of the CV-syllable is taken for 1.00). In German a phonologically long vowel is present. In a general way, it may be assumed that the ratio of the length of the consonant and that of the vowel in Ukrainian and German in CV is 1:2.

The articulation apex E_{max} in the CV often occupies the vowel area in the following way: in the Ukrainian syllable 80.4%, in German - 93.5%. The rising part forms the working phase, the falling one - the articulation relaxation which, generally, reminds one of a single muscle-tension [1] . Typical of both languages is the rise-fall alteration in the articulation tension, with some shift of the apex towards the beginning of the utterance (positive asymmetry) (Fig.1).

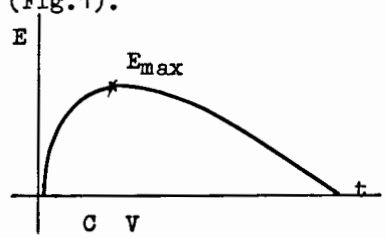


Fig.1. Articulation tension of CV-syllable

3.2. Phonetic characteristics of a CṼCV(C) word

Such words in Ukrainian and German are characterised by two-apex alteration of the articulation tension. In Ukrainian E_{max} falls on the first syllable in 97.7% of cases, in German in 87.5%. However the apex occupies the initial consonant in the Ukrainian syllable in 72.9% of cases, in German it is 14.8%. It is typical of Ukrainian speech to have a greater tension for a con-sonant, as related to a vowel within the CV, whereas in German it is vice versa - the vowel is more tense than the consonant. Presumably, in Ukrainian spech realized is a "strong-consonant" phonetic tape, while in German - a "strong-vowel" type.

The tonic apex falls on the first syllable in CṼCV(C) units in all cases. The intensity in the Ukrainian word falls on the first syllable in 85.0% of cases, in German in 88.2%.

In general, the alteration of the phonetic characteristics has a rise-fall pattern with the apex displaced towards the beginning of the utterance (similar to the CV structure) (Fig. 2).

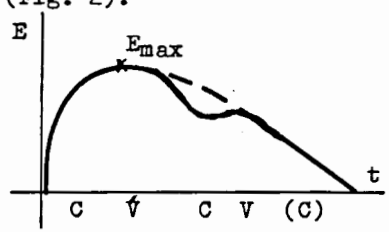


Fig.2. Phonitic structure of CṼCV(C)

3.3. Articulation tension of consonants (E_o)

Analysis of CV and CṼ- syllables made it possible to find the tension of the consonant relative to the vowel, whose value is taken to be 1.00.

The analysis also revealed the difference of the E_o consonants in terms of their formation (Table 1).

Table 1. Articulation tension of consonants (E_o) (Ukrainian and German languages)

Types of consonants (examples)	E_o Ukr.	E_o Germ.
R_w - semi-consonants (w,j)	1,20	-
R_r - liquids (r,l)	1.04	0.84
R_n - nasal (m,n)	0.84	0.79
D - voiced occlusives (b,d,g)	0.74	0.57
T - voiceless occlusives (p,t,k)	0.68	0.69
Z - voiced fricatives (v,z)	0.53	0.28
S - voiceless fricatives (s,f,h)	0.35	0.20

In German, the consonants T (p,t,k) are opposed to D (b,d,g) as fortes/lenis [3], in Ukrainian as voiceless/voiced.

3.4. Development of initial consonants in words

Analysis of ancient and modern memorials in Germanic languages has shown that at the beginning of a word the following generalized combinations occur: $\overleftarrow{ST}$-, $\overleftarrow{TR}$-, $\overleftarrow{SR}$-, $\overleftarrow{DR}$-, $\overleftarrow{STR}$- (where R is a sonorant) with a rising tension. For example: skin, tree, snake, dream, stream; foreign words having a non-rising tension being exceptions. For example: sphinx, (Germ.) Psalm, Ndola. The same consonant combinations are found in old Slavic texts, for example: skot, trije, slowo, zmii, bratr, strana. After the fall of reduced vowels combinations with non-rising tension were formed, sach as $\overrightarrow{SS}$-, $\overrightarrow{TZ}$-, $\overrightarrow{RR}$-, $\overrightarrow{RST}$- and others. For example: (Russ.) ssora, rwat', mrak, vhod, mstit' [2].

Generally, the word's beginning in Germanic and Slavonic languages presents a gradual articulation intensification as contrasted with a reduction in the word's end which resulted in the relative growth of closed syllables and consonant clusters. In ancient times, open syllables with CV among them, prevailed in these languages.

4. CONCLUSIONS

Extrapolation of tendencies of the word's beginning and end development in the prehistoric period makes it possible to arrive at the following conclusions:
a) in Germanic and Slavonic languages there has existed a tendency of the open syllable, the closed syllable being a result of language development;
b) in ancient time, the combinations of initial consonants in words were formed on the principle of rising tension, combinations with non-rising tension being secondary;
c) in the end of words, there occur different types of consonant combinations which are results of the articulation reduction of this part of the word;
d) in general, in the course of language development an intensification of tension has been taking place at the beginning of words and a reduction at their end, which implies an interaction of both tendencies. It is supposed that in the course of language development the overall utterance energy remains, in principle, constant and is redistributed within the word.

5. REFERENCES

[1] HILL, A.V. (1970), "First and last experiments in muscle mechanics", Cambridge: University press.
[2] TARANETS, V.G.(1981), "Energy theory of speech", Kiew-Odessa: Vyšča škola (in Russ.).
[3] TRUBETZKOY, N.S.(1958), "Grundzüge der Phonologie", 2.Auflage, Göttingen.

VOWEL HARMONY AS A COARTICULATORY PHENOMENON IN NANAY

Galina Radchenko

Lund University, Sweden

ABSTRACT

The present paper reports on interaction of the context (vowel harmony), target gesture and suprasegmental factor (stress). It is argued here that the notion of stable gestures and anticipatory gestures can account much of the variability of the context, given certain assumptions about the effect of stress.

An experiment was designed to evaluate the role of coarticulation and tempo on the dynamics of vowel articulation. It is shown here that motor plan for a particular segment remains the same regardless of the varying phonetic context and timing conditions.

1. INTRODUCTION

Coarticulation is defined as the influence of segmental context on the articulatory/acoustic realization of a target segment. It is assumed that because of perceptual or articulatory constraints on target and surrounding segments, there are limits on the temporal extent of coarticulation.

There are two particular speech production frameworks: the "look-ahead" models [6] and the "coproduction" models [5], [1], [4]. According to the coproduction model the underlying motor control structure for a particular segment remains essentially the same regardless of the phonetic identity of surrounding phones. In contrast to the look-ahead models changes in the observed patterns of movement in different contexts stem from local interactions between context and target gestures rather than from any change in the motor plan for the target segment. According to the look-ahead model the motor plan for the target segment and consequently the time at which coarticulation begins are revised and adapted to the context since every different context poses a different set of conditions.

The present paper supports the point of view that was put forward by Bladon & Al-Bamerni [2], who proposed that observed coarticulatory patterns might be a combination of anticipatory feature spread plus stable gestures.

In the present paper we examine coarticulation presented by vowel harmony phenomenon as an interaction between target gesture with the context. According to preliminary auditory analysis the adjustment of Nanay consonants to a certain harmonic class of vowels manifests itself in the use of dental allophones of affricates [ts], [dz] with the vowels of the first harmonic series /a/, /o/, pronounced with low jaw position and palatal allophones [dʒ], [ʧ] with vowels of the second harmonic series /i/, /ə/ pronounced with high jaw position : otsoqa 'small fish', əʧəkə 'uncle'.

The velar allophones [k], [g], [x] are used with the vowels of the second harmonic series [ə], [u] and uvular allophones [q], [ɢ], [χ] with the vowels of the first harmonic series [a], [o] : χaLa 'tribe', xələ 'stutterer '.

Most investigators of this problem mention that the influence of context upon vowels is regularly manifested as a displacement of vowel-formant frequencies away from their target frequencies [9]. An experiment was therefore designed to study vowels pronounced under varying timing conditions and in systematically varied consonantal environments.

2. PROCEDURE

To investigate the displacement of consonant articulations from their bull's-eye patterns X-ray pictures of consonants in varied vowel environments were made.To study the displacement of vowel formant frequencies away from their target frequencies, measurements were made of the first, second and third formant frequencies of vowel /a/ in three consonant environments: /p-p/, /t-t/, /x-x/ under varying timing conditions. F1, F2 and F3 of the investigated vowel were plotted against vowel segment duration, then the first and the second formant target frequencies were determined by plotting F1 and F2 against vowel duration for the three contexts of the vowel simultaneously [7].

The influence of stress and tempo on vowel properties was studied by the manipulating of the word order and rhythm of the carrier sentence frames:'CVC ------; ----- CVC ------'. Each sentence pattern was pronounced four times.

The speech material was read by a native male speaker of Nanay. and was analyzed by MacSpeech Lab 2.

3. RESULTS

X-ray pictures show the adjustment of consonants to the near vowels.

Dental affricates are characterised as flat and have a decrease of frequency components of release (the concentration of noise is at 2500 - 3000 Hz). It promotes reducing of the timber of the near vowel.

Palatal affricates may be called sharp and are characterised by high intensity of the upper frequency components (about 4000 - 6000 Hz). Articulatory palatal sounds are produced with wide pharyngeal cavity. X-ray pictures show that the constriction location of palatal affricate [ʧ] is at the zone of alveolar and hard palate,whereas constriction location of [dz] is limited by dental zone (fig. 1-2).

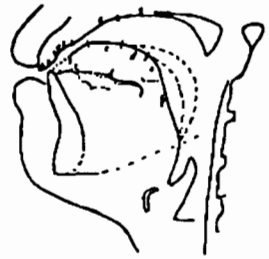


Figure 1. X-ray picture of dental affricate [dz] pronounced in the syllable [adza] by Nanay male speaker N 3.

Figure 2. X-ray picture of the affricate [ʧ] in the syllable [əʧə] pronounced by Nanay male speaker N 3.

X-ray pictures of allophones [q] and [k] show that their constriction location depends on the articulation of the near vowel. Vowels pronounced with low jaw position promote the shift of constriction location from velum to uvula (fig.3). Vowels pronounced with high jaw position /ə/, /u/ initiate a constriction location of the near consonant /k/ at velum (fig.4).

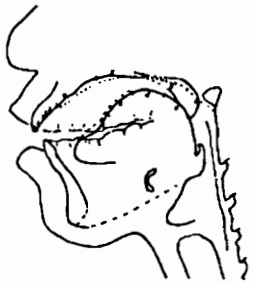


Figure 3. X-ray picture of [q] in the word d'aqa 'thing' pronounced by Nanay male speaker N 2.

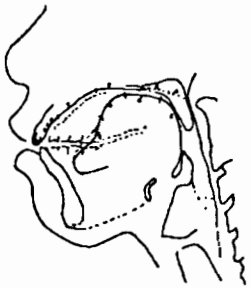


Figure 4. X-ray picture of [k] in the word d'ukən 'hardly' pronounced by Nanay male speaker N 2.

In all these cases of assimilation there are no changes at higher levels: Supra, Vocal Tract, Oral remain the same, changes undergo only at a low level: the place of location .This slight transition of the place of articulation seems not to change motor plan for target gesture, non-significant changes in gesture movements are the results of local interactions between context and target gesture.

The adjustment of vowels to the varied phonetic context is displayed in figures 5 - 6.

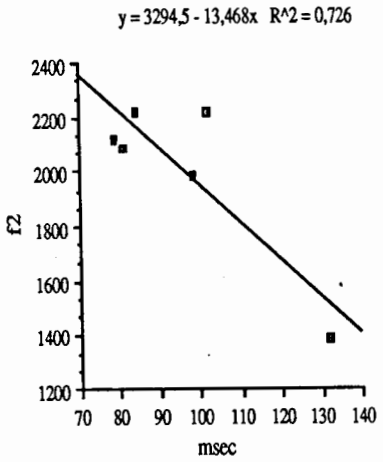


Figure 5. F2 data plotted against vowel segment duration in the context [tat] pronounced by Nanay male speaker.

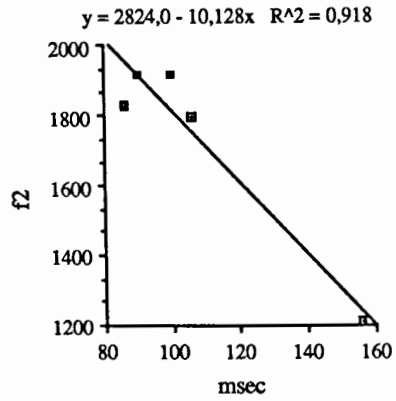


Figure 6. F2 data plotted against vowel segment duration in the context [pap] pronounced by Nanay male speaker.

Figures 5 - 6 show the values that vowel /a/ assumes for the three contexts under variable time conditions. The vowel seems to preserve its target frequency values in a variable context.The grouping of the points approximates a straight line which shows a correlation of vowel duration and formant frequency values of the vowel. The shorter duration of vowel segment is the higher values of F2 the segment assumes. This tendency is preserved irrespective of consonantal context. The values show that a target has been found to be independent of consonantal context and duration and can thus be considered to be invariant.

4. DISCUSSION

The discussed cases of coarticulation in Tungus may be viewed as realization of anticipatory speech model. Coarticulation is the result of local interaction of overlaping gestures: jaw lowering or jaw rising during the production of a vowel effects the articulation of the consonant. It is proved by a number of experimental research works that onset of jaw lowering for a vowel will start during a preceding consonant [8]. Vowel harmony may be viewed as constraints imposed by the context on the realization of a target gesture.

The changes in the segmental context are predictable and usually set up as phonotactic rules.They reflect the range of possible variability of the target gestures in a different context. The motor plan for target gestures seems to be done beforehand and changes in gesture movements in different contexts stem from local interactions between context and target gestures.

5. CONCLUSION

The present paper reports on a pilot study of vowel harmony in Tungus languages. Preliminary results have shown that restrictions imposed by a context do not change essentially a motor plan for a target gesture. A number of questions are raised that future research is supposed to answer. These questions are: what restrictions are there imposed by a context (vowel harmony) on the realization of a target gesture?; what relations are there between vowel harmony, target gesture and stress?; to what extent can context change a target gesture?

REFERENCES

[1] Bell-Berti, F., Harris, K.S. (1981), "A temporal model of speech production", *Phonetica* , 38, 9-20.

[2] Bladon, R.A.W., Al-Bamerni, A. (1982), "One-stage and two-stage temporal patterns of velar coarticulation", *Journal of the Acoustical Society of America* , 72, S104(A).

[3] Boyce, S.E., Krakow, R.A., Bell-Berti, F., Gelfer, C.E. (1990), "Converging sources of evidence for dissecting articulatory movements into core gestures", *Journal of Phonetics*, 18, 173-188.

[4] Browman, C.P.,Goldstein, L., (1989), "Articulatory gestures as phonological units", *Phonology* 6, 201-251.

[5] Fowler, C. A. (1980), "Coarticulation and theories of extrinsic timing", *Journal of Phonetics*, 8, 113-133.

[6] Keating, P.A. (1988), "Underspecification in phonetics," *Phonology*, 5, 3-29.

[7] Lindblom, B. (1963), "Spectrographic Study of Vowel Reduction", *The Journal of the Acoustic Society of America*,35, 1773-1781.

[8] Löfqvist, A. (1986), "Stability and change", *Journal of Phonetics*, 14, 139-144.

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

TENDANCES UNIVERSELLES ET STABILITÉ DES SYSTÈMES VOCALIQUES

N. Vallée, L.J. Boë et J.L. Schwartz

Institut de la Communication Parlée, URA CNRS n° 368
INPG/ENSERG – Université STENDHAL,
Domaine Universitaire, BP 25X, 38040 Grenoble cedex, France

ABSTRACT
This paper deals with a study of the structure of vowel systems in two respects: ① Our observations of certain aspects of vowel systems using the entire database of 317 language descriptions [4] lead us to confirm or refine certains tendencies and regularities in vowel systems; ② We have used a predictive model [7] of the 3 dimensional $F_1/F_2/F_3$ space to test our hypothesis that if a system is more frequent in the inventory, it is an acoustically "stable" system.
This research extends Lindblom's work about the predictive models of the organisation of the vowel space and the explanations of language universals.

INTRODUCTION

On dispose aujourd'hui d'inventaires phonologiques des langues du monde relativement importants. Ces matériaux offrent la possibilité de tester de nouvelles typologies et propositions de tendances de développement des systèmes vocaliques, nourissant ainsi la discussion sur les modèles de prédiction dans un espace de représentation. Ces modèles tentent d'expliquer l'organisation des unités à l'intérieur des systèmes, et par la même, la présence d'universaux dans les langues. C'est pour les systèmes vocaliques que ces modèles ont été le plus développés.
Dans un premier temps, nous présentons les points essentiels qui ressortent de notre typologie des systèmes vocaliques [8], établie à partir de la base de données UPSID (UCLA Phonological Segment Inventory Database) qui réunit la description phonologique de 317 langues du monde [4]. Nous avons pris la totalité de la base en excluant toute classification a priori pour repostuler ou confirmer certaines tendances universelles dans la fréquence d'occurrences des systèmes et des voyelles ainsi que dans l'organisation des 220 types de systèmes vocaliques relevés dans l'ensemble de la base.
La deuxième étape a consisté à tester, à partir de cette typologie, la "stabilité acoustique" – dans un sens qui sera précisé – des systèmes vocaliques les plus fréquents, avec un modèle de prédiction dans l'espace 3-D $F_1/F_2/F_3$ [7]. Ce modèle intègre le principe de dispersion maximale de Liljencrants & Lindblom (1972) [2] (L&L) et le complète par un critère de focalisation introduit par le biais de plans attracteurs (F_1=F_2, F_2=F_3, F_3=F_4), ainsi qu'une pondération du second formant effectif F'_2.

1. STRUCTURE DES SYSTÈMES VOCALIQUES DE LANGUES NATURELLES

1.1. Taille et occupation de l'espace

Les systèmes décrits dans UPSID possèdent de 3 à 24 voyelles. Le classement de la base nous a permis d'en extraire 220 types. Nous avons relevé dans l'inventaire une très nette dominance des systèmes à 5 voyelles (23% des langues). Les systèmes qui possèdent de 3 à 10 voyelles représentent 80% de l'échantillon et sont donc très largement majoritaires. Les systèmes les plus fréquents possèdent une large dispersion dans l'espace articulatoire traditionnel décrit par les axes d'aperture et de lieu d'articulation. Ils sont composés de voyelles que l'on retrouve quelle que soit la taille du système.
Nous avons pu mettre en évidence une différence très nette de tendance entre les types : *9 représente le maximum de timbres vocaliques distincts que peut présenter un système.* En effet, ceux qui ont un nombre élevé de voyelles ne développent pas de nouveaux timbres, mais ajoutent une complexité articulatoire à ces segments de base, en leur additionnant d'autres dimensions telles que la longueur, la nasalité, la pharyngalité — ainsi : /ã/, /ũː/, /aˤ/.
En limitant leur nombre de timbres vocaliques, on peut dire que les langues naturelles fonctionnent sur un principe de contraste pour un nombre d'unités déterminées. Cependant, il s'agit d'un contraste "suffisant" (et non pas maximal, cf [4] p.16) : l'apparition de nouvelles dimensions s'effectue surtout à partir de 10 voyelles, quand l'espace des timbres est donc trop encombré.

1.2. Organisation de l'espace

Nous allons décrire par 10 règles comment les timbres s'organisent en série (sur les traits [antérieur], [central], [postérieur] / [arrondi], [non arrondi]).

1.2.1. *Organisation horizontale / verticale*

① les degrés d'aperture sont plus nombreux que les distinctions antéro-postérieures — la tendance générale dans les systèmes vocaliques est de 3 à 5 degrés de distinction par l'aperture et de 3 distinctions sur l'axe antéro-postérieur, et ceci quelle que soit la taille des systèmes.
On peut même affirmer en général que :
①' le nombre de degrés d'aperture est supérieur ou égal au nombre de séries — avec cependant le contre-exemple notable du système /i a u/.

1.2.2. *Comparaison des séries*

• Dans un système donné :
② le nombre de voyelles périphériques est supérieur au nombre de voyelles intérieures — 100% des langues possèdent des voyelles périphériques et 44% des voyelles intérieures.
③ le nombre de voyelles antérieures non arrondies /i, ɪ, e, ɛ̃, æ, a/ est plus grand ou égal au nombre de voyelles postérieures arrondies /ɒ, ɒ̟, ɔ, o̟, o, ɷ, u/ (91% des cas).
• Dans les langues :
④ les voyelles périphériques : /i, ɪ, e, ɛ̃, æ, a, ɔ, o̟, o, ɷ, u/ sont plus fréquentes que les voyelles intérieures (avec /i, a, u/ proches des 100% d'occurrences).
⑤ les centrales non arrondies /ɨ, ɨ̟, ə, ə̟, ɜ, ɐ/ sont plus fréquentes que les antérieures arrondies /y, ʏ, ø, ø̟, œ/.
⑥ les postérieures non arrondies /ɯ, ɯ̟, ʏ, ʏ̟, ʌ/ sont plus fréquentes que les centrales arrondies /ʉ, ʉ̟, ɵ, ɵ̟/.
⑦ les centrales non arrondies ont une occurrence plus forte que les centrales arrondies.

1.2.3. *Organisation dans les séries*

⑧ les voyelles antérieures arrondies apparaissent :
– par série — les séries d'antérieures arrondies de 2 phonèmes apparaissent dans les systèmes ayant au moins 7 voyelles, celles de 3 ou 4 phonèmes apparaissent dans les systèmes à 16 et 19 voyelles,
– toujours avec les voyelles antérieures non arrondies de même aperture,
– presque toujours avec les voyelles postérieures arrondies de même aperture,
– selon l'ordre de fréquence décroissante : /y/ > /ø/ > /œ/ > /ʏ/ > /ø̟/.
⑨ les voyelles postérieures non arrondies figurent :
– généralement seules dans leur catégorie — les séries attestées contiennent au plus 3 voyelles,
– selon l'ordre de fréquence suivant : /ɯ/ > /ɯ̟/ > /ʏ/ > /ʏ̟/ > /ʌ/,
– souvent sans la voyelle postérieure périphérique de même aperture (ce qui justifie l'essentiel des cas de disparition de /u/ remplacé par /ɯ/).
⑩ les voyelles centrales s'organisent :
– plus régulièrement sur l'axe haut/bas que sur l'opposition arrondi/non arrondi,
– jamais en série sans la présence d'une voyelle centrale haute
– et dans ce cas avec une voyelle périphérique haute.

2. ÉTUDE DE LA STABILITÉ ACOUSTIQUE DES SYSTÈMES VOCALIQUES

2.1. Utilisation d'un modèle de prédiction

Schwartz & al. [7] se proposent d'améliorer 2 résultats obtenus par L&L [2] contraires aux données de la base UPSID :
– la prolifération des voyelles hautes entre [i] et [u], peu compatible avec les règles ① et ①';
– l'impossibilité de prédire une série antérieure arrondie sans une série postérieure non arrondie ou centrale "équilibrante" au sens de la théorie de la dispersion.
Dans un espace 3-D $F_1/F_2/F_3$, le modèle de Schwartz & al. reprend de L&L la minimisation des distances intervoyelles comme fonction de l'énergie des systèmes :

$$E_0 = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{1}{d_{ij}^2} \rightarrow \text{minimisée}$$

où d_{ij} est la distance formantique pondérée entre 2 voyelles i et j :

$$d_{ij} = \left[\left(F_{1_i} - F_{1_j}\right)^2 + \lambda \left(F'_{2_i} - F'_{2_j}\right)^2 \right]^{\frac{1}{2}}$$

λ est le coefficient pondérateur des formants élevés ($\lambda \leq 1$). Un poids plus important de F_1 dans le calcul des

distances améliore la prédiction des voyelles périphériques en diminuant le nombre de voyelles hautes.
Le calcul de E_0 repose donc sur un critère systémique, les voyelles devant être les plus éloignées au sens d'une distance pondérant F_1 par rapport à F'_2. Mais ce critère ne permet pas de prédire la structure de voyelles hautes /i, y, u/ (attestée dès 7 voyelles).
Schwartz & al. ont donc introduit un deuxième critère — celui-ci extra-systémique — qui favorise pour chaque voyelle les rapprochements F_1F_2, F_2F_3, F_3F_4 et diminue l'énergie du système. C'est le critère de focalisation, justifié d'un point de vue perceptif : les voyelles focales sont préférées perceptivement [6].

$$E = E_0 + \alpha(E_{12} + E_{23} + E_{34})$$

$$E_{12} = -\Sigma_i \frac{1}{(F_{2_i} - F_{1_i})^2}$$

$$E_{23} = -\Sigma_i \frac{1}{(F_{3_i} - F_{2_i})^2}$$

$$E_{34} = -\Sigma_i \frac{1}{(F_{4_i} - F_{3_i})^2}$$

α est appelé coefficient des plans attracteurs. D'un point de vue analogique les voyelles s'éloignent les unes des autres tout en étant attirées par des plans attracteurs : $F_1=F_2$, $F_2=F_3$, $F_3=F_4$.

Pour utiliser le modèle il nous a fallu disposer :
- d'un espace maximal 3-D ;
- de valeurs formantiques F_i (i=1 à 3), F_4 étant fixé à 3350 Hz, pour les 37 voyelles de la base UPSID (cf. Fig. 1).

Ces points ont été obtenus grâce à une série de travaux précédents fondés sur le modèle articulatoire de Maeda [5] et grâce à la confrontation des données formantiques de 15 études, issues de travaux sur des modèles ou sur des langues naturelles.

2.2. Hypothèses sur la stabilité

Nous ferons l'hypothèse qu'un système fréquent dans les langues du monde est un système à structure acoustiquement stable (énergie localement minimale) : si on le soumet au modèle de prédiction il doit donc conserver sa structure. La stabilité d'un système est ainsi appréciée à partir de son énergie qui dépend de 2 coefficients :
① λ coefficient pondérateur des formants tel que $0 < \lambda \leq 1$ (si λ=1 on a une distance euclidienne classique) ;
② α coefficient des plans attracteurs tel que $0 \leq \alpha \leq 1$.
Notre étude a eu pour but de trouver un couple de valeurs (λ,α) qui mette en relation la validité des critères du modèle et la stabilité attestée des systèmes les plus fréquents.

2.3. Systèmes sélectionnés pour l'étude.

A partir de notre typologie [8], un ensemble de 25 systèmes de 3 à 9 voyelles sélectionnés comme les plus fréquents ont été testés pour 3 valeurs de λ : [0.25, 0.5, 1] et 5 valeurs de α : [0, 0.25, 0.5, 0.75, 1], soit un total de 375 tests. Nous avons procédé en sorte qu'il y ait au moins deux systèmes concurrents (ou plus) par nombre de voyelles. La seule comparaison significative possible des énergies est entre les systèmes de même taille et pour des couples (λ,α) identiques.

2.4. Résultats

Notre modèle prédit la stabilité de 64% des systèmes les plus fréquents (qui représentent, après regroupement des systèmes relativement proches, 60% de la base UPSID).
Les cas non prédits stables par le modèle sont ceux qui possèdent au moins 4 voyelles sur un des bords de la périphérie et qui ne sont pas équi-réparties ou ceux qui possèdent une ou plusieurs voyelles intérieures sans posséder de voyelle centrale haute (soit 36% des systèmes testés et approximativement 20% de la base UPSID). Une voyelle centrale telle que /ɨ/ est plus facilement prédite qu'une voyelle comme /ə/ : pour de petites valeurs de λ, /ə/ se déplace vers les périphériques antérieures et pour des valeurs de λ grandes, vers les voyelles centrales hautes.
La poursuite des tests avec des valeurs de $\lambda < 0.25$ et des valeurs de coefficients différents pour chaque plan attracteur (α_{12}, α_{23}, α_{34}) pourrait améliorer les résultats recherchés.
Nous n'avons pas mis en évidence un seul couple (λ,α) comme nous l'avions souhaité pour l'ajustement du modèle, mais 2 couples de coefficients pour lesquels les résultats sont les plus satisfaisants : (0.25, 0.5) et (0.5, 0.25). Les tests ont mis en évidence que pour $\lambda > 0.5$, la prédiction des systèmes "tout périphérique" est impossible jusqu'à 9 voyelles et que le facteur α est nécessaire pour la prédiction de /y/ : pour α nul (modèle L&L), /y/ se déplace vers la position formantique de /ɨ/ ou /ɯ/.

CONCLUSION

Le modèle de prédiction a permis de simuler la plupart des systèmes sélectionnés, améliorant les résultats de L&L [2] et Lindblom [3] à propos des voyelles centrales hautes et de la prédiction de /y/, stabilisé par notre critère de focalisation F_2F_3 [1] (cf. Fig. 2).
A part le problème des voyelles trop nombreuses et celui de /ə/ (qui est aussi, on le sait bien, un problème de description phonologique), le modèle prédit des systèmes le plus souvent en accord avec les grandes tendances de développement des systèmes vocaliques. Même s'il ressort de notre étude que tous nos systèmes testés ne sont pas stables, la notion d'énergie basse pour traduire un des critères d'organisation des voyelles dans les systèmes reste un bon critère.
Avec ces résultats, notre recherche améliore les prédictions et étend les travaux de Lindblom & Liljencrants [2], *«to derive linguistic form as a consequence of various substance-based principles pertaining to the use of spoken language and its biological, sociological, and communicative aspects.»*

RÉFÉRENCES

[1] BOË L.J. & ABRY C. (1986)
Nomogrammes et systèmes vocaliques.
XV^e journées d'études sur la parole, Aix-en-Provence, 303-306.

[2] LILJENCRANTS J. & LINDBLOM B. (1972)
Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast.
Language 48, 839-862.

[3] LINDBLOM B. (1986)
Phonetic Universals in Vowel Systems.
Experimental Phonology, Ohala J.J. (Ed), Academic Press, Orlando, Floridia, 13-44.

[4] MADDIESON I. (1986)
The Size and Stucture of Phonological Inventories: Analysis of UPSID.
Experimental Phonology, Ohala J.J. (Ed), Academic Press, New York, 105-123.

[5] MAEDA S. (1979)
An Articulatory Model of the Tongue Based on a Statistical Analysis.
J. Acoust. Soc. Am. 65, Vol. S1, S22.

[6] SCHWARTZ J.L. (1987)
A propos des notions de forme et de stabilité dans la perception des voyelles.
Bulletin du Laboratoire de la Communication Parlée, Vol. 1A, 159-190.

[7] SCHWARTZ J.L. BOË L.J. PERRIER P. GUÉRIN B. & ESCUDIER P. (1989)
Perceptual Contrast and Stability in Vowel Systems: A 3-D Simulation Study.
Eurospeech 89, Paris, Vol. 1/2, 63-66.

[8] VALLÉE N. BOË L.J. & SCHWARTZ J.L. (1990)
Systèmes vocaliques : typologies et tendances universelles.
XVIII^e journées d'études sur la parole, Montréal, 32-36.

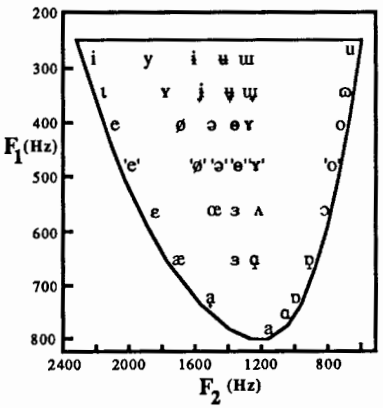


Figure 1. Répartition dans l'espace maximal 2-D F_1F_2 des 37 voyelles d'UPSID.

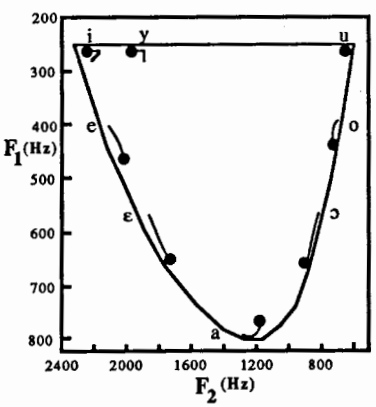


Figure 2. Exemple de simulation permettant de tester la stabilité d'un système possédant /y/.
Ici λ=0.5 et α=0.25.
Ce type de système pourait fournir sur application d'un "principe de série" (cf. règle ⑧) la base d'un système à voyelles antérieures non arrondies, antérieures arrondies et postérieures arrondies, attesté dans UPSID.

MARQUES SEGMENTALES SOCIOSITUATIONNELLES EN FRANÇAIS CONTEMPORAIN

G. Jetchev

Université de Sofia, Bulgarie

ABSTRACT

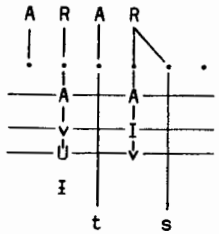
This paper discusses a possible classification of phonological variability phenomena in Modern French. Further the phonetic realizations of schwa in two variants of dialogue are treated in relationship with prosodic and syntactic boundaries.

1. TROIS TYPES DE VARIATION PHONOLOGIQUE

Si, comme l'affirme P.Encrevé<2>, la variation potentielle est à localiser dans les représentations phonologiques et si "les autosegments flottants sur la ligne segmentale sont des lieux privilégiés pour les variations socialement distinctives" (<2>, p.254, note 53), nous proposons le modèle hiérarchique suivant pour le français contemporain:

1.1. Variation des mots à éléments flottants

On peut admettre que dans les formes sous-jacentes de nombreux mots (par ex. *hôtesse*, prononcé [ɔtɛs] ou [ɔtɛs] ; cf.<9>) il y a des segments, pour le reste ancrés dans le squelette, qui comportent un élément (au sens attribué à ce terme dans <4>) flottant, c-à-d. non associé. Une telle représentation phonologique de ces segments rendrait compte des alternances entre voyelles [+ATR] et [-ATR] dans les rimes non branchantes des syllabes non finales d'un grand nombre de mots français. Pour *hôtesse* on aura la forme sous-jacente suivante:



La réalisation de ce mot en surface sera [ɔtɛs] au cas où l'élément I se serait ancré dans le squelette et [ɔtɛs] si I est resté non associé. Par contre, le segment [ɛ] dans *hôtesse* ne contient aucun élément flottant, car il occupe le noyau d'une rime branchante, ce qui fait que ce segment n'est pas sujet à variation.

1.2. Variation des mots à segments flottants

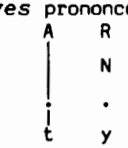
La variation de ce deuxième type s'explique par la présence, dans la représentation lexicale du mot en question, d'un segment entièrement flottant, ce qui veut dire que l'ensemble de ses éléments constitutifs se trouvent affecté par la fluctuation. Dans le cas du français, ce segment flottant peut être un schwa, une liquide devant schwa en contexte post-consonantique, une consonne de liaison. Cependant tous les schwas en français ne sont pas flottants (cf.<2>,p.230).

1.1. Deux formes sous-jacentes pour le même mot: à segment fixe/ à segment flottant

Le troisième type de variation caractérise certains mots appartenant à des classes fermées (cf. les concepts de mots syntagmatiques, mots phrastiques par opposition à mots prédicatifs <7>)et jouissant d'une grande fréquence dans le discours, pour lesquels on doit admettre au moins deux représentations phonologiques différentes: l'une contenant un segment donné en tant que fixe et l'autre, ce même segment en tant que flottant. Voici les deux formes sous-jacentes proposées dans <2>,p.214 pour le pronom clitique *tu* :

- avec voyelle fixe (par ex., dans *tu arrives* prononcé [tyariv])

- avec voyelle flottante (dans *tu arrives* prononcé [tariv])



Ces deux formes sont différemment distribuées en fonction de la variante socio-situationnelle (cf.<6>), mais aussi suivant des facteurs prosodiques et sémantico-pragmatiques (accent énonciatif, rhématisation, topicalisation, etc.; cf.<8>). Pour un inventaire, non exhaustif, des mots syntagmatiques et phrastiques du français contemporain qui auraient deux et parfois plusieurs représentations lexicales socialement distribuées, nous renvoyons à <3>, p.66-92.

2. SCHWAS ET FRONTIERES

Dans la suite nous voulons démontrer que le jeu d'association/désassociation du schwa, segment flottant ou épenthétique, dans les réalisations de surface est souvent régi par le contexte prosodique et morphosyntaxique et peut assumer, surtout dans certaines variantes sociosituationnelles, un certain rôle dans l'organisation hiérarchique de l'énoncé, en signalant le caractère jonctif/disjonctif d'une frontière.

- Deux corpus d'enregistrements authentiques, représentatifs de deux variantes du dialogue en français contemporain - l'interview (corpus A) et la conversation (corpus B) ont été analysés. Les locuteurs du corpus A se caractérisent par un degré d'intégration au marché linguistique supérieur à celui des locuteurs B. Les premiers possèdent un niveau d'instruction élevé et manient aisément plusieurs types situationnels de messages. Les productions orales des seconds se situent le plus souvent à l'intérieur d'une seule variante situationnelle: la conversation.
- Au total, 1481 occurrences de schwa ont été dénombrées dans le corpus A (dont 39,6% réalisées) et 1254 dans le corpus B (dont 33,5% réalisées). Si le taux global de réalisation des schwas dans les deux variantes de dialogue est assez proche , des divergences plus importantes ont été constatées au niveau distributions en fonction du contexte prosodique et morphosyntaxique.

2.1. Schwa en syllabe initiale de groupe prosodique

Pour faire apparaître l'incidence de la rupture prosodique entre la syllabe contenant schwa et le groupe précédent, les occurrences de schwa dans les trois contextes suivants ont été dépouillées:

- après une frontière de groupe accentuel: Fa;
- aprèsune frontière de groupe intonatif n'intégrant pas de pause silence: Fi(-p);
- après une frontière de groupe intonatif intégrant une pause réelle: Fi(+p).

Pour l'identification des Fa et Fi, les indices acoustiques et perceptifs établis par A. Di Cristo <1> ont été utilisés.

Conscientdu fait que la réalisation phonétique du schwa en français est une fonction à multiples variables, nous avons:
- neutralisé l'incidence du contexte segmental, ne retenant que les occurrences où schwa était précédé d'une seule consonne, c-à-d. les contextes:

...V/Fa/CəC...
...V/Fi(-p)/CəC...
.../Fi(+p)/CəC

(Dans la grande majorité des cas les schwas retenus étaient des schwas de monosyllabes clitiques.)
- négligé toutes les occurrences où le maintien du schwa pourrait être dû à un accent énonciatif sur la syllabe dont il constitue le noyau ou sur la syllabe suivante (cf.<5>,p.111;<3>,p.51), ce maintien étant régi par une contrainte rythmique et servant à éviter un antagonisme accentuel.

Voici les taux de réalisation de schwa obtenus:

Contexte	Corpus A	Corpus B
Fa	44,8%	28,8%
Fi(-p)	58,6%	52,6%
Fi(+p)	85,7%	72,8%

Ces résultats nous permettent de faire les conclusions suivantes:
- L'incidence d'une frontière de groupe prosodique immédiatement précédant une syllabe à schwa, sur le maintien de ce schwa dans la réalisation phonétique, est d'autant plus grande qu'il s'agit d'une frontière disjonctive.
- Cette incidence est plus nette dans le corpus A que dans le corpus B. Il apparaît donc que les locuteurs mieux intégrés au marché linguistique utilisent plus le caractère supposé de marque disjonctive du schwa réalisé dans le contexte en question.

2.2. Schwa à l'intérieur d'un groupe accentuel

A l'intérieur de groupe accentuel, trois contextes ont été distingués:
- à l'intérieur de mot: I;
- en fin de monosyllabe: Fm;
- en fin de polysyllabe: Fp.

Les taux de réalisation de schwa dans ces trois contextes étaient les suivants:

Contexte	Corpus A	Corpus B
I	40,9%	30,9%
Fm	42,5%	28,6%
Fp	9,0%	10,7%

En fin de polysyllabe, même précédé de deux consonnes, c-à-d. dans un contexte où on a l'habitude de l'entendre se réaliser, le schwa n'est prononcé que rarement. Cette même tendance a été signalée par V.Lucci <6> et par P.Léon <5>. On peut donc supposer que dans le contexte Fp, après deux consonnes, la non réalisation du schwa contribue à marquer la frontière finale d'un mot polysyllabique.

2.3. Schwa en finale de groupe prosodique

Les schwas que l'on trouve en fin de groupe prosodique sont aussi finals de polysyllabes. Cependant, nous avons constaté qu'une grande partie d'entre eux se réalisent dans le contexte

...CCə/Fa/C...

et contribuent ainsi à signaler le caractère jonctif de la frontière avec le groupe prosodique suivant.

Nous en concluons que les réalisations de schwa jouent probablement un certain rôle pour l'identification du caractère de jonction (en contexte ...CCə/Fa/C...) ou de disjonction (en contexte .../F/CəC...) d'une frontière entre groupes prosodiques. Le plus souvent, ce rôle n'est que redondant, la distinction étant assurée par un intonème (cf.<8>), mais nous supposons qu'au cas où la réalisation de cet intonème serait interdite par des contraintes phonotactiques ou syntaxiques, le jeu des schwas pourrait assumer à part entière cette fonction de marque jonctive/disjonctive.

REFERENCES

<1> DI CRISTO, A. (1985), *"De la microprosodie à l'intono-syntaxe, tome 2, Aix-en-Provence.*

<2> ENCREVE, P. (1988), *La liaison avec et sans enchaînement*, Paris: Editions du Seuil.

<3> JETCHEV G. (1988), *Marques phonostylistiques segmentales de deux variantes sociosituationnelles en français contemporain*, mémoire de maîtrise: Université de Sofia.

<4> KAYE, J., LOWENSTAMM, J. & VERGNAUD, J.-R. (1985), "The Internal Structure of Phonological Elements: A Theory of Charm and Government", *Phonology Yearbook, 2,* 305-328

<5> LEON, P.R. (1987), "E caduc: facteurs distributionnels et prosodiques dans deux types de discours", *Proceedings XIth ICPhS, 3,* 109-112.

<6> LUCCI, V. (1983), *Etude phonétique du français contemporain à travers la variation situationnelle*, Grenoble: Publications de l'Université des Langues et Lettres.

<7> MANTCHEV, K., TCHAOUCHEV, A. & VASSILEVA, A. (1986), *Traité de morpho-syntaxe française*, Sofia: Naouka i izkoustvo.

<8> ROSSI, M. (1985), "L'intonation et l'organisation de l'énoncé", *Phonetica, 42,* 135-153.

<9> WALTER, H. (1976), *"La dynamique des phonèmes dans le lexique français contemporain"*, Paris: France-Expansion.

PROSODIC AND PARALINGUISTIC FEATURES OF ONOMATOPOEIA IN KAYABÍ

Helga E. Weiss

Summer Institute of Linguistics, Brasília, Brazil.

ABSTRACT

This paper presents some aspects of the prosodic and paralinguistic features of onomatopoeia of informal oral communication in Kayabí (Tupian language of Brazil).

1. INTRODUCTION

Kayabí belongs to the Tupí-Guaraní language family of Brazil as classified by Rodrigues [7]. The language is spoken by about 600 Indians living in the rain forest of Mato Grosso and Pará of Central Brazil, where these data were gathered.

2. FUNCTION OF ONOMATOPOEIA

Onomatopoeia is used extensively in conversation and narration of texts, but less in formal and written style. Verboid onomatopoeic forms specify and add detailed information to the finite verb they accompany or substitute. This information can be aspect of the action (progressive, iterative, etc.), rhythmic pattern of the action (tempo, effort exerted, regularity of movement, etc.), or description of the actor or instrument (weight, size, number, etc.).

Verboid onomatopoeia form part of the lexical inventory of the language. They are not inflected like finite verbs, but can lexicalize into verbs by adding a causative prefix to the onomatopoeia.

Onomatopoeia referring to bird and mammal sounds are mostly echoic of their calls or noises. Many of the bird names are formed according to their echoic form.

Onomatopoeia involve the production of sounds, sound sequences, prosodic and paralinguistic features, all of which require unique physical movements and adjustments of the vocal tract, as well as rhythmic and phonotactic patterns that do not follow the phonological rules of the language. Imitations of bird and mammal calls show a greater variety of these phonetic phenomena.

The onomatopoeic words and phrases are generally followed by the inflected verb form /ʔjau/ 'say' or /imonou/ 'do', but can also occur without these, or even function on their own as a complete sentence. Thus onomatopoeic utterances show similarities to direct quotes.

One of the functions of onomatopoeia is imperative use, especially when talking to children with a limited vocabulary of the language (to make them sit, get up, sleep, etc.).

3. ARTICULATIONS AND PHONOTACTICS

The following are the phonemes of Kayabí:

p t k kʷ ʔ
m n n
f s
r
w j ɣ

i ĩ ɨ ɨ̃ u ũ
e ẽ a ã o õ

The permitted syllable pattern is CV in running speech and (C)VC utterance final.

Onomatopoeia show deviance in manner and place of articulation, adding the following segments and sequences:

[ɱ] [x] [r] [ɾ̥] [ʙ] [U] [ɬ] [Õ] [ə] [ɘ] [ɚ] [ʌ] [pɻ] [pr] [pɾ̥] [kʷr] [kɾ̥] [eᵘ] [Eᵘ].

Phonotactic differences are that the segments [ʔ] and [h] (voiceless vowel equivalent) can occur also word final, syllabic consonants [ɱ] [s̩] [x̩] [ɻ̩] [r̩] and [ɾ̥̩] as nucleus, and preglottalised fricative [ʔs] initially.

CVC can also occur utterance medial in sequences of onomatopoeic words and reduplication, accompanied by specific tempo, pitch and stress patterns. Examples are:

/'pōk 'pōk 'pōk/

(reduplication) 'chop lots'

/"pók "pɨ̄m/

(sequential) 'chop and fall'

4. PROSODIC AND PARALINGUISTIC FEATURES

4.1. Rhythm

Timing, stress, pitch, duration and iteration often co-occur to achieve the desired rhythmicality and aspect of the action.

Rapid speech (‿‿) expresses an action with fast movements, smallness, light weight, or repeated actions; whereas slow speech (‿) indicates slow movements, largeness, heavy weight, or fewer actions.

Staccato with pause breaks, often linked with lengthening, can indicate a heavy actor with slow movements; whereas legato indicates a continuous or prolonged motion. Speed-up in tempo signals increase in force of action, like 'coming-to-a-boil'. Duration can be of vowels and of syllables.

/"pɻ̄:'ɾ̥̄:'ɾ̥̄: "tóm 'tóm 'tóm/

'small fish bounce on water and fall'

/'pɾ̥̄::: 'ɾ̥̄::: 'tòm/

'large fish jump on water surface and fall'

/pɾ̥̄ 'tòm/

'large fish jumps high and falls'

/sóso'sók/ 'bird walks fast.'

/"sòk "sòk/ 'jaguar walks slowly'

/"só::ròk/ 'several pigs walk'

/sō:"rōk/

'one pig walks on leaves'

/só'rók só'rók só'rók/

'rat runs around'

/'mák 'mák 'mák 'mák/

'water starts to boil'

/'mak mak mak mak mak mak mak/

'water boils hard'

/tɨ̄/ 'small stone falls down'

/tɨ̃:::/ 'large stone falls down'

/pɨ̄m/ 'animal jumps once'

/'pɨ́m 'pɨ́m 'pɨ́m/

'animal jumps around'

/pɨ̄m 'sórok/

'monkey jumps and lands lightly'

/pɨ̀m "sò:::ròk 'tɨ̄ 'tɨ̄ 'tɨ̄/

'jaguar jumps, lands forcefully, and runs with heavy steps'

/wã́:m/ 'jaguar slanders along'

/wã̄:::m/ 'aeroplane rises up'

/'wām 'wām/

'howler monkey swings from branch to branch'

4.2. Pitch
Verboid onomatopoeic words and phrases spoken with high pitch signal a fast action (often accompanied by allegro tempo), light weight, small, lots; whereas low pitch indicates a slow action (often accompanied by legato tempo), heavy weight (accompanied by lengthening), few, or large.

/wã́ wã́ wã́ wã́m/

'mosquitos passing by fast'

/wã̀ wã̀ wã̀m/ 'aeroplane flying'

4.3. Phonation
Voice qualities used for echoic onomatopoeia are: creak (a̰), ingressive creak, falsetto creak, breathy voice (a̤) and whisper (ḁ).

/"m̰: "m̰m 'm̰ 'm̰/ 'currasow call'

/mɛ̄:?/ 'paca noise'

/kà/ 'macaw call'

/å:::/

(laryngealised and ingressive) 'howler monkey noise'

/pɨ̄'ɾɨ̄ pɨ̄'ɾɨ̄ pɨ̄'ɾɨ̄ x3/

'paca flaps ears'

/'kx̀ʊ̀ 'kx̀ʊ̀ 'kx̀ʊ̀ x3/ 'toucan call'

4.4. Articulatory Control
Some animal imitations are articulated with different articulatory gestures such as constricted larynx (small monkey), lowered larynx (large monkey), a more precise and forceful articulation accompanied by speed up of reduplications (chopping), unrounding of rounded vowels (noise of howler monkey), and other modifications of the vocal channel which are combinations of several of the above gestures.

/ḿ m̀ m̄ m̀ m̄ m̀/

(low larynx) 'piranha fish jumps up to get leaves'

/å̄:::/

(ingressive creak) 'howler monkey'

4.5. Initiation of Airstream
Some echoic forms are pronounced with an ingressive airstream, such as noises made by certain fish, monkeys, and other animals.

/è̤ è̤ è̤ è̤/

(ingressive breathy) 'capibara in water'

/ə̤̄ ə̤̄ ə̤̄ ə̤̄ ə̤̄ /

(ingressive, whispered-breathy) 'spider monkey'

4.6. Reduplication
Onomatopoeic words can be repeated up to eight times to signal increased length of time or space, the number of repeated actions, actors, or participants. Reduplication describes iterative actions, or sequential happenings.
Many of the examples given in 4.2 fall into this category. Other examples are:

/tò̰k/ 'give to one person'

/tó̰k tó̰k tó̰k/ 'give to several persons'

5. ACOUSTIC ICONICITY
Acoustic representations of iconic sounds show the following tendencies:

Initial fricatives give the idea of entering, piercing:

/sṵ̄k/ 'fall into, pierce'

A final stop seems to be telic and imitative of the forceful endpoint of the action (such as hit or fall):

/pò̰k/

'something hard or heavy falling or being cut'

A final nasal represents a continuous sound:

/tɛ́n/ 'ring, rattle'

A voiceless trill indicates a series of quick movements:

/pr̥̄/

'humming bird flying on the spot, or a small fish jumping along the water surface'

A final vowel indicates a continuous movement similar to an atelic aspect:

/tɨ̄ tɨ̄ tɨ̄/ 'walk along'

A high front vowel seems to convey the idea of smallness or fear:

/kĩ̄/ 'noise of tapir pup'

/wĩ́:n/

(weak voice) 'noise of spirits'

A high back vowel gives the idea of largeness:

/kò/ 'noise of a howler monkey'

6. CONCLUSION
In Kayabí an action that is expressed by a generic term can be further qualified by a vast choice of onomatopoeic words and phrases, transmitting detail about the performer and performance to reinforce the oral message thus making it understood and vivid.
Tape recordings of several hundred verboid and echoic onomatopoeia made by several men show that onomatopoeic forms are not coined arbitrarily by the speaker, but are part of the lexicon of the language and reflect a specialised knowledge relevant to the community's environment and everyday life.

REFERENCES
[1] BROWN, R. (1963), *"Words and things"*, New York: The Free Press.
[2] CRYSTAL, D. (1976), *"Prosodic systems and intonation in English"*, Cambridge: Cambridge University Press.
[3] LADEFOGED, P. (1980), "What are linguistic sounds made of?", *Language*: 56, 485-502.
[4] LAVER, J. (1980), *"The phonetic description of voice quality"*, Cambridge: Cambridge University Press.
[5] LEHISTE, I. (1970), *"Supra-segmentals"*, Cambridge: Cambridge, Mass.: MIT Press.
[6] OHALA, J. (1983), "Cross-language use of pitch: an ethological view", *Phonetica*: 40, 1-18.
[7] RODRIGUES, A. D., (1958), "Classification of Tupí-Guaraní", *IJAL*: 24, 231-244.

THE "OLDER" VOICE

Harry Hollien, Ph.D

University of Florida, Gainesville, FL 32611

ABSTRACT

-A number of speech and voice changes are associated with advancing age -- even in individuals who are normal. Apparently, they are due to tissue loss and reduction in mobility (the physiological model). However, a second theory is needed to account for other alterations: it is the Male-Female Coalescence Model of Vocal Aging. Specifically, it has been established that, at puberty, the sexes become biologically less like each other; these processes appear to reverse during female menopause (and its counterpart in males) with the sexes becoming more alike. Thus, the cited model supplements the older theory and permits more accurate predictions of those speech changes which correlate with advancing age.

1. INTRODUCTION

-Certain voice and speech alterations appear to accompany the aging process. However, the nature and extent of these changes is not well understood. For one thing, it has been argued [11] that they are due to pathologies that are associated with old age. However, this position has been sharply countered by a large number of authors [6,8,17]. It is conceded, of course, that the elderly can exhibit pathologies of many types (and that some of them can lead to changes in speech and voice). Yet, it also can be expected that, while cohorts of normally aging people do not suffer from these deficit related changes, they nonetheless will exhibit shifts of some type -- and there is evidence to support this postulate. For one thing, it has been found that auditors are fairly good at accurately estimating the age of talkers from their speech [9,13,16]. If some sort of change had not occurred, re: the older talkers, these judgements simply could not have been made. But, what are these changes? Do they result from growing old (chronological age), from physiological changes, or from some other set of factors? Of course, chronological age plays a major role in the process as the shifts must be related to the passage of time. However, are they well documented physiologically or do other factors also operate? Two theories appear useful in this regard; they are the physiological model of vocal aging and the male-female coalescence model.

2. THE PHYSIOLOGICAL MODEL

-In the past, most investigators have subscribed to a theory that can be referred to as the "physiological model" of vocal aging. Although not always articulated as such, this theory explained the normal aging process as one that results from a reduction in the efficiency of the human biosystems as a function of time. Specifically, the changes which are observed in the elderly are said to be due to tissue atrophy/reduction and an associated loss of mobility. (See Figure 1.)

-In a sense, the physiological model was developed (at least informally) in response to the question "What is old?" [4,17]. In this regard, it was observed [1-4,8] that the chronological age of a person does not always appear to best represent their "actual" age. Indeed, if these (and other) authors are to be believed, the aging process is neither linear nor invariable. That is, it would appear necessary to directly assess the mental, sensory, motor and communicative capabilities of older individuals in order to determine their true age. It is not argued, of course, that no degradation occurs; rather it would appear that the process is one that varies -- and sometimes dramatically -- from person-to-person. Hence, the physiological model of aging served to replace the traditional chronological metric. Unfortunately, even today, this model is far from complete. A brief review of the relevant factors may be found in Hollien [8], and, of course, these relationships are being continually updated. On the other hand, even though this physiological model accounts for many of the changes which occur as a person ages, there still are discrepancies.

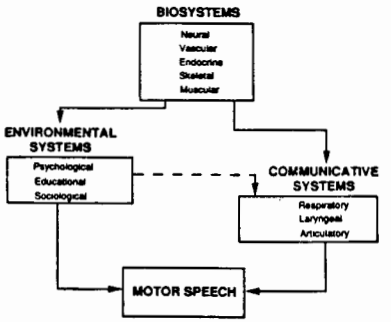


Figure 1

2.1 Relevant Data

-It is without question that most of the research carried out on the communication characteristics of older persons has been focused on the deficits resulting from one form of pathology or another. Moreover, even when normal subjects are the focus of interest, it is the human voice that tends to be studied -- and the discussion to follow will reflect these biases. Specifically, the vocal/speech correlates of aging include speaking fundamental frequency (especially), vocal intensity, speech spectra, timing and, in some instances, perceptually based information.

-Speaking fundamental frequency (SFF or FØ) has been studied extensively in populations of all ages (see Figure 2). As can be observed, SFF levels shift markedly during adolescence with this lowering much greater for males than females. The data for mature males suggest that FØ is further lowered during adult life but then begins to rise as middle age is concluded -- and perhaps rises sharply as male cohorts reach an age where they can be classed as elderly. This pattern is one that could be predicted on the basis of the physiological model. The configurations for females follow a slightly different course. The downward shift in SFF is seen at puberty -- even though it is less extensive than that for males -- and then SFF levels appear not to change much during life, excepting perhaps for a very slight rise in the elderly. On the other hand, there now are data which support the notion of a downward shift in female SFF as a function of old age [5,12,18]. This change is one which is contrary to that predicted by the physiological model. In that instance, the atrophy associated with aging would be expected to force a rise in female SFF -- just as it does in males. Apparently, this does not happen.

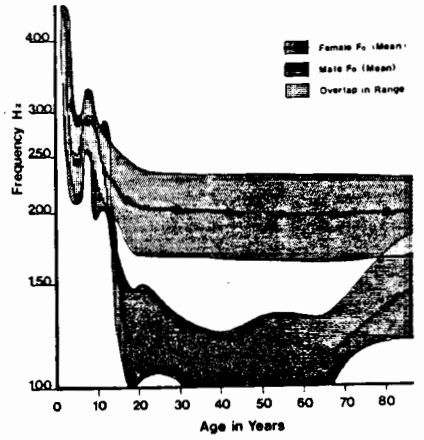


Figure 2.

-Other 'old age' related speaking characteristics also have been reported. For example, the vocal loudness levels of older individuals are thought to be greater than those for younger people [13,15] and others [10] argue this position even for older populations who show no evidence of significant hearing loss. Physiological findings, however, are difficult to reconcile with the above statements. For example, certain investigators [14,15] have reported a decline in intraoral breath pressure and vital capacity with increasing age. On this basis, it does not seem logical that increases in vocal loudness would be observed in the elderly. However, if they were, they might be gender related.
-Vocal fry is the lowest voice register to be found on the frequency continuum produced by the human voice [7]; it may contribute to voice tremor [15] -- and voice tremor would be predicted by the physiological model. Finally, other voice and vocal tract changes have been observed; breathiness is an example [15]. This factor, coupled with a slowing of articulation and a reduction in phonational frequency range, could serve to explain why individuals of advanced age can be recognized as older simply from listening to their speech. Additionally, many of the relationships noted are consistent with the physiological theory. On the other hand, there are enough data in variance with this theory that it would appear necessary to modify or supplement it.

3. THE MALE-FEMALE COALESCENCE THEORY

-The male-female coalescence model of aging is not at all unique to human communication. Indeed, it is employed to explain many of the (other) events/changes associated with aging. Briefly, this model suggests that menopause (and its physiological counterpart in males) is a functional reversal of the sex oriented changes that occur at puberty. As stated, it would appear that hormonal effects at pubescence operate differentially on males and females with the greatest changes occurring in the male. Thus, the sexes become biologically less like each other at puberty whereas, at menopause (and during subsequent life-changes associated with advancing years) these processes appear to shift in the opposite direction -- i.e., males and females become more like each other. Of course, it is conceded that this model will explain only a portion of the observed processes and that aging affects behavior, physiology and communication in other ways. For example, tissue atrophy, reduction in strength due to muscle deterioration, changes in neural function and reaction to environmental changes, also will degrade communicative skills. This model simply accounts for a limited set of relationships in the process -- a set that previously has been a little difficult to explain.
-Application of the coalescence model would result in predictions reflecting reversals in some of the shifts which took place at puberty. For example, it could be employed to predict that SFF for the male would rise as a function of advancing age but that it would not do so for females. That is, the interaction between elements related to both theories would permit the suggestion that SFF in females either would not shift dramatically or be lowered -- and this effect is now documented [12,18]. Finally, other gender related changes could be expected, in part based on the differential increase in strength at puberty. Unfortunately, even though there are suggestions that these differences occur, definitive information currently is lacking -- primarily (1) due to the fact that little research (re: communication) has been carried out on these issues and (2) due to the confounding effects of those pathologies which exist in many of the elderly.

4. CONCLUSIONS

-The male-female coalescence model of vocal aging has proved to be an excellent supplement to the basic physiological model. While, this coalescence theory is in-and-of-itself physiologically based, it nevertheless serves to modify the larger theory in ways which permit more robust predictions to be made on the nature of vocal aging.

5. REFERENCES

[1] BILASH, JE; ZUBECK, JP (1960), "Effect of Age on Factorially Pure Mental Abilities", J. Gerontol., 15:175-182.
[2] BOTWINICK, J (1973), "Aging and Behavior", New York, Springer.
[3] CASSELL, K (1979), "A Time Clock in Your Genes", Science Dig., 86:57-60.
[4] CHOWN, SM; HERON, A (1965), "Psychological Aspects of Aging in Man", Psychol. Rev., 16:417-450.
[5] de PINTO, O; HOLLIEN, H (1982), "Speaking Fundamental Frequency Characteristics of Australian Women", J. Phonetics, 10:367-376.
[6] GREENBERG, J (1979) Old Age: What Is Normal? Science News, 115:284-285.
[7] HOLLIEN, H (1974) On Vocal Registers, J. Phonetics,2:125-143.
[8] HOLLIEN, H (1987) "Old Voices: What Do We Really Know About Them?" J. Voice, 1:87-101.
[9] HUNTLEY, R, HOLLIEN, H; SHIPP, T (1987) "Influence of Listener Characteristics on Perceived Age", J. Voice 1:49-52.
[10] HUTCHINSON, JM; BEASLEY DS (1976) "Speech and Language Functioning Among the Aging", in Aging and Communication, Baltimore, University Park Press.
[11] JEROME, EA (1950) "Age and Learning", in Handbook of Aging, Univ. Chicago Press, 661-697.
[12] KROOK, MIP (1988) "Speaking Fundamental Frequency Characteristics of Normal Swedish Subjects", Folia Phoniat, 40:82-90.
[13] PTACEK, PH; SANDER EK (1966) "Age Recognition From Voice", J. Speech Hear Res., 9:273-277
[14] PTACEK, PH, et. al. (1966) "Phonatory and Related Changes with Advanced Age", J. Speech. Hear. Res., 9:273-277.
[15] RYAN, WJ; BURK, KW, (1974) "Perceptual and Acoustic Correlates of Aging in Males", J. Comm Disorders, 7:181-192.
[16] SHIPP, T; HOLLIEN, H (1969) "Perception of the Aging Male Voice", J. Speech Hear Res., 12:704-710.
[17] TIMIRAS, PS, (1978) "Biological Perspectives on Aging", Amer Scient, 68:605-612.
[18] YAMAZAWA, H; HOLLIEN, H "Speaking Fundamental Frequency Patterns of Japanese Women", in press.

ON INGRESSIVE GLOTTALIC AND VELARIC ARTICULATIONS IN XHOSA

J.C. Roux

Dept. of African Languages, University of Stellenbosch, Stellenbosch, South Africa

ABSTRACT

The aim of this paper is to present some data on the bilabial implosive and on click articulations in Xhosa, a language belonging to the Bantu group of languages. Variations in implosive production will be enunciated paying specific attention to the inadequacy of phonological features to account for phonetic differences among languages. Distributional characteristics of certain click types in Xhosa wil then be considered. An articulatory phonetic motivation for the occurrence of these specific forms will be proposed.

1. IMPLOSIVE BILABIAL

Xhosa exhibits one bilabial implosive [ɓ], orthographically presented as **b**, when not present in a nasal combination, eg. **abantu** [abant'u]. Impressionistic phonetic descriptions futhermore refer to two other bilabial plosives occurring in this language, viz. a bilabial plosive with full breathy voice in nasal compounds, eg. **imbuzi** [imbuzi], and a bilabial plosive with delayed breathy voice occurring in non-nasal environments, eg. **-bhala** [ɓ̤ala] [1]. A computer assisted phonotactic analysis [3] of the sound system of Xhosa based on grapheme to phoneme conversions following the above mentioned conventions indeed indicated a high occurrence of the implosive vis-a-vis the other two plosives types. An analysis of 294 965 /CV/-syllables yielded 8,2 % implosives in the /C/-position, with 1,2 % breathy voiced plosives and 0,42 % delayed breathy voiced plosives in the same position. In real life, however, the phonetic qualities of the implosive in Xhosa seem to change quite extensively, inter alia, as a function of tempo. Figure 1 represents the articulation of an intervocalic bilabial implosive, produced at a relatively slow (deliberate) speaking rate. This articulation may be regarded as a "classical" implosive sound. Total closure of the vocal folds followed by some amount of pre-voicing prior to the bilabial release may clearly be noted. This observation renders some support for the traditional view [1] that the vocal cords may start to vibrate due to a downward movement of the larynx through the column of subglottal air. However, it is also clear that distinctive timing sequences prevail: the voicing follows a glottal closure which in itself may be necessary to induce rarefaction. Figure 2 presents the same intervocalic sequence, this time embedded in the word **abafana** "boys", produced at a relatively faster speaking rate. Although the articulation is also clearly implosive in nature, both auditorily and articulatorily, there is a marked difference in the acoustic spectrum during the closure phase of the implosive. Voicing continues throughout the closure phase suggesting no specific timing with respect to glottal closure. The extent of this phonetic variation is difficult to determine. Free variation may

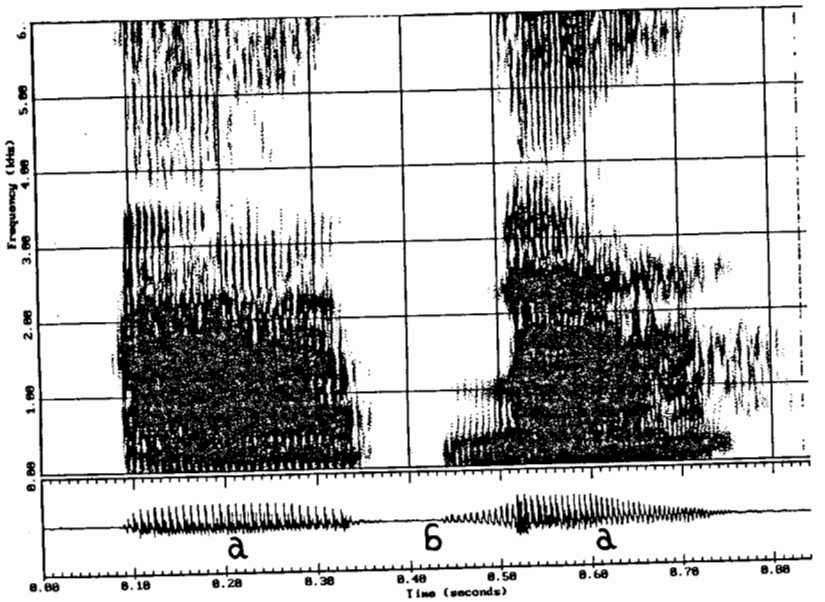


Figure 1 Implosive bilabial in /aba/

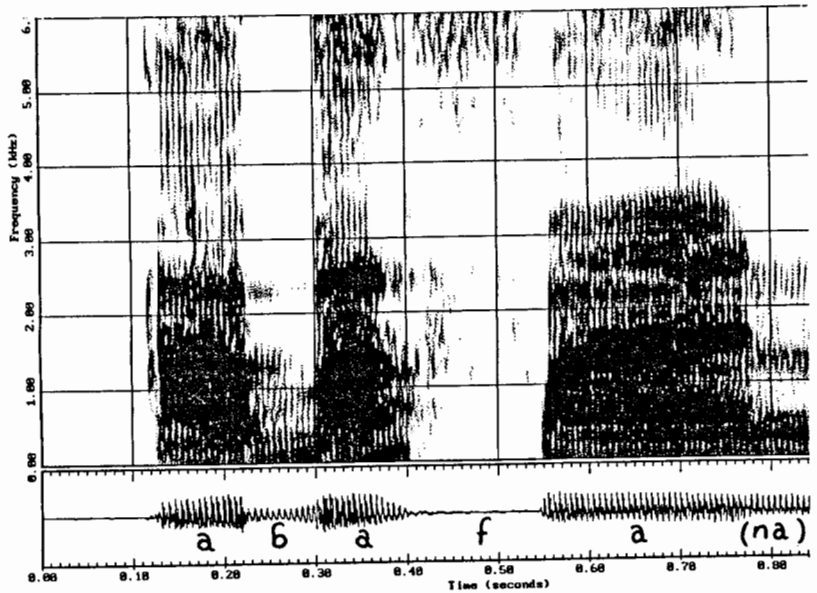


Figure 2 Implosive bilabial in /abafana/

take place within individuals irrespective an increase in tempo, whilst in other cases the variation seems to be tempo related. Instances were also recorded [3] where implosives totally lost their ingressive qualities in fast speech to be articulated as voiced bilabial plosives as are usually found in nasal combinations. This phenomenon, however, may possibly be explained in terms of unattained articulatory targets.

The variation in the phonetic qualities of implosives in Xhosa brings to mind similar types of differences described by Ladefoged [2] for implosives in Hausa and Kalabari, two languages spoken in Nigeria. In Hausa, implosives display laryngealized voicing throughout the closure, whilst implosives are fully voiced during the closure with no tendency toward creaky voice or laryngealization in Kalabari. Apart from sharing the voiced version with Kalabari, Xhosa augments the possible implosive articulatory repertoire with a distinct closure of the glottis followed by an amount of prevoicing prior to the release. These observations clearly have, as Ladefoged [2] has pointed out, serious implications for a theory of (universal) distinctive features as it becomes virtually impossible to account for linguistically significant differences between languages. It therefore remains an open question on what the exact phonetic content of a distinctive feature such as, for example, [implosive] should be. This, in principle lends support to Ladefoged's well known view that "..phonological features are certainly not sufficient for specifying the actual sounds of a language; nor are they in a one-to-one relationship with the minimal sets of parameters that are necessary and sufficient for this purpose." [2].

2. CLICKS

Xhosa exhibits three basic click types, all of which are produced with ingressive air mechanisms. These basic click types (dental [|] orthographic **c**, alveo-lateral [||] orthographic **x**, and alveo-palatal [!] orthographic **q**) may furthermore be accompanied by various secondary features such as aspiration, voicing, nasalization, and breathy voicing. It is important for the following discussion to be aware that clicks are produced with two occlusions, i.e. at a point in the front region of the oral cavity, as well as a closure at the velum. Subsequent backward movement of the tongue across the velar area is necessary to induce rarefaction [4]. The computer-assisted phonotactic analysis mentioned above [3] revealed interesting phonotactic patterns, some of which suggest phonetic conditioning. It should, however, be pointed out that /CV/-syllable combinations containing any one of fifteen possible clicks (phonetic varieties included) as initial segments are relatively rare, representing only 2,78 % of the total corpus of 294 965 combinations. Hence, typification of the Nguni languages (Xhosa and Zulu) as unique in this respect, is obviously not based on quantitative values. Of these clicks, the plain version (1,45 %) and the nasalized versions (0,62 %) constitute the bulk of these occurrences. With Xhosa entertaining a five vowel system, the following most frequent combinations occur in decreasing (vertical) order for each type (e and o represent mid-low vowels):

	DENTAL	ALV-LAT	ALV-PAL
PLAIN			
	|a	||e	!a
	|e	||a	!o
	|i	||o	!u
NASALIZED			
	$\tilde{|}$i	$\tilde{||}$a	$\tilde{!}$a
	$\tilde{|}$e	$\tilde{||}$e	$\tilde{!}$u
	$\tilde{|}$a	$\tilde{||}$i	$\tilde{!}$i

(The "missing" two vowels in each case have been omitted, due to their extremely low occurrence rate, i.e. a rate of less than 0,003 %. In some instances these vowels do not even occur whatsoever [3].)

In all cases above /a/ seems to have default status occurring irrespective of any other vowel category.

Dentals, both plain and nasalized, show a clear preference for front vowels, which implies that styloglossus (and possibly palatoglossus) activity which is responsible for tongue-velar closure (to induce rarefaction), is overrided by activity of the genioglossus muscle maintaining the position of the tongue in the anterior region of the oral cavity.

Plain alveo-lateral articulations seem to favour mid-low vowels. If it is taken into account that the alveolar closure is maintained during lateral release where the side of the tongue is lowered to a position towards the middle of the oral cavity, then it may be expected that mid vowels will tend to follow.

In plain alveo-palatal articulations the preference seems to be for back vowels to follow these clicks. In these articulations the tip of the tongue is very active performing a partly retroflex movement in which the inferior longitudinal muscle as well as the genioglossus are most probably involved. Considering the back and downward movement of the dorsal part of the tongue (to induce rarefaction) as well as the retroflex movement of the tip of the tongue, it seems quite plausible that the following articulation could also be in the posterior area of the oral cavity, hence the preference for back vowels.

In nasalized alveo-lateral articulations the preference for mid vowels seems to give way to front vowels. This preference is also shared by nasalized dentals. It is probably save to assume that in both cases the pull of the tongue towards the dental and alveolar regions for the primary occlusions overrides any pull or lift towards the posterior area. Activity of the palatoglossus muscle which lowers the soft palate as well as a lack of levator palatini activity may comcommitantly contribute to a lack of activity in the back of the oral cavity, giving rise to a preference for front vowels.

Finally, nasalized alveo-palatal clicks frequently combine with high front and back vowels. It should be borne in mind that the articulation of the alveo-palatal click requires a high raising, and even bunching of the body of the tongue in the palatal and velar regions. These two occlusions are relatively close to one another, compared to dental clicks where the points of occlusion are relatively far apart. Add to this the active lowering of the velum, hence a narrowing of the area between the roof of the oral cavity and the body of the tongue, then it comes as no surprise that high vowels are preferred in a position following a nasalized click.

The patterns described above generally seem to hold true for voiced as well as for nasalized breathy click articulations, albeit the incidence of occurrence is extremely limited. The only counterexamples are found with aspirated clicks where dentals and alveo-palatals seem to interchange. Aspirated dentals seem to prefer a following back vowel, whilst aspirated alveo-palatals tend to have a preference for front vowels in the following position. No obvious phonetic explanation seems to be at hand for this phenomenon.

4. REFERENCES

[1] FINLAYSON, R, JONES,J. ET AL. (nd), *An introduction to Xhosa phonetics*, Hout Bay: M.Lubbe Publishers.

[2] LADEFOGED, P. (1980), "What are linguistic sounds made of?", *Language*, 56, 485-502.

[3] ROUX, J.C. (1990), *Phonetic elements of the sound system of Xhosa* (Written in Afrikaans), Pretoria, HSRC-Report, 267 pp.

[4] TRAILL, A. (1985), *Phonetic and Phonological Studies of !Xoo Bushman*, Hamburg: Helmut Buske Verlag.

PHYSIOLOGICAL PROPERTIES OF "BREATHY" PHONATION IN A CHINESE DIALECT
—A FIBEROPTIC AND ELECTROMYOGRAPHIC STUDY ON SUZHOU DIALECT—

Ray Iwata*, Hajime Hirose**, Seiji Niimi**, Satoshi Horiguchi***

Faculty of Humanities, Shizuoka University, Japan*
RILP., Faculty of Medicine, University of Tokyo, Japan**
Tokyo Metropolitan Neurological Hospital, Japan***

ABSTRACT

Physiological features of the difference in phonation types were investigated on Suzhou Chinese by use of fiberoptic endoscopy and electromyography. The findings suggest that "normal" vs. "breathy" opposition in phonation type in Suzhou should be brought about by antagonistic setting in the larynx.

1. INTRODUCTION

It is known that there is an interesting interaction between initial consonants, vowels and tones in Wu dialects in China. Recent phonologists often mention the term "phonation type" in treating this phenomenon, but the physiological reality of it is still unclear. In Wu dialects, a breathy (or murmured) syllable is initiated by so called "muddy" initial (usually indicated by a phonetic letter [ɦ], like in [pɦ], [tɦ], [kɦ], [sɦ] and [ɦ]), followed by a breathy vowel with low pitch initiation; whereas a normal (or clear) syllable is initiated by a "clear" initial (i.e. voiceless inaspirates and aspirates), followed by clear vowel with high pitch initiation.

Experiments were conducted on Suzhou dialect, one of the main dialects in Wu area, to reveal the physiological aspects of the difference between "breathy" and "normal" phonation. Seven lexical tones are discriminated in Suzhou as described below by tone letters. Among seven tones Tones IVa and IVb, are characterized by shorter duration in their syllables than in other syllables as indicated by one numeral(5) or two numerals with an underline(23)[1].

Yin Tones	Yang Tones
Ia 55	Ib 24
IIa 52	
IIIa 412	IIIb 231
IVa 5	IVb 23

The normal phonation and clear initials are associated with Yin tones (indicated by "a"), and the breathy phonation and muddy initials with Yang tones("b").

2. PROCEDURE

Laryngeal views were observed by a flexible fiberscope and were recorded on VTR at a rate of 30 frames (60 fields) per second. The intraoral pressure (Po) was simultaneously measured by introducing a miniature pressure transducer through the nostril to the mesopharnx. Electromyographic(EMG) recording was made on the same day but separately from the fiberoptic experiment. The electrodes were inserted into the cricothyroid (CT), thyroarytenoid (Vocalis, VOC) and sternohyoid (SH). The EMG signals were rectified and integrated over a period of 5 ms. and sampled at a rate of 1 kHz.

In the experiments syllables with zero initials and dental stops, [i]/[ɦi], [ti]/[tɦi], were uttered, first in isolation and second in the carrier sentences:
A: [li55 kø41_____kɦə21 kə25 z⊥31] "He looks at this character_____"
B: [kɦə 21 kə25 z⊥ 31 in55 ta25 ____iI 25 iã412] "This character, its pronunciation is the same as ____"

3. Results

3-1 Acoustic evidence

Spectrographic observations show that the breathy phonation is characterized in vowels by frication components at the higher frequency range with the damping of the upper formants. Closure duration and VOT for [t]/[tɦ] were measured and the results of the measurements are shown below. VOT is identified here as the interval from the release point to the onset of the periodical vocal wave since it was often the case with the muddy stops that the exact value of VOT was hard to detect by oscillographic and spectrographic inspection.

Closure Duration(ms.)

	Avg.	Min-Max.	Std.	No.
[t]	165.8	131.0-225.2	20.9	30
[tɦ]	128.2	99.9-171.9	23.3	21

t-test: p<1%

VOT(ms.)

	ToneI,II,III			ToneIV		
	Avg.	Std.	No.	Avg.	Std.	No.
[t]	12.2	2.4	22	8.5	1.3	7
[tɦ]	20.6	3.9	14	10.8	0.6	6

t-test: p<1%

Closure duration is significantly longer in unaspirated stops than in muddy stops. VOT in the muddy stop ([tɦ]) is invariably positive. It was reported that there was no significant difference in VOT between muddy stops and voiceless unaspirated stop[2]. But the present analysis has revealed that the difference is statistically significant if the VOT is defined as above.

3-2 Intraoral air pressure(Po)

Maximum Po in initial stops is lower in [tɦ] than in [t]. The difference is significant at the level p<5% in carrier sentence A; p<2% in carrier sentence B; and totally p<1%.

Po(mmH_2O)

	sentenceA			sentenceB		
	Avg.	Std.	No.	Avg.	Std.	No.
[t]	106.9	18.3	16	122.9	14.8	8
[tɦ]	91.4	15.4	12	107.6	12.8	6

3-3 Direct observation of the larynx

Fig.1 shows representative frames of the glottal views in the normal([ti55]) and breathy ([tɦi24]) phonation. It can be observed in the figures that the anterior-posterior dimension of the supraglottal structure is remarkably decreased for breathy phonation. The constricted gesture, which can be called "ary-epiglottic constriction"[3], is observed throughout the entire syllable with an increasing degree, but is weakened at the end of the syllable. Note that the adductive movement of the false vocal folds, which is a characteristic feature in the "glottal stop"[1], does not take place in this type of constriction. It appears that the whole larynx moves downward in breathy phonation. The state of the glottis as well as other related features is summarized below.

1) Vocal initiation([i]/[ɦi])

The glottis appears to be closed both in normal and breathy phonation at the initiation of the syllables. In the normal phonation a syllable is often initiated by the glottal stop which is characterized by the adduction of the false vocal folds. In the breathy phonation the glottal stop is definitely absent.

2)Consonantal initiation([ti]:[tɦi])

No remarkable difference is observed in the glottal feature between the two consonant types. There could be four states of the glottis: (1)both cartilaginous and membranous(ligamental) portions of the glottis are open, (2)only cartilaginous portion is open, (3) only membranous portion is open, (4)both of them are closed; and every one of the four states is observed both in [t] and [tɦ], causing no glottal vibration.

3-4 Electromyographic findings

Fig.2 shows the averaged EMG signals for CT, SH and VOC.

VOC: In normal phonation VOC is activated at the vocal initiation, while in breathy phonation it is

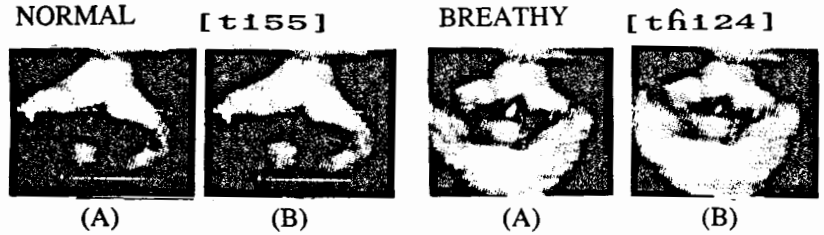


Fig.1 Selected frames of the laryngeal views for Suzhou [ti55](normal) and [tɦi24](breathy) uttered in isolation. In both syllables laryngeal views are selected from beginning part of the vowels; (A) approximately corresponds to the point of oral release, and (B) to 40-70 ms. after release.

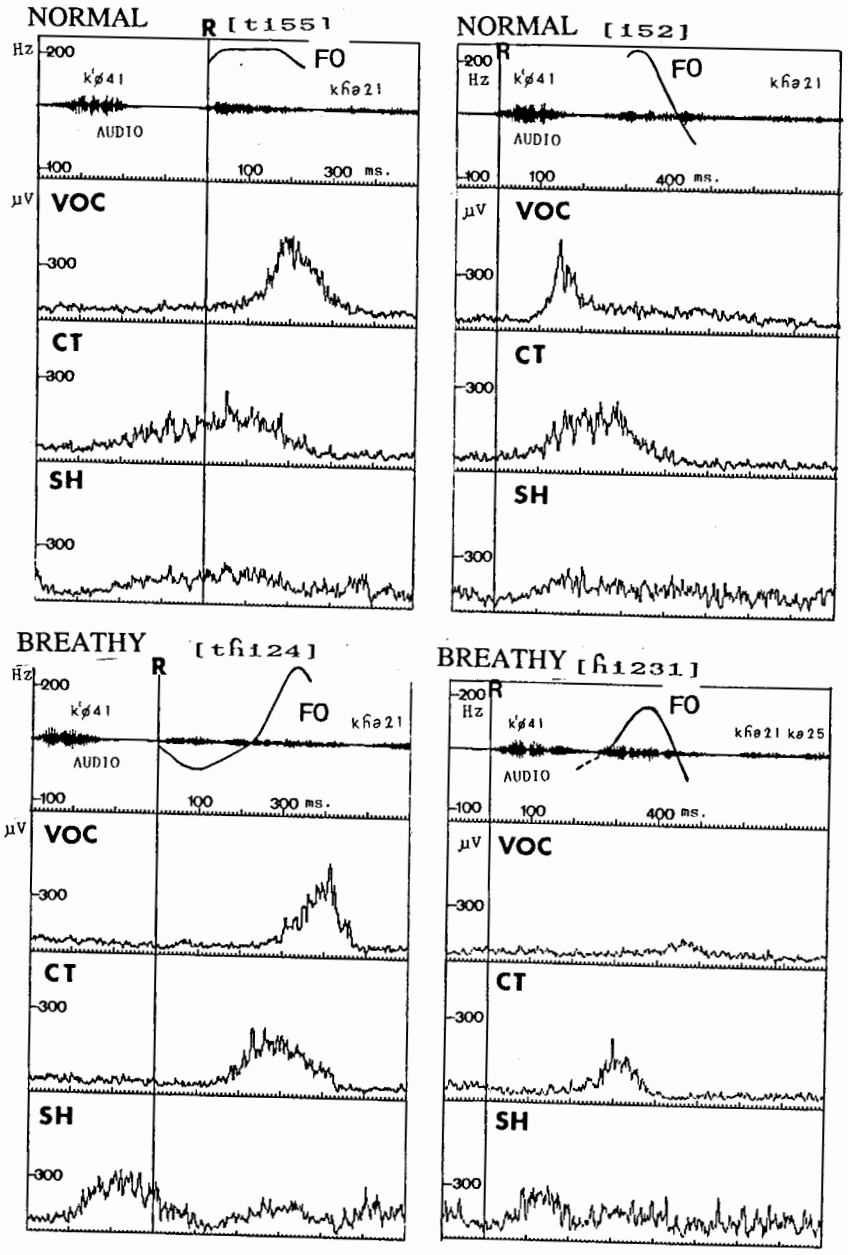


Fig.2 Averaged EMG signals of VOC, CT and SH for "normal" phonation and "breathy" phonation. The vertical line in [ti55] and [tʰi24] indicates the moment of /t/ release ; and in [i52] and [ʰi231], the moment of /k/ release for the syllable preceding the test word. A typical contour of F0 is shown for each test word.

suppressed at the initiation. This evidence is in good conformity with the fiberoptic observation. In the syllables with high or rising F0 contour VOC activity increases at the syllable final portion showing a reciprocal pattern with CT. This is related with the pitch control of the tones as well as the vocal termination of the syllable[1].

SH, CT: The activity of SH and CT is basically antagonistic at the beginning of the syllables: SH is activated for breathy phonation and CT for normal phonation, though SH activity increases to a extent as CT is activated in normal phonation. Both SH and CT show the early initiation in their activities; in the syllables with dental stops, they start their activities at around the closure point or even earlier. In other words, high and low pitch initiations are preceded by early activation of CT and SH.

In pitch rising and falling, however, CT and SH are not antagonistic. CT evidently participates in pitch raising (see EMG for [tʰi24] and [tʰi231]), but SH does not show any marked activity in pitch falling(see EMG for [ti52]).

3-5 Evidence observed in bisyllabic words

The muddy initials have been reported to be realized in fully voiced consonants in connected speech. In the experiment a set of bisyllabic words which have normal or muddy initials in the second syllable were also examined. VOT in muddy stops is in most cases negative and the glottis(both cartilaginous and membranous portions) is closed. The closure duration and peak Po in muddy stops are significantly shorter /lower than in normal stops.

4. Discussion

It is suggested that the difference in phonation types should be produced by the antagonistic setting in the larynx.

The breathy phonation is characterized by "ary-epiglottic constriction", with the downward movement of the larynx. The activity of SH, and presumably other extrinsic muscles as well, undoubtedly contribute to form the constriction and the downward shift of the larynx. These muscles adjust the framework of the larynx as a whole, then externally or vertically effecting the tension of the vocal folds[4][5]. Note that VOC and CT are suppressed at the initiation of the breathy phonation. Conceivably increased activity of the extrinsic muscles would shorten and thicken the folds by exerting the forces externally on them, its adductive tension being decreased[3]. The "breathy" quality of the syllable might be brought about by a "slack" state of the vocal folds, which would provide a favorable condition for low pitch initiation. And this may also be a reason the vocal folds start vibrating in the intervocalic positions.

The normal phonation, on the other hands, is initiated by the increased activity of VOC in the vocal initiation and that of CT in the consonantal initiation, the former of which is often accompanied by the adductive gesture of the false vocal folds. VOC contributes to increase the adductive tension of the vocal folds by supplying the medial compression[6]. CT is primarily a pitch raiser, but note that its activity is initiated quite early. It is assumed that CT also participates in increasing the adductive tension of the folds[6]. Thus a "stiff" state of the vocal folds in normal phonation is unlikely to cause the vibration and would provide a favorable condition for high pitch initiation.

REFERENCES

[1] Iwata,R.,H.Hirose,S.Niimi and S.Horiguchi(1990): "Syllable final "glottal stop" in Chinese dialects", Ann. Bull.RILP,No.24.

[2] Shi,F.(1983):"Acoustic features of muddy initials in Suzhou dialect" (in Chinese) Yuyan Yanjiu, 1983-1.

[3] Lindqvist,J.(1972): "A descriptive model of laryngeal articulation in speech" STR-QPSR,1972,2-3.

[4] Sonninen,A.(1968):"The external frame function in the control of pitch in the human voice" Ann. N. Y. Acad.Sci.,Vol.155.

[5] Ohala,J.(1972): "How is pitch lowered?" JASA,52.

[6] Hirose,H.,T.Gay(1973):"Laryngeal control in vocal attack" Folia phoniat.25.

INTRASPEAKER VARIATION ON THE SEGMENTAL LEVEL: A TRANSCRIPTION-BASED APPROACH

A.P.A. Broeders* & W.H. Vieregge**

*National Forensic Science Laboratory, Rijswijk, Netherlands
**Dept. of Language & Speech, University of Nijmegen, Netherlands

ABSTRACT

This paper discusses a set of procedures which may be used to examine intraspeaker variation on the segmental level. The primary tool employed for this purpose is the consensus transcription. A variation index is proposed which captures the amount of intraspeaker variation around the modal realization of each variable. The procedures described should provide a principled approach to the investigation of intraspeaker variation with special relevance to the subject of speaker identification.

1. INTRODUCTION

While it is generally recognized that intraspeaker variation poses a major problem in speaker identification, comparatively little is known about the way in which this type of variation manifests itself in the speech of individual speakers. It is not clear, for example, whether speakers differ consistently in the amount and nature of intraspeaker variation associated with their speech. In recent years, there has been a marked increase in the number of studies dealing with inter- and intraspeaker variation, many of them undertaken with the prime object of answering questions in the field of speech technology. In spite of the current interest in speaker characteristics, there is still a remarkable scarcity of data at even the most basic level about the actual extent of variability in the speech of individual speakers. The present study seeks to develop a systematic approach to this question. However, unlike many other studies in this field, ours is not inspired by issues arising from speech technology and may therefore be of only marginal interest to it. We are aiming to devise an approach which is primarily relevant to auditory speaker identification. The primary tool employed for this purpose is the consensus transcription. Presented below are the preliminary results of this approach.

2. PURPOSE OF THE STUDY

Our main objective is to gain a better understanding of the magnitude and nature of intraspeaker variation through the use of a consensus transcription. Some of the questions we would like to answer are: Do some speakers consistently exhibit more variation than others?; Is it possible to express speaker variation in quantitative terms, and if so, how much material is required to arrive at a reliable index of intraspeaker variation?; Are some variables more consistent than others?; Is variation constant against time?

In order to investigate these questions non-contemporary speech samples were collected from 6 speakers of Dutch and subsequently transcribed according to the principles outlined below.

3. CONSENSUS TRANSCRIPTION

The concept of the consensus transcription is not new. Shriberg et al. [2] recommend it as a procedure which can be used to eliminate errors due to inattention and other shortcomings of the transcriber. They found that, of the corrections made by transcribers in a consensus transcription, 90% of those in vowel segments and 80% of those in consonant segments were considered by the transcribers to be due to inattention on their part during the original transcription process. Also, Ting et al. [3] have shown that within a group of transcribers mutual corrections lead to greater agreement between transcribers.

In the present instance, all speech samples were first transcribed by pairs of Language & Speech Pathology students of the University of Nijmegen - all of them qualified speech therapists - as part

Speaker	<3>	Svara.	Voicing	Elision	<ts>
MJ	3 (.39),3	0 (.79),1	+ (.50),2	0 (.67),2	ts (1.00)
JK	˚3(.78),2	ə (.96),1	+ (.78),2	0 (1.00),0	ts (1.00)
MK	ʃ˚(.94),1	0 (.62),1	+/v(.39),1	0 (.79),2	ts (1.00)
NO	˚3(.33),3	ə (.79),1	- (.95),1	0 (1.00),0	ts (.97),1
NR	ʃ˚(.83),2	ə (.88),1	- (.39),2	0 (1.00),0	ts (.97),1
WS	˚ʃ(.33),4	0 (.58),1	v/-(.39),1	0 (1.00),0	s (.90),2

The same statistics were determined for the various contexts of the variables 1, 2 and 6. They are omitted here for reasons of space.

The descriptive statistics presented above give a first indication of the various degrees of intraspeaker variability encountered in the material produced by the six speakers. They will make it possible to examine any changes in the realization of the variables with time. More specifically, we will be able to determine whether the modal realization changes or remains constant in both qualitative and quantitative terms. What is less satisfactory about the format used so far is the amount of information it contains about the non-modal realizations. It tells us how many realizations there are in addition to the mode and what their combined relative frequency is but it would be more interesting to know whether they are very similar to the mode in qualitative terms or very different. In other words, we would like to be able to develop a variation index which can capture the degree of similarity between the modal and non-modal realizations. The solution proposed here is one based on the use of a distance matrix as developed by Vieregge & Cucchiarini [4]. A weighted variation index Vw can be calculated by means of the following formula:

$$Vw = \sum_{i=1}^{N} Ri \,.\, di$$

Here, Ri stands for the relative frequency of the various non- modal realizations and di for the articulatory distance between a realization Ri and the Mode, calculated on the basis of the number of articulatory features in terms of which the two realizations differ. The value of the index is arrived at by summing the products of the relative frequency of each non-modal realization and its distance measure. It will be clear that the weighed variation index Vw represents a measure of the articulatory variation around the mode which is superior to the gross variation index obtained by summing the relative frequencies of the non-modal realizations because it takes account of the articulatory difference between the mode and the non-modal realizations.

7. CALCULATION OF THE VARIATION INDEX

We will illustrate the calculation of the variation index for one of our variables, <3>. Between them, the 6 speakers used 10 different realizations of this variable. The following matrix was used to calculate the differences:

	ʃ+	3+	3˚+	ʃ	3	3˚	z˚	z	ʃ−	ʃ˚ˇ
r1	ʃ+	1.0	0.5	0.5	1.5	1	2	2.5	1	0.5
r2	3+		0.5	1.5	0.5	1	2	1.5	2	0.5
r3	˚3+			1	1	0.5	1.5	3	1.5	0
r4	ʃ˚+				1	0.5	2.5	3	0.5	1
r5	3					0.5	2.5	2	1.5	1
r6	3˚						2	2.5	1	0.5
r7	z˚							0.5	3	2.5
r8	z								3.5	2
r9	ʃ−									1.5
r10	ʃ˚ˇ									

of the final project of a 120-hour course in phonetic transcription taught by the second author. They were instructed to produce a consensus transcription in accordance with the IPA conventions [1] which, they were told, would later be assessed by their teacher. The final version of the consensus transcription, which forms the basis of the present study, was produced by the two authors. After several tuning sessions, during which a number of minor notational problems were ironed out and maximum uniformity in transcriptional practice was achieved, the authors worked through the student-made transcriptions on an individual basis. However, apparent inconsistencies in the author versions were carefully re-examined to produce the ultimate consensus transcription used for this study.

4. COLLECTION OF MATERIALS

The speech samples were produced by 6 educated speakers of standard Dutch, all employed by the University of Nijmegen and living in the Nijmegen area, though originally hailing from various parts of the country. The amount of regional accent in their speech varied from mild to reasonably strong. There were three women and three men, their ages ranging from 25 to 50. The six speakers read three texts on each of three days, with a one- week interval. On each day, the three texts were read three times in succession at three points in time, i.e. at 9am, 1pm and 5pm, giving a total of 9 readings per speaker per day, and a grand total of 27 readings for each speaker for the three days. Although the texts were different, they were identical in terms of the variables under investigation, so that in effect 27 tokens of each instance of all variables are available for analysis. However, the preliminary results presented below are based on a subset of 6 non-contemporary readings from the total of 27 readings.

5. VARIABLES INVESTIGATED

Nine segmental variables were investigated. They were selected on the basis of their expected variability in Dutch. They are (n = the number of tokens per reading):

1. <r>, in four contexts, viz.:
   r1: C - n=3
   r2: - (C) # n=6
   r3: # - V n=3
   r4: V - V n=4
2. <x>, in two contexts, viz.:
   x1: - r n=2
   x2: # - V n=3
3. <z> n=5
4. <v> n=5
5. <ʒ> n=3
6. Svarabhakti, in two contexts, viz.:
   S1: l - n=3
   S2: r - n=1
7. Assimilation of voice before /b/ and /d/ n=3
8. Elision of /n/ after schwa n=4
9. <ts>, as in Dutch *politie* (English *police*) n=5

6. DESCRIPTIVE STATISTICS

In order to arrive at a first overall measure of the degree of intraspeaker variation, the following statistics were determined per variable and per speaker over the six non-contemporary readings:
1. the Mode, (M), i.e. the most common realization of the variable;
2. the relative frequency of the mode, (fM);
3. the number of realizations other than the mode, (p), any number of hapax legomena (i.e. unique realizations) being counted as 1.

They are expressed below in the format M (fM),p, or M1/M2 (fM),p, for a bimodal distribution. (The conventions used for the Voicing variable are + for voicing, - for devoicing and v for a media realization.)

Speaker	<r>	<x>	<z>	<v>
MJ	ɣ̊ (.57),6	x (.73),3	z̥ (.50),3	f/v (.33),2
JK	ʁ̊ (.36),6	ɣ (.57),4	z (.93),1	v (.77),2
MK	r (.59),7	χ (.90),1	s (.93),1	f (.97),1
NO	R (.25),10	x (.60),4	z (.57),5	f (.57),2
NR	R (.39),6	x (.87),2	s (.60),4	f (.63),3
WS	ɣ̊ (.35),9	x (.57),3	z̥ (.40),3	v (.53),3

For a full discussion of the principles underlying the distance matrix the reader is referred to Vieregge & Cucchiarini [4]. Suffice it to say here that measures used to calculate the distances are based on the articulatory difference between the sounds. Note that the value 0 is assigned to the distance measure between the realizations r3 and r10, the devoiced realization of a voiced fricative and the voiced realization of its voiceless counterpart.
Speaker MJ's raw variation score for this variable is (.39),3. There were 7 instances of the modal realization r3, 4 occurrences of r1, 4 of r2 and three hapax legomena, r5, r7 and r10. The variation index is then calculated as follows:

Vw = (.222 x .5) + (.222 x .5) + (.058 x 1.5) + (.058 x 1) + (.058 x 0) = .37

It is interesting to compare this index with the combined relative frequency of the variation around the mode, which was .61. Below, the variation index Vw is given for the remaining 5 speakers, followed by the raw variation index V.

	M	Vw	V
JK	ʒ̬(.78)	.26	.22
MK	ʃ̬(.94)	.03	.06
NO	ʒ̬(.33)	.39	.67
NR	ʃ̬(.83)	.17	.17
WS	ʃ̬(.33)	.51	.63

It appears that the weighed variation index Vw can deviate quite considerably from the raw variation index, especially if the mode has a low frequency of occurrence, as in the case of speakers NO and WS. While the relative frequency of the mode is the same for these speakers, NO's weighed variation index is considerably lower, which reflects the greater similarity to the mode of NO's non-modal realizations.

8. CONCLUSION

As observed in the introduction, the results presented above are based on a small portion of the available data. The emphasis here has been on some of the procedures used to describe intraspeaker variation in a systematic fashion. The consensus transcription is proposed as the most suitable format for the initial analysis of the speech samples collected. The use of a distance matrix based on articulatory differences between realizations affords a principled approach to a further, quantitative analysis of the variation encountered in the material. Major problems remain to be resolved before a meaningful comparison is possible of the readings produced at different times. It is this comparison which should provide answers to the central question of the consistency of intraspeaker variation patterns.

REFERENCES

[1] ROACH, P.J. (1989), "Report on the 1989 Kiel Convention", *Journal of the International Phonetic Association, 19*, 67-80.
[2] SHRIBERG, L.D. et al. (1984), "A Procedure for Phonetic Transcription by Consensus: A Research Note", *Journal of Speech and Hearing Research, 27*, 456-465.
[3] TING, A. et al. (1970), "Phonetic Transcription: A Study of Transcriber Variation", *Report*, Wisconsin Research and Development Center, Madison.
[4] VIEREGGE, W.H. & C. CUCCHIARINI (1988), "Evaluating the Transcription Process", *Proceedings of the 7th FASE Symposium Speech 88*, Edinburgh, 73-80.

PHONÉTISATION AUTOMATIQUE : ÉVALUATION ET ENRICHISSEMENT D'UNE GRAMMAIRE DE RÈGLES POUR LE FRANÇAIS

R. Belrhali, L. Libert, L.J. Boë

Institut de la Communication Parlée, URA CNRS n° 368
Grenoble, France

ABSTRACT

Our project was the establishment of a grammar for the automatic phoneticization of French. By constituting a lexicon of 60.000 words and systematically examining their transcriptions, we formulated a large body of new rules, which were added to a pre-existing base set, making a total of 900 rules. The resulting system gives a correct phoneticization of 99.75% of the base lexicon. We here present the analysis method used on this large lexicon, as well as a selection of the rules derived.

1. INTRODUCTION

La phonétisation du français peut être décrite essentiellement par règles de correspondance entre graphèmes et phonèmes. Cette correspondance se décrit par la contrainte de contexte sur la chaîne graphémique et va même jusqu'à la correspondance lexicale. Les phénomènes de phonétisation tiennent comptent de niveaux linguistiques supérieurs au mot que sont : la valeur catégorielle (souvent [suvɑ̃], chantent [ʃɑ̃t]), la fonction syntaxique (les portions [pɔʀsjɔ̃], nous portions [pɔʀtjɔ̃]), la structure syntaxique : les liaisons, (un savant [Ø] aveugle (nom - adjectif), un savant [t] aveugle (adjectif -nom)), la valeur sémantique (fils [fis], fils [fil]).

Ce travail concerne le niveau lexical de la phonétisation du français.

Avec la mise en place de bases de données faciles d'accès, de manipulation et de reconfiguration, il est maintenant possible d'élaborer, tester et améliorer les formalisations possibles. Notre effort s'est concentré sur les relations entre codes orthographiques et de vastes corpus de notations phonétiques.

2. PRÉSENTATION DU LOGICIEL TOPH

L'outil *TOPH* (Transcription Orthographique-Phonétique) est un phonétiseur multilangue qui propose une syntaxe pour décrire des grammaires de phonétisation. Ce transducteur fonctionne sur texte libre. Il permet de réécrire une chaîne d'entrée graphémique en une chaîne de sortie phonémique. Les avantages de TOPH par rapport à d'autres logiciels (cf. [3], [4], [6], [8]) sont certains. Nous pouvons mentionner sa facilité d'utilisation (traces d'application, statistiques) ainsi que la formalisation de ses règles transparente à l'utilisateur, permettant des modifications aisées. L'expert formalise son raisonnement sous la forme d'une grammaire déterministe de règles de réécriture contextuelles.

A chaque classe de règles il introduit un ordre local défini par l'ordre d'écriture des règles. La grammaire comprend :

1°- des ensembles prédéfinis de caractères orthographiques décrivant des phénomènes de nature très différente :

- *ensembles linguistiques* : "Consonnes non nasales" = (b, c, ç, d, f, g, h, j,k, l, p, q, r, s, t, v, w, x, z)
- *ensembles d'exceptions* : "Exception : fin en g " = (barlong, bastaing, basting, bourg, oing, seing, dugong, écang, étang, hareng, harfang, joug, kaoliang, long, pacfung, parfaing, rang, sampang, sanderling, sang, shampoing, shampooing, trépang, tripang)

2°- des *commentaires* pouvant être une chaîne quelconque bornée par '!' et insérée dans n'importe quelle portion de la grammaire.

3°- des *règles* partitionnées en classes ; la classe d'une règle étant déterminée par le premier caractère de la chaîne à transcrire.

3. MÉTHODOLOGIE

Afin d'enrichir la grammaire de phonétisation existante, un lexique de grande taille est nécessaire. Nous avons donc, dans un premier temps, constitué une *base de données* de 60 000 mots implantée sur *Macintosh*. L'environnement *Hypercard* et le langage *Hypertalk* ont rendu possible la mise au point de programmes de recherche de chaînes orthographiques de longueur quelconque. Elles ont été recherchées dans trois positions : *initiale, interne, finale*. A partir des listes obtenues nous avons systématiquement relevé la transcription phonétique de la chaîne étudiée en prenant comme référence de prononciation le *Petit Robert 1*. Nous avons ensuite vérifié l'existence de la (ou des) règle(s) correspondante(s) à la (ou aux) transcription(s) phonétique(s) de la chaîne de caractères étudiée. Dans le cas contraire, nous avons écrit de nouvelles règles.

Illustration de la méthode de travail par un exemple : la classe du 'b'.

Nous avons obtenu 2394 mots commencant par 'b-', 7210 mots contenant au moins un '-b-' en position interne et 32 mots se terminant par '-b'. Après le relevé de la prononciation du graphème 'b' dans tous les mots et dans toutes les positions nous avons établi les règles suivantes :

(les caractères syntaxiques sont notés en gras)

- (radou, lom) +b+ ("#", s) = [ ]

Cette règle concerne radoub et les mots se terminant par 'lomb' comme plomb, coulomb, surplomb, aplomb, dont la réalisation du '-b' en position finale est muette (ces mots peuvent être suivis d'un 's', marque du pluriel).

- +b+ (s, t) = [p]

Cette règle concerne tous les mots contenant la suite de caractères 'bs' ou 'bt' et dont le 'b' se réalise [p] ; il s'agit ici d'un cas d'assimilation régressive.

- (sub) +b+ (sidence, sidiaire, sist) = [b]
- (lam) +b+ (swool) = [b]

Ces deux dernières règles sont des exceptions à la précédente.

- +b+ (c, k) = [p]

Cette règle concerne tous les mots contenant la suite de caractères 'bc' ou 'bk' dont le 'b' se réalise [p]. Nous avons ici un autre cas d'assimilation régressive. La seule exception à cette règle est la suivante :

- (su) +b+ (carpatique) = [b]
- +bb+ = [b]

Toutes les géminées de la classe du 'b' obéissent à une règle unique.

- (“#”) +b+ (“#”) = [be]
Cette règle est uniquement applicable à la lettre de l'alphabet.
- +b+ = [b]
Il s'agit de la règle la plus générale.

Classement suivant l'ordre d'application des règles :
(radou, lom) +b+ (“#”, s) = [ ]
(lam) +b+ (swool) = [b]
(sub) +b+ (sidence, sidiaire, sist) = [b]
+b+ (s, t) = [p]
(su) +b+ (carpatique) = [b]
+b+ (c, k) = [p]
+bb+ = [b]
(“#”) +b+ (“#”) = [be]
+b+ = [b]

4. RÉSULTATS

La grammaire de base [1] contenait 200 règles et 12 ensembles d'exceptions. Actuellement *900 règles* et *16 ensembles d'exceptions* (décrivant 1 000 mots) permettent de phonétiser automatiquement les 60 000 mots de notre base de données avec un taux de réussite de *99,75%* (problème de polyphonie des mots du type 'plus' pouvant se prononcer [plys] ou [ply]). La langue, matériau vivant, est en constante évolution d'où la nécessité d'une *réactualisation systématique* de notre base de données et de la grammaire.

5. CONCLUSION

Au-delà des applications évidentes en synthèse et reconnaissance de la parole, le passage du niveau orthographique au niveau phonétique renvoit à des problèmes linguistiques fondamentaux et constitue un champ de validation privilégié des formalisations linguistiques et phonétiques. Le développement des Industries de la Langue constitue à la fois une stimulation et une possibilité directe d'application.

6. RÉFÉRENCES

[1] AUBERGE V. (1985)
Contribution à la phonétisation automatique des langues alphabétiques : le langage "TOPH". Rapport de DEA, CRISS, Département d'Informatique et Mathématiques appliquées aux Sciences Sociales, Université des Sciences Sociales de Grenoble.

[2] CATACH N. (1984)
La phonétisation automatique du français. Edition du CNRS., Paris.

[3] DIVAY M. & GUYOMARD M. (1979)
Le compilateur de règles de réécriture TOP et son utilisation à la transcription du français en vue de la synthèse. *10èmes J.E.P.-G.A.L.F.*, Grenoble, 202-211.

[4] FERVERS H., LE ROUX J.,& MICLET L. (1976)
Programme de transcription orthographique-phonémique en langue française. ENST. Paris.

[5] GACK V.G. (1976)
L'orthographe du français. Edition Selac.

[6] LETY M. (1980)
Transcription orthographique-phonétique : un système interpréteur. Thèse de 3ème Cycle. Université Scientifique et Médicale de Grenoble.

[7] PETIT ROBERT 1 (1990)

[8] PROUTS B. (1980)
Contribution à la synthèse à partir du texte ; transcription graphème-phonème en temps réel sur micrprocesseur. Thèse de 3ème Cycle. Université Paris -Sud-Centre d'Orsay.

MAVL/VOT. De l'utilisation de la notion de moment d'apparition des vibrations laryngiennes pour la description phonetique

Jean-Pierre Goudaillier

Laboratoire de phonétique, U.F.R. de Linguistique,
Université René Descartes, Paris, France

ABSTRACT

Thanks to examples in French the aim of this paper is to show that from a typological point of view the concept of MAVL is more operative than VOT in what concerns stop consonants.

Depuis plus de 25 ans déjà, la notion de V.O.T. (Voice Onset Time) telle qu'elle a été proposée dans un premier temps par LISKER et ABRAMSON [9] et ultérieurement par KLATT [8] est utilisée tant d'un point de vue phonétique que phonologique (cf., entre autres, l'emploi qui en est fait dès 1968 dans *S.P.E.*), voire même à des fins typologiques [7]. Au-delà de son utilité même [1], ce concept trouve cependant ses limites, lorsqu'il s'agit de décrire l'ensemble des réalisations possibles pour les articulations de type occlusif. La distinction entre *VOT+* (positif) et *VOT-* (négatif), même si elle peut être "ajustée" en *short voicing lead* et *long voicing lead* d'une part et en *short voicing lag* et *long voicing lag* d'autre part, ne peut pas, à mon sentiment, rendre compte des divers cas de figure que l'on peut rencontrer, ne serait-ce que pour les occlusives /p/, /t/, /k/ et /b/, /d/, /g/ en français, langue qui servira ici d'exemple. Une description complète de ces consonnes nécessite de tenir compte des cas de dévoisement partiel et/ou total pouvant affecter les occlusives phonologiquement 'sonores', ceci tout aussi bien au niveau phonétique que phonologique [2][7]. A cet effet, non seulement le concept de V.O.T. doit être utilisé mais aussi celui de M.A.V.L. (Moment d'Apparition des Vibrations Laryngiennes) [3]. Dans une perspective typologique, le concept de M.A.V.L. se révèle être plus performant que celui de V.O.T., ce que montrent les exemples présentés ici-même.

Comment déterminer les moments d'apparition des vibrations laryngiennes d'une consonne occlusive ? Une illustration est fournie par les Planches 1 et 2 qui comportent les tracés du phonogramme (ligne M) et de l'électroglottogramme (ligne EGG) de 6 séquences prononcées par des enfants francophones. Le [b] de [œ̃balɛ·] (Figure 1; Planche 1) est entièrement voisée; ceci veut dire que sa phase d'occlusion et celle de relâchement sont toutes les deux accompagnées de vibrations des cordes vocales (Dans un tel cas la phase de relâchement, qui est constituée d'une explosion très brève non accompagnée d'un V.O.T. positif, est difficilement repérable sur ce type de tracés : l'explosion se confond alors avec une vibration des cordes vocales, ce tant sur le phonogramme que sur l'électroglottogramme. Toutefois, le passage entre la consonne et la voyelle est visible, étant donné l'augmentation d'amplitude notée au début de la voyelle). La durée de l'ensemble est de *85 ms* et aucune interruption des vibrations laryngiennes n'a lieu lors du passage de la consonne [b] à la voyelle [a]. C'est un *type 1* de M.A.V.L.. Le [g(h)] de [œ̃g(h)at(h)ǫ] (Figure 2; Planche 1) et la consonne dentale [d̥(h)] de [œ̃d̥(h)e] (Figure 3; Planche 1), quant à eux, ne comportent pas de voisement pendant leurs phases de relâchement : l'explosion n'est pas voisée et le bruit de friction suivant cette dernière ne l'est pas non plus. Un V.O.T. positif de *+15ms* est observé dans les deux cas. Pour [g(h)], qui correspond à un *type 2* de M.A.V.L., l'occlusion est entièrement sonore. Elle dure *45ms*. Pour l'autre consonne (Figure 3) l'occlusion n'est voisée que partiellement : seuls *45ms* (soit *75%*) de celle-ci, qui dure au total *60ms*, sont voisés; la fin de cette occlusion est donc sourde pendant *15ms*. Il s'agit d'un *type 3* de M.A.V.L.. L'occlusion du [b̥(h)] de [b̥(h)ɑ̃ʒl] (Figure 4; Planche 2) est entièrement sourde. Cette articulation compte par ailleurs un V.O.T. positif de *+5ms*. On est en présence ici d'un *type 4* de M.A.V.L..

Pour ce qui est des consonnes phonologiquement 'sonores', les quatre types de M.A.V.L. peuvent être récapitulés comme suit : M.A.V.L. 1 : occlusion voisée + relâchement voisé; M.A.V.L. 2 : occlusion voisée + relâchement non voisé; M.A.V.L. 3 : occlusion mi-sonore + relâchement non voisé; M.A.V.L. 4 : occlusion non voisée + relâchement non voisé (pour les types 2,3 et 4 le (h) note simplement l'absence de tout voisement pendant la phase de relâchement et non un quelconque souffle ou "aspiration").

Pour le [p(h)] de [œ̃p(h)anje·] (Figure 5; Planche 2) et le [p(h)] de [p(h)o·s] (Figure 6; Planche 2) les durées d'occlusion sont respectivement de *80ms* et *115ms*. Si le V.O.T. positif est inférieur à 40ms, la consonne a un *type 5* de M.A.V.L.. Ceci est le cas du [p(h)], puisque son V.O.T. ne dure que *+20ms*; si le V.O.T. positif est superieur à 40 ms, on est en présence d'un *type 6* (ceci est le cas du [p(h)] avec son V.O.T. de *+55ms*).

Les deux types de M.A.V.L. attribués aux consonnes phonologiquement 'sourdes' peuvent être récapitulés comme suit : M.A.V.L. 5 : occlusion non voisée + relâchement non voisé (avec V.O.T. inférieur à 40ms); M.A.V.L. 6 : occlusion non voisée + relâchement non voisé (avec V.O.T. supérieur à 40ms).

Les différents types de M.A.V.L. peuvent par ailleurs être schématisés ainsi qu'il est indiqué à la Planche 3. Si l'on veut analyser les phénomènes d'assibilation notés dans certaines variétés du français (franco-canadiennes plus particulièrement; Québec, Ontario, Nouveau-Brunswick, etc.), il convient d'inclure à cette schématisation des types supplémentaires [3][6].

Dans une perspective typologique, sur quels points l'approche en termes de M.A.V.L. est-elle plus performante que celle basée sur le V.O.T. ? Si l'on revient sur les types 2 et 3, on peut aisément constater que ceux-ci ne peuvent donner lieu à aucune mesure de V.O.T.. Pour le type 2 il faudrait tenir à la fois compte d'un V.O.T. positif de +15ms et d'un autre, quant à lui négatif, de -45ms. Or, l'idée -même de Voice Onset Time ne permet pas d'avoir des unités phonétiques comportant à la fois un V.O.T. négatif et un V.O.T. positif. C'est l'un ou l'autre. Il est donc impossible de traiter de telles unités, si ce n'est en utilisant le concept de M.A.V.L.. Il en est de même pour l'exemple de M.A.V.L. de type 3, d'autant plus qu'une phase sans vibrations laryngées de 15ms de durée survient entre ce qu'il conviendrait d'attribuer à un V.O.T. négatif de 45ms s'arrêtant à 15ms de la fin de l'occlusion d'une part et d'autre part la fin de l'occlusion elle-même, tout ceci étant accompagné d'un V.O.T. positif de 15ms. Il est donc impossible d'analyser en fonction du Voice Onset Time le [g(h)]

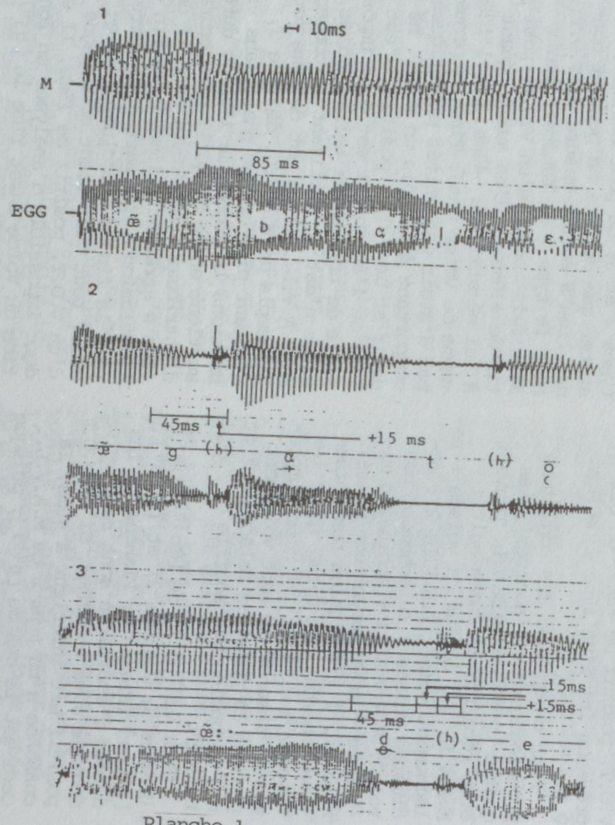


Planche 1

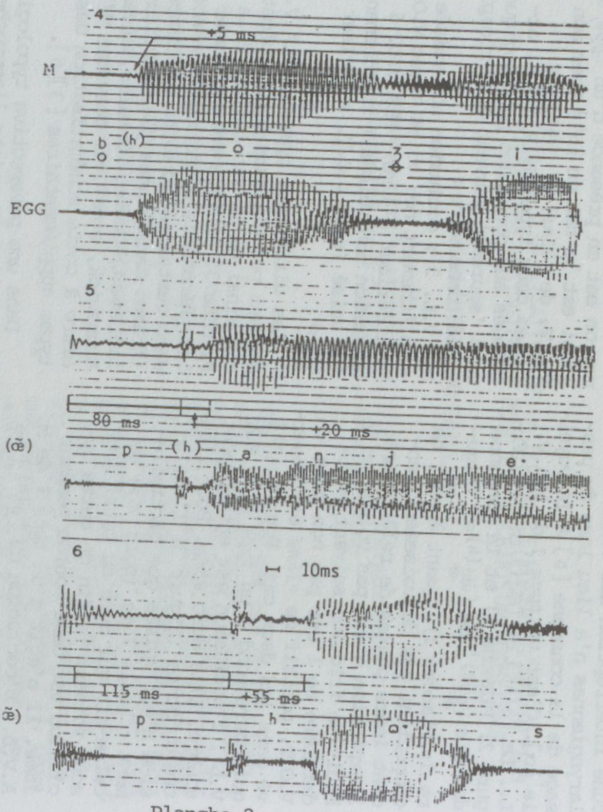


Planche 2

et le [g̊(h)] (Figures 2 et 3); le V.O.T. ne peut être compris que de manière *binaire* : négatif/positif. C'est ceci même qui le rend inopératoire pour l'analyse d'occlusives partiellement désonorisées, que ce soit pendant leur phase d'occlusion ou pendant leurs phases d'occlusion et de relâchement. Sans le concept de M.A.V.L., qui permet d'affiner la description phonétique des occlusives en individualisant, pour le français, *4* types différents, au lieu de *2*,il serait impossible de rendre compte de faits d'acquisition [5] ou de différenciations sociolinguistiques, plus particulièrement d'ordre sexuel [4].

REFERENCES

[1] GOUDAILLIER,J.-P. (1981),"Exemple de traitement de l'opposition de "sonorité"...",*12è JEP* (Montréal), 377-391.

[2] GOUDAILLIER,J.-P. (1986),"Eléments de phonologie...", *Langues et Linguistique*, 12, 131-180.

[3] GOUDAILLIER,J.-P. (1986),"Voisement et assibilation...", *12è ICA* (Toronto), Section A3-7.

[4] GOUDAILLIER,J.-P. (1988),"Sonorité des occlusives et différenciation sexuelle",*BSL*,83/1,323-330.

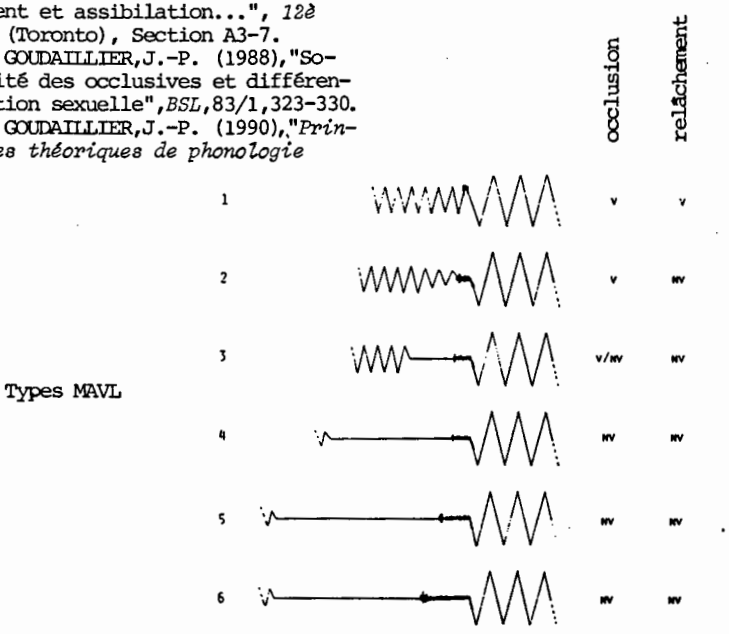
[5] GOUDAILLIER,J.-P. (1990),"*Principes théoriques de phonologie fonctionnelle expérimentale (P.F.E.) Théorie, illustrations et application aux occlusives d'enfants francophones français et québécois*", Hamburg : Buske Verlag.

[6] GOUDAILLIER,J.-P.,BENTO,M. (1990) "M.A.V.L./V.O.T. ? Proposition pour un classement phonétique en termes de Moments d'Apparition des Vibrations Laryngiennes des occlusives françaises et québécoises",*18è JEP* (Montréal), 64-68.

[7] KEATING, P.E. (1984), "Phonetic and phonological representation of stop consonant voicing", *Language*, 60/2, 283-319.

[8] KLATT, D.H. (1975), "Voice Onset Time, frication, and aspiration in word-initial consonant clusters", *Journal of Speech and Hearing Research*, 18, 277-290.

[9] LISKER L., ABRAMSON A.S. (1964), "A cross-language study of voicing in initial stops : acoustical measurements", *Word*, 20, 384-422.



Types MAVL

Planche 3

THREE TYPES OF PROSODIC CORRELATIONS IN SOUTH GERMAN DIALECTS

L. NAIDICH

INSTITUTE OF LINGUISTICS, LENINGRAD, USSR

ABSTRACTS

The paper deals with three types of prosodic correlationes in South German dialects: correlation of gemination in High Alemannic, opposition of syllables with different place of quantitative syllable peak in North and Central Bavarian, correlation of syllable cut in Low Alemannic and High Franconian. This prosodic systems are considered as different stages of the evolution of Germanic quantity.

There are three major types of prosodic systems in South German dialects. High (South) Alemannic dialects - a group which most of Swiss German dialects belong to - are known to preserve a rather archaic quantitative order and syllabic structure. In the consonant system of these dialects two phoneme types are opposed to each other: the weak consonants - lenes - and the strong - fortes -, this opposition concerning all the consonants: stops, fricatives, resonants. Geminates, double consonants typical of Swiss German dialects, occuring in the intervocalic position or sometimes between resonant and vowel, are phonologically identified with fortes. This can be demonstrated by means of the rule formulated by L.Zinder [9]: sounds standing in complimentary distribution and alternating in the same morpheme constitute the same phoneme; e.g. in our case šit 'Holzscheit'- -šit-tə ' Holz spalten'. The occurence of long vowels before fortes (long) consonants or consonant clusters ([šāff] 'Schaf', [lȳt] 'Leute', [dōrff] 'Dorf') seems to testify "free quantity". This syllable shape (overlong syllable) is especially frequent in some South Swiss German dialects with "Auslautverhärtung", i.e. the neutralization of fortes-lenes opposition in favour of fortes at the word end. As for another "unusal" (in terms of modern West Germanic languages) syllable type in monosyllabics - short vowel before lenis, - they are frequent, when stressed, in few dialects (Gadmental, Haslital, Obwalden in the South and some dialects in the North-East of German speaking Switzerland)[2].In most Swiss German dialects, however, they are possible only in some word kinds - imperative verb forms, expressive words: [šlag] 'schlage!', [red] 'rede!' [kxog] swear-word [5] This syllable shape is quite seldom in case of a resonant in the word end. Whereas in monosyllabics lenis is in most cases preceded by a long vowel - because of the so-called High Alemannic lengthening (or Leichtschlussdehnung), i.e. lengthening of short vowel before lenis, short vowels in the open syllable of di- and polysyllabic words are quite usual, [šabə] (ša-bə) 'schaben', [redə] (re-də) 'reden'. While the syllabic division and shape are independent on the vowel length, the type of consonant plays here a decisive role. In the intervocalic position the syllabic boundary always occurs within fortis which is phonetically geminated whereas the lenis starts the next syllable: sit-tə 'Seite' -si-də 'Seide'. Thus a specific prosodic order in these dialects is present which could be called correlation of gemination.

A similar prosodic system is observed in South Bavarian dialects, whereas in Central and North Bavarian the syllable structure obeys the rule: long vowel+lenis (short) consonant-short vowel+fortis (long) consonant (Pfalz's law), these two syllable types often alternating in morphological paradigms, e.g.the opposition of singular and plural forms of substantives: [grīf]-[griff] 'Griff',[fīš]-[fišš]'Fisch', [šdōg]-[šdek]'Stock'. This correlation has some peculiarities differing it from similar prosodic quantitative opposition in other Germanic languages. The alternation of fortes and lenes depending on the length of preceding vowels takes place in consonant clusters as well: [gōsd]-[gesst] 'Gast'- -'Gäste'; long as opposed to short can be diphthongs and affricates too:[hāud]-[hait] 'Haut' - 'Häute', [khōbf]- -[khepf] 'Kopf'-'Köpfe'; resonants do not take part in this correlation being always weak: [hūnd] - [hunt] 'Hund' - 'Hunde'. We consider this correlation as a prosodic one, as an opposition of syllable types differing in the place of syllabic quantitative peak, although the character of this opposition remains disputable [3].

In Low Alemannic and High Franconian a third prosodic order is represented - the syllable cut correlation or the opposition of close and loose contact between a stressed vowel and the following consonant, typical of many West Germanic languages and dialects, e.g. modern standard English, German and Dutch. The prosodic character of this correlation becomes apparent in syllabication depending on the vowel length which in turn is the phonetic sign of contact: [līdə] (lī-də), -[lidə] 'leiden' - 'lauten' (Low Alemannic - Alsatian dialect).

Modern dialects reflect different stages of prosodics development. They show general trends of syllable structure evolution common to all Germanic languages, but also some specific High German features

An important consequence of the Second Sound Shift was the elimination of the opposition voiceless-voiced in the consonant system of South German dialects. The reduction of the opposition of three consonant raws : lenes-fortes-geminates to that of only two: lenes-fortes with geminates as positional variants took place - as many scholars believe -

already in the Old High German [6]. Thus the consonant system here was based on the opposition of lenes derived from Germanic voiced and fortes derived from Germanic voiceless stops which were shifted according to the Second Sound Shift, Germanic geminates and long consonants resulted from the West Germanic consonant lengthening. The shifted fricatives were long, strong and in the intervocalic position geminated. Germanic p>ff merged with Germanic ff, k>xx merged with Germanic xx, t>ʒʒ. Thus the group of fortes was enlarged, the opposition of fortes-lenes became universal for the whole system. Only in the dental raw the triple opposition d-t-tt and also s-ss-ʒʒ was preserved for a longer time. After the coincidence of ʒ, ʒʒ (<t) with s, ss and t (<d) with tt (<dd, þþ) the consonant system was simplified, the opposition of fortes-lenes intimately connected with syllabication and syllable shape became strengthened This consonant system is preserved to-day in Swiss German dialects.

Further evolution of syllabic structure was directed toward the interrelation of vowel and consonant quantity inside the word. According to the assumption of E.Kranzmayer a trend to the equal quantity of all the words got developed [4]. As the Swiss German dialects indicate, the first step of this development could be the vowel lengthening in monosyllabics before lenes - first of all of open vowels before resonants. This statement conradicts the widespread concept according to which the lengthening in monosyllabics occured by analogy with that in open syllable, Many new monosyllabic words ending in fortes resulted from the apocope which took place in Central and North Bavarian dialects: [šiff]<[šiffə]. They contrast with old monosyllabics built according to the pattern: long vowel+lenis. As a result of all these processes a quantitative- prosodic order referred to above as the correlation of the syllable peak place was formed. There are reasons to suppose that a prosodic correlation like this (sometimes called isochrony) always preceded the syllable cut correlation in the history of West Germanic languages [8].

The next step was the elimination of geminates which became phonologically redundant. In Central German dialect area including also some South German dialects - High Franconian, Low Alemannic - these processes were connected with the merge of fortes and lenes - consonant weakening[7].Because of very few oppositions in the consonant system of these dialects the syllable cut (contact) correlation became an important means of differentiation: liuten > lit-ten (with delabialization iu [y] > i and vowel shortening)> modern Alsatian [lidə] with close contact 'lȁuten' - līdan> modern Alsatian [līdə] with loose contact 'leiden' [1].

REFERENCES

[1] FOURQUET, J (1964), "*Zur Deutung der Isophonen der Quantität*", "Phonetica", *v.11*, 3-4,155-163.

[2] GABRIEL, E. (1960), "*Die Entwicklung der althochdeutschen Vokalquantitäten in den oberdeutschen Mundarten*" - "*Studien zur osterreichisch-bairischen Dialektkunde*", 5

[3] HINDERLING, R. (1980), "*Lenis und Fortis im Bairischen*",*ZDL, H.1*,25-51.

[4] KRANZMAYER, E. (1956), "*Historische Lautgeographie des gesamtbairischen Dialektraumes*", Wien: Österreichische Akademie der Wissenschaften.

[5] NAIDICH, L. (1979), "*Prosodicheskiye yavleniya v shveytsarsko-nemetskix dialektax*", In:"*Issledovaniya v oblasti sravnitelnoy aktsentologii indoyevropeyskix jazikov*",Leningrad: Nauka, 238-250.

[6] REIFFENSTEIN, I. (1980), "*Geminaten und Fortes im Althochdeutschen*","*Münchener Studien zur Sprachwissenschaft*", *H.18*,61-77.

[7] REIS, M.(1974), "*Lautheorie und Lautgeschichte. Untersuchungen am Beispiel der Dehnungs- und Kürzungsvorgange im Deutschen*", München, *Internationale Bibliothek für allgemeine Linguistik*",14.

[8] VALENTIN, P. (1969) ."*L'isochonie en nha. ancien*", In: "*Mélanges pour Jean Fourquet*", Paris, München: Klincksieck, 334-347.

[9] ZINDER, L.R. (1979), Obshtchaya fonetika. 2nd edition , Moskow: Vysshaya shkola.

MORPHOPHONEMICS OF WEAK-STEM FORMS IN MODERN CHALDEAN

Solomon Sara,S.J.

Georgetown University, Washington, D.C.

ABSTRACT

This paper explores the derivation of the surface forms of the weak-stem verb forms in Modern Chaldean. Weak-stem forms are those forms that include the glides /y,w/ among their triconsonantal radicals. The intercalation of vowels in the verb paradigms include /y,w/,and the labile nature of these glides, produce surface forms that are at variance with the corresponding strong-stem paradigms.

O.O INTORDUCTION

Modern-Chaldean is one of the currently spoken, and much changed, dialects of Classical Syriac/Aramaic. There are many such dialects, but the dialect that is the focus of this presentation is the dialect of /mangeši/, a town in the northern part of Iraq.

1.0 DATA AND PROCEDURES.

There is no lexicon for the Modern Chaldean dialect of /mangeši/ that will provide a list of all the weak-stem lexical items. I have depended on lists of words found in several grammars or lexica of other dialects as my data sources. The lexical items that were provided in these sources needed to be phonemically modified to the /mangeši/ sound patterns. These sources provided a representative sample of lexical items in which /w,y/ occur in verbal paradigms.

2.0 STEMS WITH /Y,W/

/w,y/ are contrastive units in this dialect as shown in the examples:

/yarixa/ 'long'	/wariδa/ 'root'
/sya?a/ 'fence'	/swa?a/ 'satisfy'
/gnaya/ 'set'	/gnawa/ 'steal'

3.0 /Y/ IN INITIAL POSITION

The terms INITIAL, MEDIAL, and FINAL positions refer to the first, second and third consonantal segments of the root respectively. The following stems illustrate /y/ in initial positions in roots:

/yrx/ 'long',	/ypy/ 'bake'
/ywl/ 'give',	/ywš/ 'dry'
/yld/ 'give birth'	/ylp/ 'learn'
/yqδ/ 'burn'	/yqr/ 'heavy'
/yrx/ 'length'	/ytw/ 'sit', etc.

In the conjugation of the verbal forms various changes take place that affect the initial /y/ of the stem. An instructive way of observing this process is to compare the strong and the weak paradigms of the verb in a conjugational framework. Only the singular paradigm is given here, since the plural suffixes do not affect the changes in the stem differently from the singular:

3.1 Strong/Weak

/d-r-s/ 'study'	**/y-l-p/ 'learn'**
/drasa/ 'to study'	/?ilapa/ 'to learn'
1s./darsɪn/'I study'	/yalpɪn/ 'I learn'
2s./darsɪt/ 'you study'	
/yalpɪt/ 'you learn'	

3s./darɪs/ 'he studies'
/yalɪp/ 'he learns'
1s./drɪsli/ 'I studied'
/lɪpli/, /?ilɪpli/ 'I learned'
2s./drɪslux/ 'you studied'
/lɪplux/, /?ilɪplux/ 'you learned'
3s./drɪslɪ/ 'he studied'
/lɪplɪ/, /?ilɪplɪ/ 'he learned'
imp.2s./dros/ 'study'
/lop/, /?ilop/ 'learn'
imp.2p./drusu/ 'study'
/lupu/, /?ilupu/ 'learn'

R1. y-Deletion.
y ---> 0 /#__C /ylɪpli/ --> /lɪpli/
R2. y/w-Syllabification .
y---> i /V__C/ /ylop/---> /ilop/
R7. ?-Insertion.
0 ---> ? /#__V /ilop/ --> /?ilop/
Both options occur and are acceptable: The glide /y/ is deleted in the initial cluster. The second option is that /y/ is converted to /i/. This necessitates the addition of a glottal stop /?/ before the vowel, since all syllables begin with a consonant, and vowels do not begin syllables in this dialect.

3.2 /Y/ IN MID-POSITION

The following stems illustrate /y/ in mid-position:

/dyš/ 'trample'	/zyd/ 'increase'
/zyp/ 'josstle'	/cyk/ 'stuff'
/cym/ 'close'	/nys/ 'bite'
/dyn/ 'judge'	/xyt/ 'sew'
/lys/ 'chew'	

Strong/Weak

/d-r-s/ 'study'	**/d-y-n/ 'judge'**
/drasa/ 'to study'	/dyana/ 'to judge'

1s. /darsɪn/ 'I study'
/denɪn/ 'I judge'
2s. /darsɪt/ 'you study'
/denɪt/ 'you judge'
3s. /darɪs/ 'he studies'
/dayɪn/ 'he judges'
1s. /drɪsli/ 'I studied'
/dɪnni/ 'I judged'

2s. /drɪslux/ 'you studied'
/dɪnnux/ 'you judged'
3s. /drɪslɪ/ 'he studied'
/dɪnnɪ/ 'he judged'
imp.s. /dros/ 'study' /don/, 'judge'
imp.p. /drusu/ 'study' /dunu/, 'judge'

R2. y/w-Syllabification.
y---> i/V__C /daynɪn/ --> /dainɪn/
R5. Tense vowel lowering.
i--->e/V__ /dainɪn/ --> /daenɪn/
R6. Vowel deletion.
V1V2 ---> V2 /daenɪn/ --> /denɪn/

3.3 /Y/ IN FINAL POSITION

The following stems illustrate /y/ in final position:

/bny/ 'build'	/gby//beg'
/jhy/ 'tired'	/dry/'put '
/ypy/ 'bake'	/hwy/ 'be'
/xzy/ 'see'	/tpy/ 'stick'
/kly/ 'stay'	/cmy/ 'put out'
/kry/ 'shorten'	

Strong/Weak

/d-r-s/ 'study'	**/k-l-y/ 'stay'**
/drasa/ 'to study'	/klaya/ 'to stay'

1s./darsɪn/ 'I study'
/kalɪn/ 'I stay'
2s./darsɪt/ 'you study'
/kalɪt/ 'you stay'
3s./darɪs/ 'he studies'
/kalɪ/ 'he stays'
1s./drɪsli/ 'I studied'
/kleli/ 'I stayed'
2s./drɪslux/ 'you studied'
/klelux/ 'you stayed'
3s./drɪslɪ/ 'he studied'
/klelɪ/ 'he stayed'
imp.s./dros/ 'study'
/kle/ f. 'stay', /kli/ m. 'stay'
imp. p. /drusu/ 'study' /klo/ 'stay'

R3. y-Syllabification before non-low vowel.
y —> i/__V /kalyɪn/ --> /kaliɪn/
R5. Tense vowel lowering.
i—e/V__ /klɪili/ --> /klɪeli/

R6. Vowel deletion.
V1V2 --> V2 /klɪeli/ --> /kleli
/kalɪn/ --> /kalɪn/

4.0 /W/ IN INITAL POSITION
The following stems illustrate /w/ in initial position:
/wjb/ 'duty' /wrq/ 'paper'
/wrd/ 'rose' /wrδ/ 'root'
No verbal form beginning with /w/ is available for conjugation.

4.1 /W/ IN MEDIAL POSITION
The following stems illustrate /w/ in medial position:
/hwy/ 'become' /zwn/ 'buy'
/xwr/ 'white' /twr/ 'break'
/?wδ/ 'do' /qwy/ 'hard'
/šwr/ 'jump'
Strong/Weak
/d-r-s/ 'study' /z-w-n/ 'buy'
/drasa/ 'to study' /zwana/ 'to buy'
1s./darsɪn/ 'I study'
/zonɪn/ 'I buy'
2s./darsɪt/ 'you study'
/zonɪt/ 'you buy'
3s./darɪs/ 'he studies'
/zawɪn/ 'he buys'
1s./drɪsli/ 'I studied'
/zwɪnni/ 'I bought'
2s./drɪslux/ 'you studied'
/zwɪnnux/ 'you bought'
3s./drɪslɪ/ 'he studied'
/zwɪnnɪ/ 'he bought'
imp. s./dros/ 'study' /zwon/ 'buy'
imp. p./drusu/ 'study' /zwunu/ 'buy'

R2. y/w-Syllabification.
w---> u/V__C /zawnɪn/--> /zaunɪn/
R5. Tense vowel lowering.
u--->o /V__ /zaunɪn/ --> /zaonɪn/
R6. Vowel deletion.
V1V2 ---> V2 /zaonɪn/ --> /zonɪn/

4.2 /W/ IN FINAL POSITION
The following stems illustrate /w/ in final position:
/ytw/ 'sit' /gnw/ 'steal'
/xrw/ 'spoil' /qrw/ 'near'
/slw/ 'cross'
Strong/Weak
/d-r-s/ 'study' /g-n-w/ 'steal'
/drasa/ 'to study' /gnawa/ 'to steal'
1s. /darsɪn/ 'I study'
/ganwɪn/ 'I steal'
2s. /darsɪt/ 'you study'
/ganwɪt/ 'you steal,
3s. /darɪs/ 'he studies'
/ganu/ 'he steals'
1s. /drɪsli/ 'I studied'
/gnuli/ 'I stole'
2s. /drɪslux/ 'you studied'
/gnulux/ 'you stole'
3s. /drɪslɪ/ 'he studied'
/gnulɪ/ 'he stole'
imp.s. /dros/ 'study'
/gnu/ 'steal'
imp.p. /drusu/ 'study'
/gnuwu/ 'steal'

R2. y/w-Syllabification.
w---> u/ V__C /ganɪw/ --> /ganɪu/
R6. Vowel deletion.
V1V2 ---> V2 /ganɪu/ --> /ganu/

5.O SUMMARAY
The /w,y/ segments change in the sequences of verb paradigms according to specific rules that are determined by the contexts in which these segments occur. There is similarity between the rules rules for the glides but no identity. There are more changes that are operative with the high front glide /y/ than the changes of the back high glide /w/. The discussion of the changes was limited to the verbal paradigm, and may be extended to other lexical paradigms with comparable expectations. There are other contexts in which these changes take place, and are under investigation.

6.O SUMMARY OF RULES:

R1. y-Deletion.
y ---> 0 /#__C
[-syll]
[-cons] [-syll]
[+high]-->[0] /#__[+cons]
[+tens]
[-back]
[-roun]

R2. y/w-Syllabification.
y/w ---> i / V__C
[-syl, -cons, +high, +tense]
||
\/
[+ syll]
/
[+syll] [-syll
[-cons] _____ [+cons]

R3. y-Syllabification before non-low vowel.
y ---> i /__V
[-syll]
[-cons] [+syll]
[+high]--> [+syl] /__[-cons]
[+tens]
[-low]
[-back]
[-roun]

R4. y-Deletion between two non-low vowels.
[-syl -cons +hi +tense -bk -rnd]
||
\/
0
/

[+syll] [+syll]
[-cons] [-cons]
[-low] ____ [-low]

R5. Tense vowel lowering.
i--->e /V__
u--->o /V__

[+syll]
[-cons] [+syll]
[+high] ---> [-high]/[-cons]__
[+tens]

R6. Vowel deletion.
V1V2 ---> V2
[+syll] [+syll]
[-cons] [-cons]
1 2 ---> 2

R7. ?-Insertion.
0 ---> ? /#__V
[-syll]
[-cons] [+syll]
0 --> [-voic] /#___ [-cons]
[-cont]

REFERENCES

[1] GARBELL, I. (1965). *The Jewish Neo-Aramaic dialect of Persian Azerbaijan: Linguistic analysis and folklore texts.* (Janua Linguarum, Series Practica, 3.), The Hague: Mouton & Co.

[2] KROTKOFF, G. 1982. *A Neo-Aramaic dialect of Kurdistan*, New Haven: American Oriental Society.

[3] MACLEAN, A. J. 1901. *A dictionary of the dialects of Vernacular Syriac*, Oxford: The Clarendon Press. reprinted 1972. Amsterdam: Philo Press.

[4] SARA, S. I. 1974. *A description of Modern Chaldean*, The Hague: Mouton & Co.

STOP ASSIBILATION IN QUEBEC FRENCH; AN ANALYSIS BY ARTICULATORY SYNTHESIS

H.J.Cedergren[1], D.Archambault[2] & G.Boulianne[1]

[1]Université du Québec à Montréal, Canada
[2]INRS-Télécommunications, Québec, Canada

ABSTRACT

This paper discusses the use of articulatory synthesis as a research tool for determining the relationship between phonetic data and phonological structure. Under the assumption that the mapping between phonological and phonetic representations is accomplished by rules of phonetic implementation we use a computational model to examine alternative accounts of the derivation of surface affricates in Quebec French. The model's behaviour is shown to parallel natural speech data.

1. INTRODUCTION

One of the defining characteristics of Quebec French is the surface reflex of coronal stops /t/,/d/ in the context of a following anterior high vowel /i/,/y/. Underlying coronal stops are realized as the assibilated affricates [t^s,d^z] in these contexts. Forms such as *dix* "ten" or *petit* "little" are typically realized as [d^zIs] and [pt^si]. These forms are usually accounted for by rules of the form

/t/ ---> [t^s]/___{i,y}
/d/ ---> [d^z]/___{i,y}

Alternative accounts of the mechanism responsible for the derivation of these surface affricates have been proposed. The first is a phonological account which assumes that the forms [t^s], [d^z] are contour segments with dual feature matrices consisting of a stop and fricative components [4]. The second is a phonetic account which assumes that the fricative component of the surface affricate is the consequence of the transition gesture between the constriction location of the apico-alveolar stop and that of the following vowel [3,8].

An articulatory synthesis system [1], which uses phonological feature matrices as input and explicitly models feature/production relations, is used to examine the explanatory value of both hypotheses. In the next sections we give a brief description of the system and present details of our computational simulation of /t/ assibilation.

2. SYNTHESIS SYSTEM

The articulatory synthesis system which has been implemented consists of three major components: first a series of modules in which different phonological, phonetic and articulatory knowledge structures are represented; an articulatory model [9]; and finally a central representation structure which controls modules interfacing and is accessible to the user.

We assume a multidimensional non-linear representation of underlying phonemic and phonetic segments, with fundamental distinguishing properties: (1) phonological features are assumed to be abstract binary classificatory features, i. e. segments may be [+/- labial]; while phonetic features are n-ary [x], i.e. [+labial] phonological segments may correspond to either [projected], [neutral] or [retracted] labial positions. (2) Phonological representation is assumed to be underspecified; that is, only contrastive feature values are marked underlyingly. Redundancy rules add essential feature values. (3) Physiological representation is in the spatial domain; therefore the phonetic/physiology passage defines a relation between discrete phonetic parameters and acoustically important area-function parameters [5].

Rules written in Delta account for internal properties and module interfacing [6]. The phonological rule-set assigns initial feature specifications and defines feature alignment relations. Phonetic rules translate abstract phonological representations into n-ary phonetic features and assign inherent duration to each segment. Rules of the physiological module account for phonetic feature/production relation by specifying corresponding motor-sensory goals and intrasegmental dynamics. The calculation of articulatory trajectories is accomplished by optimization techniques.

3. ACCOUNTING FOR /t/ ASSIBILATION

The system which we have described allows us to explicitly examine the phonology-phonetics interface question as the implementation of physiological gestures derived from a sparsely specified abstract feature representation and to test linguistic hypotheses about levels of representation. We illustrate this issue by modelling /t/ assibilation in Quebec French.

The properties of our computational model (which simulates the properties of actual articulatory systems) permit us to address two aspects of the assibilation problem: (1) do the articulators, during the transition gesture form /t/ to /i/, occupy for a critical duration a location which permits assibilation?; (2) does delayed glottal adduction following offset of closure result from speaker control or aerodynamic conditions of the post /t/ constriction?

3.1 Acoustic Data

A corpus of natural speech was gathered from a single speaker. Three types of stimuli were included in the data-set: (1) occurrences of /t/ followed by (a) a vowel which conditions assibilation, (b) vowels which do not; (2) occurrences of /t/ followed by /s/ in the same vocalic contexts; (3) occurrences of /s/ not preceded by /t/.

From these data we defined the duration for the various segments. We also found that while durations for vowels and fricatives vary under stress, durations of plosives remain constant.

3.2 Radiographic Data

X-ray tracings of /t/ when followed by /i,y/ were analyzed in order to gather data on articulatory gestures [10]. These data suggest that speakers do not aim at a particular target during assibilation, rather that the transition of the tongue is made directly from the /t/ to the /i/. This information on the motor-sensory goals was integrated into the physiological module.

3.3 Experiment

The input strings, such as, /ati/ and /atsi/ were submitted to the system. There were no specific rules provided to account for assibilation. The rule set did not include any rule inserting a fricative segment between the stop and the following vowel. Further, the rate of occlusion offset was set to be uniform for all vowels, and glottal aperture duration did not depend on consonant type.

The synthetic response produced an excrescent fricative segment between the stop and the vowel. The spectrograms of synthesized sequences replicate the major characteristics of measured sequences taken from natural speech. Figure 1 shows the spectrogram of the synthetic sequence derived from the /ati/ input string. Frication noise is found between 3 and 8kHz.

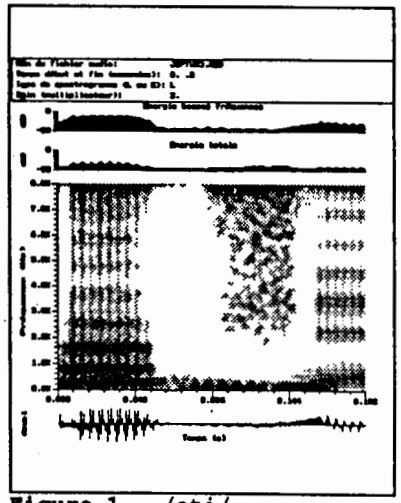


Figure 1 /ati/

The measured durations for the vowels and the frication noise correspond well to those that are stipulated in the rules (cf Table 1). The durations for /t/ are smaller than expected, because they refer only to the period of occlusion, and do not include the transition of the vowel to the onset of closure. Assibilation was produced even when the vowels /e/ and /a/ followed the /t/. However, the spectrogram reveals that the initiation of vocal cord vibrations was quicker than after high front vowels, giving frication periods inferior by 6 to 14%.

TABLE 1. Acoustic durations

Sequence	Duration (ms)			
	prec. vowel	/t/	fric.	foll. vowel
/ati/	48	49	47	37
/ati/ acc.	48	48	46	67
/atsi/	48	47	141	46
/atsi/ acc.	48	46	162	69
/aty/	48	49	46	37
/ate/	48	46	44	41
/atɛ/	48	61	41	39

4. CONCLUSION

The explicit modelling of hypotheses on the mechanism of assibilation has allowed us to evaluate their likelihood. We find that it is not necessary to program assibilation as a phonological process: it results from aerodynamic conditions that are satisfied during the course of the transition form /t/ to /i/, as a consequence of voice onset delay.

These results, however, should be interpreted in the light of traditional warnings on the use of simulation techniques. For they depend on the validity of the model and the articulatory synthesizer that are evidently an approximation of reality [2].

Further tests should be carried out with a full range of vowels. However it would appear that the computational model can be of further use to research in linguistics to test theories.

5. ACKNOWLEDGMENT

This research was supported by the Fonds de développement académique du réseau de l'Université du Québec.

6. REFERENCES

[1] ARCHAMBAULT, D., G. BOULIANNE & H.J. CEDERGREN (1990), "L'analyse de la relation langue-parole pour un système de synthèse articulatoire", *XVIIIèmes JEP, Montréal, 28-31 mai.*

[2] BROWMAN, C.P. & L. GOLDSTEIN (1990), "Targetless" Schwa: an articulatory analysis", *Haskins Lab. SR-101/102*, 194-219.

[3] CHARBONNEAU, R. & B. JACQUES, (1972), "[ts] et [dz] en canadien-français", in A. Valdman, ed., *Papers in Lingustics and Phonetics to the Memory of Pierre Delattre*, La Haye: Mouton, 77-90.

[4] DUMAS, D. (1978), *La phonologie des réductions vocaliques en français québécois*, Ph.D. diss. U. de Montréal.

[5] GAY, T. & B. LINDBLOM & J. LUBKER (1981), "Production of Bite-Block Vowels: Acoustic Equivalence by Selective Compensation", *Journ.Acous.Soc.Am.*,vol.69, no 3, 802-810. [6] HERTZ, S.R. (1988), "The Delta programming language: an integrated approach to non-linear phonology, phonetics, and speech synthesis", *Working Papers of the Cornell Phonetics Lab. 2*, 69-122.

[7] LADEFOGED, P. (1988), "The many interfaces between phonetics and phonology", *UCLA Working Papers in Phonetics*, no. 70, 13-23.

[8] MARCHAL, A. (1980), *Les sons et la parole*, Montréal: Guérin.

[9] MERMELSTEIN, P. (1973), "Articulatory model for the study of speech production", *J.Acoust.Soc.Am.*, vol.53, no.4, 1070-1082.

[10] SIMARD, C. and C.ROCHETTE (1985), *Etude des séquences du type consonne constrictive plus voyelle en français, à l'aide de la radiocinématographie et de l'oscillographie*. Quebec: Centre international de recherche sur le bilinguisme.

A STUDY ON DISTINCTIVE FEATURES AND FEATURE HIERARCHIES THROUGH "PHONEME ENVIRONMENT CLUSTERING" (PEC)

M. Dantsuji* and S. Sagayama**

*Faculty of Letters, Kansai University, Suita, Osaka, JAPAN
**Dept. of Speech Processing, ATR Interpreting Telephony Research Laboratories, Kyoto, JAPAN

ABSTRACT

The present study concerns the study of distinctive features by means of "Phoneme Environment Clustering" (PEC). The PEC algorithm, originally developed for automatic speech recognition, selects the optimal set of allophones and estimates missing contexts automatically. We have examined approximately 2,000 segments from 216 phonemically balanced words uttered by a male informant of Japanese using PEC. The results show that the feature [sonorant] is separated from others in the earliest stages of the process of the tree structure and coincide with the feature hierarchies proposed in the field of current non-linear phonology.

1. INTRODUCTION

In this paper we would like to describe an attempt to reconsider the hierarchical structure of distinctive features by means of a "Phoneme Environment Clustering" (PEC). Early generative phonologists adopted the distinctive features of the Jakobsonian framework [6]. Later, they revised the distinctive features in many respects. In the framework of SPE, distinctive features are mainly described from an articulatory point of view [1], and the same inclination has been maintained among current approaches. This, however, does not mean that acoustic and auditory aspects have lesser importance, but rather that it was difficult to make an exact and precise description of the acoustic characteristics of distinctive features at that time.

With respect to the hierarchy of distinctive features, several kinds of feature hierarchies have been proposed.

We would like to introduce another kind of hierarchy based on the acoustic distance. PEC, which was originally developed for automatic speech recognition, is one such experiment and attempts the establishment of the feature hierarchy.

2. THE CONCEPT OF PHONEME ENVIRONMENT CLUSTERING (PEC)

We can consider a number of possible factors, which may affect the sound patterns of a given language, such as a preceding phoneme, a phoneme before a preceding phoneme, a center phoneme (the current phoneme itself), a succeeding phoneme, a phoneme after a succeeding phoneme, speakers, pitch frequency, power, speaking rate, stress position, phoneme position in the utterance, background noise, emotion and so forth. The combination of these factors makes an abstract space which is called the environment space E. Each allophone is assumed to be a point e in the space E. On the other hand, each allophone is observed as an acoustic pattern which can be assumed to be a point v in a vector space (V), after some normalization of pattern durations as well.

If we have a set of phonetically labeled acoustic segments, each is a point e in the environment space E as well as a point v in the pattern space V. Denoting a mapping function from the space E to V by $\varphi : E \longrightarrow V$, the acoustic pattern of each allophone $v=\varphi(e)$ varies from sample to sample and has a certain spread in the space V. This spread is measured by some distortion measure, such as an averaged Euclidean distance from the centroid, and denoted by d(v). The image in a subspace Ei of the phoneme environment space E through the mapping function is also a subspace $Vi = \varphi(Ei)$ in the vector space V. Its spread in V is denoted by d(Vi).

The aim of the phoneme environment clustering is to find the optimal set of n subspaces $\{Ei\}_{i=0}^{n-1}$ to cover all variations of acoustic segments. It is defined as the minimization of the total distortion defined by:

$$D = \sum_{i=1}^{n} d(\varphi(Ei))$$

where

$$E = E1 \cup E2 \cup E3 \cup \ldots \cup En$$
$$\text{and } Ei \cap Ej = 0 (i \neq j)$$

That is, PEC aims to find an optimal division of the phoneme environment space to minimize the total sum of the distortions of images of environment subspaces. This formulation means a sort of piecewise approximation of a mapping function such that, if an arbitrary phoneme environment is given, its pattern is predicted with a minimum error. Since it is not easy to obtain the real minimum, the solution to the above problem is approximated by successive splitting of the environment subspaces, which has significant advantages such as, the clustering algorithm is simple, all produced subspaces are convex, the splitting process derives a binary decision tree, and so forth.

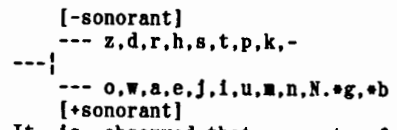
3. EXPERIMENTS ON PEC AND DISTINCTIVE FEATURES

As has been mentioned above, the process of successively splitting subspaces forms a tree structure which is interpreted as a similar grouping of phonemes and the phoneme environment. The concept of PEC can be applied as well to the distinctive features, which are components of phonemes. For Example, Fant (1973) stated that "the phonetic value of a distinctive feature can be regarded as a vector in a multidimensional signal space. The variability due to context shall be expressible by rules which define how the feature vector is changed when the conditioning elements are varied" [5]. Therefore, distinctive features may be extracted to some extent using the PEC procedure.

We have examined how sets of phonemes are divided into allophones in the process of PEC. Experiments were carried out under the following condition.

1)Informant and texts: Approximately 2,000 segments out of 216 phonemically balanced words for one male adult. 2)Acoustic parameters: cepstrum, delta-cepstrum, log-power, delta-log-power. 3)Dimension: 34. 4)Regression window: 90 ms triangular. 5)Window length:30 ms. 6)Window shift: 10 ms. 7)Sampling frequency: 12kHz. 8)Environment factors: 5. 9)Distance measure: weighted Euclidean distance.

The results indicate that allophones depending on phonetic environment are extracted at lower nodes. Phonemes as sets of allophones appropriately correspond to upper nodes which bind the lower nodes of allophones. Still upper nodes tie several phonemes into bundles and these bundles correspond to natural classes. Following diagrams represent parts of the tree structure which was formed through the process of successive splitting using PEC.

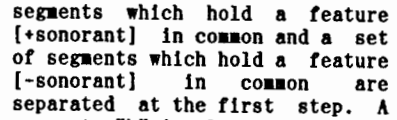


It is observed that a set of segments which hold a feature [+sonorant] in common and a set of segments which hold a feature [-sonorant] in common are separated at the first step. A segment "h" is classified as a member of segments having [-sonorant] in this analysis. In the case of Japanese, the phoneme /h/ occurs as allophones [ç], [ɸ], [x], and [ɦ] in addition to [h], and this phoneme is not usually classified as a glide. Therefore, there is no problem in classifying this segment as [-sonorant].

With respect to /r/, this segment is an approximant (semi-vowel) in the case of English, and this would be classified as [+sonorant]. In the case of Japanese, however, this segment has quite a number of allophones and free variations. For example, /r/ is often represented as a kind of plosive at word initial positions, and as a flap at word-medial positions. It is assumed that this segment is accordingly classified as [-sonorant] in this instance.

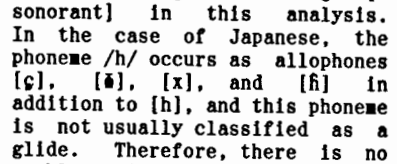
Attaching an asterisk (*) to g and d implies a special case. These segments are originally voiced plosives and should be classified as [-sonorant]. At the stage of labeling preconditioned the phoneme environment clustering, transition portions of formants were not included in vowels but included in voiced plosives. Therefore, some properties of vowels, which should be classified as [+sonorant], are assigned to these segments in this analysis. Furthermore, /g/ and /b/ seldom occur as voiced plosives [g] and [b]. Rather, they occur as voiced fricatives [ɣ] and [β] or velar nasal [ŋ] called "bidakuon". These are also assumed to be factors.

In the next step, the segments that have features [-high, -consonantal] in common were separated from [+sonorant].

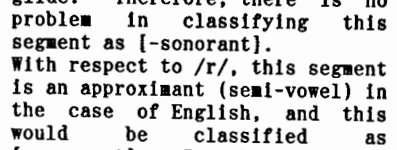


In this analysis /w/ is classified as [-high], although it is classified as [+high] in the case of English. In the case of English, [w] is produced with a constriction between the upper and lower lips and the back of the tongue and soft palate as well, and is a so-called voiced labial-velar approximant. On the other hand, in the case of Japanese, the degree of raising the back of the tongue is lower even at the word initial position, and it is pointed out that is still lower at the word medial position. Therefore, the informant of this analysis reflects such properties of Japanese, and /w/ was classified as [-high].

The group which holds [-high, -consonantal] is subdivided into a group which has a feature [+round], viz./o/ and /w/, and a group which has a feature [-round], viz. /a/ and /e/.



The segments that have a feature [-round] in common are still subdivided into individual phonemes of /a/ and /e/ by a feature [+/- low]. The low vowel /a/ and the non-low vowel /e/ are separated by this feature.



Other groups of segments are also subdivided into individual phonemes in a similar way.

4. DISCUSSION AND CONCLUSION

Recently, there is a tendency to revise not only partial problems but also the total framework of feature systems in many ways. One of the main concerns among them is setting up a hierarchy structure or groupings for the feature arrangement. Until now, several kinds of feature hierarchies or groupings of features have been proposed. For example, in a Jakobsonian framework, Fant (1973) discussed a feature hierarchy depending on the economy of description [5]. From the automatic recognition study, Dantsuji (1989) proposed a feature hierarchy making use of auditory distance [3]. In a generative phonology framework, for example, Clements (1985) discussed feature hierarchy geometrically organized from a phonological point of view considering articulatory aspects. and Sagey (1986) elaborated this feature hierarchy from phonetic and physiological facts [2,9]. These phonetic and physiological facts mean that speech sounds are produced with the movement and action of a physiologically limited number of articulators, as was pointed out by Maddieson and Ladefoged (1989), etc. [7]. Movable articulators are lips, tongue tip, tongue blade, tongue dorsum, tongue root, soft palate, larynx and so forth. Therefore, as terminal features [high]. [back] and [low] have, for example, relevance to the movement of the dorsum of the tongue, they are dominated by a non-terminal node dorsal. As labial, coronal and dorsal are related to the place of articulation, these nodes are dominated by a higher node place. Furthermore, the place node and soft palate node are dominated by a still higher node, the supralaryngeal. However, major class features such as [sonorant] and [consonantal] are directly dominated by a root node which is the highest position of the hierarchy, or situated as special features that constitute the root node.

On the other hand, the analysis by PEC establishes another type of feature hierarchy which reflects the acoustic distance. Features such as [sonorant] and [consonantal] are extracted at quite early steps in this experiment. For example, [sonorant] is extracted at the first step of the clustering. These matters indicate that the acoustic distance between segment groups corresponding to the feature [+sonorant] and [-sonorant] is considerably great. Therefore, this confirms the view that the feature [sonorant] is placed at a higher position of the feature hierarchy, as proposed in current literature of non-linear phonology based on articulatory and physical facts.

5. REFERENCES

[1]CHOMSKY,N. AND M.HALLE (1968), "The Sound Pattern of English", New York: Harper and Row.
[2]CLEMENTS,G.N.(1985), "The Geometry of Phonological Features",Phonology Year Book 2,225-253.
[3]DANTSUJI,M.(1989),"A Tentative Approach to the Acoustic Feature Model", Revue de Phonétique Appliquée, #91,92,93, 147-159.
[4]DANTSUJI,M. AND S.SAGAYAMA(1989),"A Study on Acoustic Aspects of Phoneme Environment Clustering and Distinctive Features", IEICE Technical Report SP 89-79, 25-32, (in Japanese).
[5]FANT,G.(1973), "Speech Sounds and Features", Cambridge, MA, The MIT Press.
[6]JAKOBSON,R, C.G.M.FANT AND M.HALLE (1952), "Preliminaries to Speech Analysis: The Distinctive Features and their Correlate". (1969, Cambridge, MA, The MIT Press)
[7]MADDIESON, I. & P. LADEFOGED (1989),"Multiply articulated segments and the feature hierarchy", UCLA Working Papers in Phonetics, 72, 116-138.
[8]SAGAYAMA,S.(1989), "Phoneme Environment Clustering for Speech Recognition", ICASSP-89, 397-400.
[9]SAGEY,E.(1986),"The Representation of Features and Relations in Non-linear Phonology", Diss., MIT.

LE DEBIT DE PAROLE : UN FILTRE UTILISE POUR LA GENERATION DES VARIANTES DE PRONONCIATION EN FRANCAIS PARISIEN

Anne Lacheret-Dujour

LIMSI-CNRS BP 133 91 403 Orsay-Cedex France

ABSTRACT

This paper describes the development of a grapheme-to-several phoneme strings module according to speech rate in French. Some examples of the phonological variations linked to the speech rate and the basic principles of the system are presented.

1. INTRODUCTION

Puisqu'il n'existe pas de prononciation standard en reconnaissance de la parole multilocuteur, l'intégration de modules de génération automatique des variantes de prononciation dans des systèmes de reconnaissance phonétique est nécessaire pour l'accès au lexique. De tels systèmes ont été réalisés au LIMSI : GRAPHER [5] et VARION.0 [1]. Les tests de VARION.0 ont révélé la complexité du problème lié à la génération automatique des variantes de prononciation : du fait de la production équiprobable et maximaliste des variantes, l'explosion combinatoire des chemins allophoniques produits est inévitable. Il est donc irréaliste de penser pouvoir utiliser de façon optimale ces systèmes dans des complexes de reconnaissance; l'utilisation d'heuristiques, qui permettent le cas échéant de bloquer la génération de certaines variantes, est indispensable.
Nous avons développé pour le français parisien un module de transformation graphème-phonème avec variantes, VARION.1 [3], dans lequel la génération des allophones est conditionnée par le débit de parole (lent ou rapide). Les règles du module ont été développées à partir de l'observation de corpus de parole continue en situation de lecture, prononcés par quatre locuteurs à différents débits (lent, normal et rapide) [2]. L'objet de cet article est de présenter deux exemples de variantes liées au débit prises en compte par le système : la prononciation variable du schwa et la fusion vocalique. L'architecture générale du module de règles (logiciel utilisé et formalisme adopté) est également décrite.

2. LE SYSTEME DE REGLES

2.1 Génération des variantes en fonction du débit de parole : présentation des règles.

* *Elision facultative du schwa*

Indépendamment du débit, l'élision en début de mot n'est envisageable que si le schwa est précédé d'une seule consonne dont le contexte gauche est autre que le graphème *'e'* en finale de lexème (*la dame demande* –► */ladamdəmãd/* et non */ladamdmãd/*).

Quand plusieurs *'e'* se suivent les séquences graphémiques suivantes sont à distinguer :
- *'monosyllabe#ne' (mais je ne sais pas)* ou *'ne#monosyllabe' (ce pantalon ne te va pas)*. Dans ces contextes, seul le *'e'* de la négation peut être élidé en débit lent. En débit rapide, la négation peut être omise totalement. Le schwa restant peut ne pas être prononcé dans la mesure où le contexte gauche le permet (*je ne sais pas* –► */Sspa/*).
- Dans la séquence *'ce#que'*, seul le premier *'e'* peut être élidé quel que soit le débit (*je sais ce que tu penses*).
- Dans les autres séquences, un *'e'* sur deux peut tomber à partir du premier si le contexte gauche le permet, à partir du second dans les autres cas (*mais redemande-le*).

Les règles sur la chute du schwa isolé en fonction du débit de parole varient relativement à sa position dans le mot :
- En début de polysyllabe ou dans un monosyllabe, le *'e'* est toujours maintenu en débit lent, il peut être élidé en débit rapide (*demander* –► */dəmãde/* en débit lent, */dəmãde/* ou */dmãde/* en débit rapide.
- En milieu de polysyllabe, le schwa est élidé obligatoirement en débit rapide, facultativement en débit lent (*seulement* –► */soelmã/* en débit rapide, */soeləmã/* est également prévu en débit lent).
- En fin de polysyllabe, l'élision facultative, obligatoire ou interdite des finales *'-e', '-es', '-ent'* dont le contexte droit est consonantique dépend du nombre de consonnes à gauche de la finale. Les règles sont les suivantes pour les contextes consonantiques gauches ci-dessous :
R1 [+1cons]: la finale est élidée quel que soit le débit (*une robe verte* –►*/ynrɔbvɛrt/*).
R2 [+cons,-liq][+liq]: la final est toujours maintenue en débit lent, facultativement en débit rapide, entraînant avec elle la chute de la liquide qui la précède (*ils peuplent Paris* –► */ilpœpləpari/* en débit lent, */ilpoeppari/* est une variante possible en débit rapide).
R3 [+liq][+cons,-liq]: la chute du *'e'* est facultative quel que soit le débit (*une valse de Vienne*).
R4 [+2cons]: la chute de la finale est tolérée en débit rapide uniquement (*un texte de base*).
Quand la finale *'-ent'* est suivie d'un mot à initiale vocalique, le *'e'* est toujours élidé en débit rapide, en débit lent il peut être entendu si la liaison est effectuée (*ils aiment y aller* –► */ilzɛmətijale/*).

Un certain nombre d'exceptions sont à noter à ces règles générales, pour lesquelles, quel que soit le contexte et le débit, le *'e'* est toujours maintenu (*femelle, relier*). Il en va de même dans le déterminant *'le'* accentué, dans le démonstratif *'ce'* suivi d'une voyelle (*ce en quoi*). En revanche, dans les formes du futur, le graphème *'e'* correspond à un phonème 0 si le contexte gauche est autre que *'[+obstr][+liq]' (il aidera* –► */ilɛdra/*). Il en va de même pour le pronom *'je'* postposé au verbe (*qui suis-je?*) ainsi que dans la tournure interrogative *'est-ce' (qui est-ce qui vient?*).

* *Fusion vocalique*

Si 2 voyelles identiques sont séparées par une frontière de mots et éventuellement un *'h'* aspiré, elles peuvent être réduites en un seul et même long segment. Il s'agit de *fusion vocalique.* Dans le module, la fusion est produite pour toutes les voyelles en débit rapide uniquement puisqu'aucune considération de type syntactico-sémantique n'est prise en compte pour affiner les règle. Si les graphies correspondant aux phonèmes *'o', 'e', ø'*, qui en syllabe inaccentuée non fermée par la liquide *'r'* peuvent être facultativement réalisées ouvertes ou fermées, sont à l'initiale de polysyllabes éventuellement précédées d'un *'h' muet (offert, aimable, heureux*), si elles ont comme contexte gauche respectivement les phonèmes *'o', 'e'* ou *'ø'*, leur degré d'aperture doit être identique à celui du contexte gauche (*le corbeau officie le mardi* –► * */lɔkɔrbɔofisi lɔmardi/*. La fusion vocalique est alors une variante libre produite en débit rapide : */l(ɔ)kɔrbo:fisil(ɔ)mardi/*.

2.2 Le système de règles VARION.1.

Les principaux objectifs lors du développement de VARION.1 étaient les suivants :
(1) Formaliser les variantes par le biais de règles et non en faisant usage d'un lexique. Car, la structure de la langue française ne justifie pas l'emploi d'un dictionnaire couteux en espace mémoire; les irrégularités rencontrées sont dues pour une faible part à la structure morphologique de la langue; l'accent lexical et le ton, qui jouent un rôle distinctif dans certaines langues, sont inexistants en français. Enfin, l'utilisation d'un lexique nécessite une maintenance rigoureuse afin de traiter correctement les néologismes.
(2) Effectuer une transcription graphème-phonème sans passer par l'intermédiaire de formes de base au sens chomskyen du terme; concept qui sous-entend une notion d'écart par rapport à une norme abstraite dont la définition est loin d'être claire.
(3) Adopter un formalisme permettant une écriture simple, économique et compacte des règles, qui doivent être facilement testables et le cas échéant modifiables. Pour ce faire, nous avons utilisé le compilateur de règles LEX [4] qui permet d'effectuer n'importe quel traitement linguistique sur une chaîne de caractères donnée en entrée (Fig.1).
Les règles, dans lesquelles ont peut inclure des traitements procéduraux, sont écrites sous forme déclarative et compilées en langage C. Le programme effectue ainsi un certain nombre d'actions spécifiées par l'utilisateur. Il génère ensuite un automate déterministe d'états finis. Le temps requis pour l'exécution des règles dépend de la taille du texte à phonémiser. Dans un fichier, les règles sont ordonnées, non cycliques, elles sont déclenchées de gauche à droite de la forme à transcrire, elles sont du type :

A *o("B");*

En tout point du texte donné en entrée où l'on trouve le caractère *'A'*, appliquer la fonction o(s) pour le remplacer par le caractère *'B'*. On substitue ainsi un buffer phonémique à un buffer graphémique au fur et à mesure de la transcription. *'A'* peut être un mot, un graphème ou même une séquence de mots. *'B'* peut correspondre à 0, 1 ou plusieurs phonèmes.
La figure 2 représente la phonémisation de la séquence graphémique *"Le corbeau officie seulement le mardi"* en débit lent et en débit rapide.

3. CONCLUSION

Les tests du système sur de la parole lue mettent en lumière l'amélioration des résultats (98.8% de variantes prononcées et prévues par le système en débit lent, 98.4% en débit rapide contre 95.5% pour VARION.0 tout débit confondu). Néanmoins, l'analyse de ces résultats amène les conclusions suivantes : des connaissances prosodiques supplémentaires (la distribution des pauses par exemple) sont nécessaires pour améliorer les performances d'un tel système. Il en va de même des connaissances syntaxiques. Enfin, le débit de parole est une variable relative. Les variations de débit ne sont pas toujours exécutées de façon identique d'un sujet à un autre. Chez un locuteur donné les variations sont également possibles. De ce fait, à un débit particulier, les choix allophoniques peuvent varier d'un groupe de locuteur à un autre et chez un même sujet lorsqu'il répète la même séquence de parole. Il est donc nécessaire d'étudier, outre les facteurs linguistiques, les mécanismes extralinguistiques (sociolectes, idiolectes) et para-linguistiques (situation de discours, émotivité du locuteur, etc) qui sous-tendent les stratégies allophoniques pour des classes de locuteurs données. Une telle étude permettrait de déclencher les ensembles de règles appropriés à un groupe de locuteurs spécifique.

Légende des symboles utilisés

R: règle.
[+1cons]: une et une seule consonne.
[+2cons]: deux consonnes ou plus.
(e/es/ent/s): finales facultatives.
[-son]: consonne moins sonante.
[+ocl]: consonne occlusive.

REFERENCES

[1] A. Dujour, Octobre 1987, *"Conversion graphèmes-phonèmes avec variantes du français par règles"*. DEA interuniversitaire de phonétique, Université de Paris 7 (Jussieu).

[2] A. Lacheret-Dujour, Octobre 1989, *"Automatic Generation of Phonological Variations"*, EUROSPEECH, vol 2, pp 376-379, Paris.

[3] A. Lacheret-Dujour, Juin 1990, *"Contribution à l'analyse de la variabilité phonologique pour le traitement automatique de la parole continue multilocuteur"*. Thèse de Doctorat de l'Université de Paris 7 (Jussieu).

[4] Manuel d'utilisation UNIX : Programming Utilities and Libraries SUN microsystems, 1988, *"Lex, a Lexical Analyser Generator"*. pp 203-267.

[5] F. Néel, M. Eskénazi, J. Mariani, 1986, *"Module de traduction phonétique avec variantes"*. Actes du Séminaire : "Lexique et traitement automatique des langues" du Greco Communication parlée et du Galf, pp 129-138, Toulouse.

FIGURES

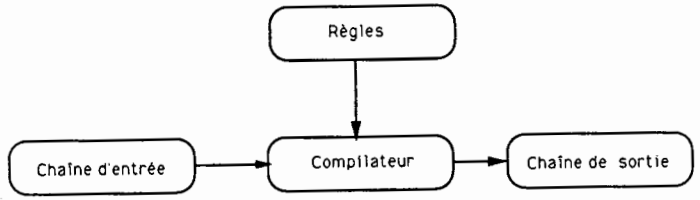


Fig.1: Le compilateur de règles, LEX.

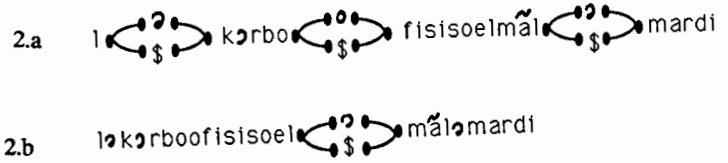


Fig.2: Phonémisation de la séquence *Le corbeau officie seulement le mardi,* **en débit rapide (Fig.2.a) et en débit lent (Fig.2.b)**

LA LIAISON A ORLEANS (FRANCE) ET A MONTREAL (QUEBEC)

Daan de Jong

Université Libre d'Amsterdam / Université de Montréal

This paper summarizes some of the most important findings from a large scale, ongoing research project after sociolinguistic variation in liaison usage in Orléans French (France) and Montréal French (Québec).

0. INTRODUCTION[1]

La liaison est la réalisation d'une consonne latente finale devant un mot à initiale vocalique (voir 1). Devant un mot à initiale consonantique ou en fin de phrase, une consonne latente n'est jamais réalisée (voir 2):

1) chez eux [ʃezØ]
   petit ami [ptitami]
2) chez lui [ʃelɥi]
   il est petit [ilɛpti]

La réalisation d'une consonne latente devant une voyelle dépend de plusieurs facteurs. Nous examinerons de quelle façon la variation dans l'emploi de la liaison est influencée par les facteurs suivants: la structure syntaxique, la classe sociale, l'âge et le sexe. De plus, nous comparerons l'emploi de la liaison dans deux variétés du français: le français d'Orléans en France, et le français de Montréal au Québec.

Les données viennent de deux corpus de français parlé: le corpus d'Orléans de Blanc et Biggs [1] (dont nous avons dépouillé 45 entrevues), et le corpus de Montréal de Sankoff e.a. [10] (dont nous avons dépouillé 33 entrevues). Les deux corpus consistent en des entrevues relativement informelles. Les informants se divisent de façon égale sur 5 classes sociales et 3 groupes d'âge. Le nombre de femmes et d'hommes est à peu près égal. Les deux corpus ont été enregistrés presqu'en même temps (en 1969 et en 1971). Ceci rend les deux corpus comparables.

1. LE ROLE DE LA SYNTAXE

Un premier facteur affectant la fréquence d'emploi de la liaison est la structure syntaxique [8, 11]. De Jong (1990) démontre que la structure syntaxique doit d'abord être transformée en une structure prosodique hiérarchique consistant en trois couches de constituants prosodiques: le Groupe Clitique (GC), la Petite Phrase Phonologique (PPP) et la Phrase Phonologique Maximale (PPM). Dans le premier constituant, la liaison est très fréquente, dans le deuxième elle est d'une fréquence moyenne, et dans le troisième elle est rare.

La dérivation en consituants prosodiques présuppose une analyse de la phrase en terme de la théorie X-bar. Ainsi, la fin (droite) de chaque tête (X) délimite le domaine de la liaison fréquente (ou obligatoire). Nous considérons comme tête les catégories majeures N, A ou V, et aussi les catégories mineures P, Comp et Aux [6,7]. Ainsi, la phrase ils ont été aidés par des enseignants admirables est divisée comme suit en GCs: (ils ont) (été) (aidés) (par) (des enseignants) (admirables). ils et ont sont dans un même GC. Donc, on peut prédire que la liaison après ils sera très fréquente, sinon obligatoire. La même chose vaut pour des et enseignants.

La PPP est dérivée en choisissant seulement les catégories N, A ou V comme fin de domaine. La phrase citée ci-haut sera divisée comme suit en PPP: (ils ont été aidés) (par des enseignants) (admirables). Au niveau de la PPP, ont et été sont dans le même constituant prosodique: on peut prédire que la liaison après ont se fera avec une fréquence moyenne.

Finalement, la PPM est dérivée en prenant chaque fin d'une projection maximale comme la fin d'un constituant prosodique, ce qui donne le résultat suivant: (ils ont été aidés) (par des enseignants admirables). Le z final de enseignants est dans le même domaine que admirables, alors on peut prédire qu'occasionnellement ce z final peut être réalisé.

La liaison doit être plus fréquente dans la GC que dans la PPP, et dans la PPP elle est plus fréquente que dans la PPM. Nous avons testé cette hypothèse sur 45 entrevues du corpus d'Orléans. La figure 1 montre clairement que l'hypothèse est confirmée. En plus, cette figure montre, que la hiérarchie GC > PPP > PPM vaut pour toutes les classes sociales. Finalement, nous voyons que l'emploi de la liaison décroit de façon régulière avec la classe sociale.

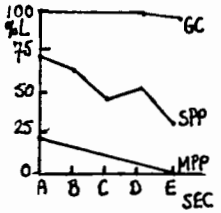


Figure 1. Pourcentage de liaison (%L) dans 5 classes socio-économiques (A) la plus élevée, (E) la moins élevée).

2. LE FACTEUR LEXICAL

L'analyse précédente ne suffit pas à elle seule à prédire l'emploi de la liaison. Dans De Jong (1988, 1991), nous avons présenté une analyse statistique des données relevées sur le corpus d'Orléans au moyen d'un modèle loglinéaire. Cette analyse a démontré que plusieurs propriétés du mot contenant la consonne latente, affectent la fréquence de la liaison. Ainsi, la liaison était

significativement plus fréquente après des mots monosyllabiques qu'après des mots polysyllabiques. La liaison avec /t/ était plus fréquente que la liaison avec /z/. La catégorie grammaticale avait aussi un effet significatif. Finalement, la liaison se faisait plus souvent après les mots très fréquemment utilisés qu'après les mots peu fréquemment utilisés. Des résultats comparables ont été obtenus pour le Corpus de Montréal.

3. LES FACTEURS EXTRALINGUISTIQUES

La fréquence d'emploi de la liaison est aussi significativement influencée par plusieurs facteurs extra-linguistiques. Nous avons examiné le rôle de la classe sociale, de l'âge et du sexe, pour Orléans et pour Montréal. Les principaux résultats sont résumés dans les figures 2A, 2B et 2C, qui montrent que l'emploi de la liaison décroît avec la classe sociale, augmente avec l'âge, et que les femmes utilisent plus de liaison que les hommes.Ces données montrent aussi, contrairement à ce qui est dit dans Encrevé (1988: 50) que la liaison variable n'est pas limitée aux classes sociales supérieures, mais se retrouve dans toutes les classes sociales. Ces figures montrent aussi que la liaison se comporte tout à fait comme les variables socio-linguistiques décrites dans Labov (1972), et absolument pas comme 'une variable socio-linguistique inversée' (Encrevé 1988: 45). Finalement, ces figures montrent que cela vaut aussi bien pour Orléans que pour Montréal (voir aussi [2,4].

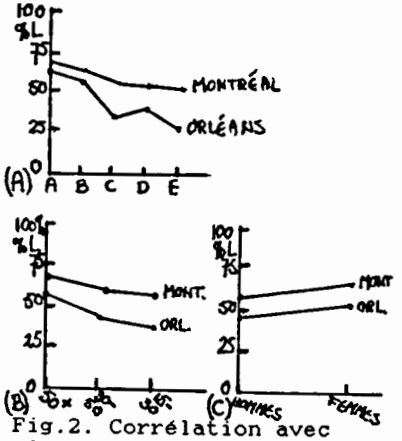


Fig.2. Corrélation avec classe sociale (A), âge (B) et sexe (C).

4. DIFFERENCES ENTRE ORLEANS ET MONTREAL

A part des ressemblances, il y a aussi des différences importantes entre Montréal et Orléans: Par exemple, à Montréal, la liaison après suis se fait souvent non avec /z/, mais plutôt avec /t/. La fig.5 montre que l'emploi du /t/ après suis est fréquente dans les trois classes inférieures, mais presque absente dans la classe supérieure.

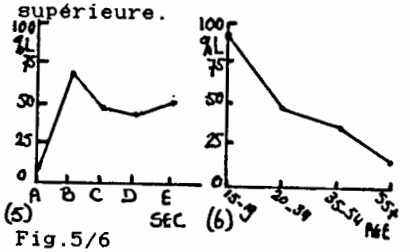


Fig.5/6

La fig.6 montre que

l'emploi de la liaison avec /t/ est beaucoup plus fréquente parmi les jeunes que parmi les plus âgés, ce qui suggère que la liaison avec /t/ après suis est un nouvel emploi qui est en train de se répandre (voir aussi [4]).

Une autre particularité du français montréalais est l'optionalité de la liaison après le pronom indéfini on et après le pronom personnel ils. Les figures 7 et 8 montrent que l'emploi de la liaison après on et ils augmente avec la classe sociale.

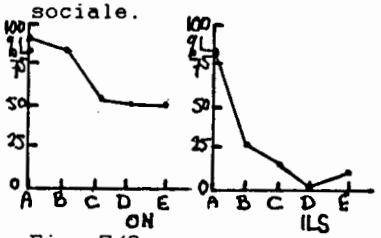


Fig. 7/8

5. NOTES

1. Je veux remercier Pierrette Thibault, David Sankoff et Henrietta Cedergen pour m'avoir donné accès au corpus de Montréal. Je tiens également à remercier le Département de linguistique et de philologie de l'Université de Montréal, où j'ai pu écrire cet article comme professeur invité. Les recherches pour cet article ont été rendues possibles grâce à l'Académie Royale Néerlandaise des Sciences et des Lettres.

6. REFERENCES

[1] Blanc, M. & P. Biggs (1971). 'L'enquête socio-linguistique sur le français parlé à Orléans'. Le français dans le monde 85. Pp.16-25.
[2] De Jong, D. (1988). Sociolinguistic Aspects of French Liaison. Ph.D. Thesis. Amsterdam: Free University.
[3] De Jong, D. (1990). 'The syntax-phonology interface and French liaison'. Linguistics 28. Pp.57-88.
[4]De Jong, D. (1991). 'Sociophonological aspects of Montréal French Liaison'. To appear in: Proceedings of the 21st Linguistic Symposium on Romance Languages.
[5] Encrevé, P. (1988). La liaison avec et sans enchaînement. Paris: Le Seuil.
[6] Guéron, J., & T. Hoekstra (1988). 'Les chaînes-T et les verbes auxiliaires. Lexique 7.
[7] Jackendoff, R. (1977). X-bar Syntax: A Study of Phrase Structure. Cambridge: MIT Press.
[8] Kaisse, E.M. (1985). Connected Speech. The Interaction of Syntax and Phonology. Orlando, etc.: Academic Press.**
[9] Labov, W. (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.
[10] Sankoff, D., G. Sankoff, S. Laberge & M. Topham (1976). 'Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale' Cahiers de Linguistique de l'Université du Québec 6. Montréal: Presses de l'Université du Québec. Pp. 85-125.
[11] Selkirk, E.O. (1986). 'On derived domains in sentence phonology'.

VERB STRESS IN SPANISH

S. Alcoba

Universidad Autónoma de Barcelona.

ABSTRACT

In this study, extrametricality is put forward and the hierarchy of stress markedness is precisely identified among verbal forms in the Thematic Vowel (TV) affix, keeping the stress rules parameters in a pancategorical sense, in order to avoid some incoherences which we shall refer to in the solution of J.W.Harris, the most solid and best directed proposal.

The hypothesis in connection with the morphological determination of the verbal stress in Spanish is in a state of crisis. What is mentioned in [4], page 84, that "Segmental phonological representation and morphological identification are jointly necessary and sufficient to determine placement of word stress for all verbforms" is nowadays considered to be a challenge which needs to be and can be overcome: [5], [7] and [8] point in this direction.

1. EXTRAMETRICALITY OF THE FINAL METRIC ELEMENT.

1.1. Established generalisations.

The generalisations in connection with the stress in Spanish words which I consider to be established are the following:

First. Spanish words, as far as the accent is concerned, are hierarchically divided into three classes: on the one hand, a wide selection of words with regular stress on the penult syllable, or type A (*sa'banas, ex'tensas, bal'cones, a'zules*); on the other hand, a collection of words with irregular stressing, divided into two subsets: one consisting of words with the stress on the antepenult syllable, or Type B (*'sabanas, 'comicas, 'arboles, 'utiles*), and the other of words with the stress on the last syllable, or Type C (*fara'laes, ca'fes, hu'ries, marro'quies, domi'nos, ban'tues*). The roots or affixes of type B are marked in the lexicon with a responsible diacritic feature that expresses the markedness of the stress in Type B words. The unmarked lexical entries will be of Type A/C.

Second. All morphological words, of whatever category, fall in with the so-called window of three syllables restriction (WTS) which makes the existence of words *(X'----) with the stress on one syllable further back than the third from the end impossible in Spanish .

Third. The Spanish stress is sensitive to quantity (Braching Condition of [4]). An stress of Type B is not possible if the penult syllable has a branching rhyme: *(X'-VC-), *(X'-VG-), *(X'-GV-). Nor is a Type B stress possible in words ending in a final rhyme GV, *(X'--GV) (**'continua, con'tinua, conti'nua*). Unther these circumstances the window for stress is reduced to the last two syllables.

Fourth. The domain of stress assignment is the morphological word. The rules or parameters for the assignation of the stress explore the whole word (theme and non-cyclic or inflectional affixes) and establish the stress in accordance with the foregoing conditions.

The hypothesis of the word as the domain of the stress rules presupposes that the parameters explore lexical forms in a derivative stratum in which the non-cyclic affix constituents (word markers and paradigmatic costituents: πG^{o}, πN^{o},

πTMA and πPN) are specified.

Fifth. The representation of the stress is placed within the framework of the theoretical model developed in [3].

1.2. Harris's stress rules.

The stress rules in [7] to generate the stress grids of Spanish words are collected in (1) and illustrated i (2) where, for simplicity's sake, only line O of the stress grid is shown.

(1)*1. Stressable elements are syllable nuclei (rhyme heads).*
2. The rightmost stressable element is extrametrical iff word-final or followed by an inflectional consonant.
3.Form constituent(s) on line 0 and mark head(s) on line 1; Parameter settings:
a.unbounded,right-headed (general case).
b.binary, left-headed, right-to-left (special case).
4. Form constituent(s) on line 1 and mark head(s) on line 2; Parameter settings: unbounded, right-headed.
5. Conflate lines 1 and 2 (=remove asterisks in columns that have no line-2 asterisks).

(2).

con.tes.tas con.tes.ta.mos
(. *)<*> (. . *) <*>

con.tes.ta.bas con.tes.ta.BA.mos
(. . *)<*> (* .) (* .)<*>

1.3. Limits and drawbacks to Harris's hypothesis.

In the manner in which extrametricality is formulated in (1.2), the distinction is maintained between words of Type A (*bal'cones, a'zules)* and of Type C (*domi'nos, ban'tues*) and these last remain pending an exceptional explanation (cfr. [7]: 257 and n. 4). Among the verb forms the final resolution of the oxiton forms of the "weak preterite" and of the "future" also remains outstanding. (cfr. [7]: 257).

Furthermore, it is obvious that the difficulty in generating the infinitive itself *contes'tar* by reason of the incoherence deriving from calling the final segment a derivative. In addition, if the paradigmatic morphemes -BA-, -RA/SE-, -RE- and -STE- are considered as being Type B to explain the accent on forms such as *contes'ta.BA.<mos/is>*, how can the stress of *contes'ta.<BA(s/n)>* be explained? In the same way as occurs in *contes'ta.STE.<is>* as against *contes'ta.<STE>*.

2. EXTRAMETRICALITY OF FLEXIBLE METRICAL ELEMENTS.

2.1. Hierarchy of markedness in verb forms.

I suggest that it is the outermost cyclic affix of the verb forms, the various forms of TV which carry the diacritic of markedness. The morpheme TV (a, i, i) of the Theme of the Preterite (cfr. [1] and [2]) will be of the type A: weak preterite, imperfect indicative, imperfect and future subjunctive, gerund and participle. The morpheme TV (A, E, E) of the Theme of the Present will be of type B: present indicative, present subjunctive and imperative. And the morpheme TV (a, e, i) of the Theme of the Future will be of Type C: future indicative and conditional.

I suggest furthermore, that extrametricality be understood in terms of (3):

(3). **Extrametricality** (replaces (1.2))
A stressable element is extrametrical iff it matches an inflectional constituent.

The formulation of (3) does not have a higher theoretical cost than (1.2): "It is perfectly straightforward to distinguish between «inflectional» and «non-inflectional» morphemes in Spanish... the set of «inflectional» morphemes contains exactly class markers and the plural morphemes in non-verbs plus tense/mood/aspect and person/number suffixes in verbs... the set of «inflectional» morphemes correspons exactly to the noncyclic affixes in the Halle-Vergnaud theory of phonological organization". (cfr. [7]: 253).

If (1.2) is a pancategorial formulation, so is (3); but (1.2) has to treat as exceptions words of Class C, which is not necessary with the formulation of (3): words which lack flectional constituents will not have extrametrical elements (cfr. in [6]: 38, the analysis of *ale'man* and *'huesped*, lacking word marker or inflectional elements and therefore extrametrical elements).

Now, given the formulation of (3), if it

is postulated that the domain scanned by the stress rules (1.3-5) is the derivative theme (according to [8] p. 11, "The domain of Spanish stress is the lexical word. Clearly, the TMA and PN verbal suffixes are inflectional, and thus not included in the domain of the lexical word") and not the word, it would be possible to do without (3). But the arguments of [7] are strong enough. Syllabification must precede the stress for this to show its sensitivity to WTS and the syllabic quantity (cfr. [7] p. 28). The inflectional elements, when they increase the number of syllables in the word, affect the placing of the stress by virtue of the WTS restriction (*re'gimenes, averi'guamos, averi'güemos)* and the quantity sensitivity of the Spanish stress (*averi'guais, averi'güeis*) (cfr. [6] p. 29-30). The class of inflectional elements is a closed class and by extension perfectly identifiable as well as by its non-cyclic nature, while the derivative themes make up an open class of elements which, in some cases, is difficult or impossible to establish: in the metalinguistic uses of prepositions, conjunctions and proclitics; in fragments (*'pros, 'contras*); in acronyms and abbreviations (*'talgo, 'ONU, U'NESCO*); and in apocoristics and formations by reduction or shortening, (*'Emi, 'Santi, 'cole, 'repe, 'porfa*), which form words with regular systematic stress, Type A, or which tend to become regularised with use, (*'radar > ra'dar*).

Therefore, it is necessary to retain the morphological word as a sequence scanned by the stress rules and a principle of extrametricality (3) specific to Spanish, albeit of little theoretical value, because it is established in generic terms. The concept of extrametricality in [8], p. 21, "The 'desinence' is extrametrical", although apparently similar to (3) is very different, according to [8] itself, p. 12 which explains: "Desinences in nonverbals are excluded by extrametricality, while in verbals both clitics and inflectional endings simply fall outside the domain". Roca's stress rules do not scan the extrametrical elements. Extrametricality, as defined by Roca, is redudant with his domain proposal (cfr. [8] p. 11).

I therefore suggest that (3) be understood in the sense of [6], p. 38, "the class marker is within the domain of scansion of the stress rules, but extrametrical". The difference between the stress domain hypothesis and the extrametricality hypothesis may be subtle but it is crucial, as Harris himself observes because the arguments previously put forward make the first hypothesis untenable but do not affect the second. The rules scan all the metric elements of the word but only the non extrametrical elements count.

Thus, the analysis of the examples in (2) would now be that of (5) where the forms of the Theme of the Future would be exceptional to (3) in which the element immediate to the Theme is not declared extrametrical.

(5)a. *General stress, Type A.*

```
con.tes.t+a.#bas        con.tes.t+a#ba.mos
(.   .    *) <*>        (.   .    *)<*><*>
```

b. *Marked stress, Type B.*

```
con.tes.t+A.#-/e+s
     (*    .) <*>

con.tes.t+A. #-.mos
     (*     .)<*><*>
    ----------------
WTS:    -->   *)

con.tes.t+A.#- .  is
     (*     .)<*><*>
    ----------------
WTS:    --> *)

con.tes.t+A.#e. mos
     (*     .)<*><*>
    ----------------
WTS:    -->  *)
TVø:    -->  -  *)

con.tes.t+A.#e:  is
     (*     .)<*><*>
    ----------------
WTS:    -->  *)
TVø:    -->  -  *)
```

c. *Special stress, Type C: exception to (3)*

```
con.tes.t+a.re.(mos/is)
(.   .   .  *) <*>
con.tes.t+a.ra(s/n)
(.   .   .  *)
```

2.2. Outstanding questions.

In this analysis the oxiton forms of the weak preterite (*contes'te, contes'to; compren'di, compren'dio*) remain outstanding where there has to be a special solution parallel to those concerning the strong preterites (*an'duve, an'duvo; con'duje, con'dujo*).

The case of the infinitive may be more complicated, although not for the reasons put forward in [6] p. 50-51. The problem rests in establishing the TV affix Theme of the forms of the infinitive. By reason of the stress, it could be considered as the Theme of the Preterite. On the other hand, by virtue of the TV form, it should be considered as the Theme of the Future or as the Theme of the Present; but then, how would the stress be explained?

3. REFERENCES

[1] ALCOBA, S. (1989), "Tema verbal y formación de palabras en español", en *Actas del XIX Congreso Internacional de Lingüística y Filología Románicas*, (en prensa) Univ.de Santiago de Compostela, 1989.

[2] ALCOBA, S. (1990), "Morfología del verbo español: conjugación y derivación deverbal", en *Actas del VI Congreso de Lenguajes Naturales y Lenguajes Formales*, (in press) Univ. de Barcelona, 1990.

[3] HALLE, M. and J.R. VERGNAUD (1987), *An Essay on Stress*, Mass. The MIT Press.

[4] HARRIS, J.W. (1983), *Syllable Structure and Stress in Spanish*, MIT Press, Cambridge.

[5] HARRIS, J.W. (1987), "The Accentual Patterns of Verb Paradigms in Spanish", *Natural Language and Linguistic Theory* 5, 61-95.

[6] HARRIS, J.W. (1989)a, "Spanish Stress: the Extrametricality Issue", unpublished ms., MIT.

[7] HARRIS, J.W. (1989)b, "How Different Is Verb Stress in Spanish", *Probus*, 1, 241-248.

[8] ROCA, I. (1990), "Morphology and Verbal Stress in Spanish",unpublished ms., Univ. of Essex.

A Study of Vowel Coarticulation in British English

James L. Hieronymus

Centre for Speech Technology Research, Edinburgh University
80 South Bridge, Edinburgh EH1 1HN, Scotland

Abstract

Coarticulation in continuous speech causes vowel formant frequencies to be affected by nearby phonemes. Generally continuous speech causes the vowel formant targets to be centralized relative to their isolated word counterparts. The present study concentrates on 660 phonetically hand labelled sentences from one male talker of the RP accent of British English. This allows the study of coarticulation without the confounding effects of accents, speech habits and differing individual formant ranges. The 12 monophthongal vowels of RP British English /i, I, ae, ɛ, ɑ, ʌ, ɔ, o, U, u, ə, ɜ/ have been studied using formant frequency and amplitude tracks and duration, and sentential stress (sentence stress as opposed to lexical stress). Generally the vowels are most affected by nearby semi–vowels /l, r, y, w/. No simple relationship between adjacent phoneme place of articulation and the vowel target change has been found when all the vowels are treated together. However, the data shows the presence of "robust vowels" which are not greatly effected by nearby semi–vowels. These vowels are not simply stressed vowels, but depend on duration and others factors being studied. The weak effect of duration is that the pre–pausal lengthened vowels are in the "robust" category, but shorter vowels can either be robust or ordinary. The categories of function word and content word do not account for robustness.

Introduction

Most coarticulation studies have considered isolated words. An early study by Shearme and Holmes [1] showed that vowels in continuous speech very seldom had steady states and often did not overlap the Peterson–Barney [2] 95 percentile contours in any part of their frequency trajectories in time. Generally the vowels are much more centralized in continuous speech and the vowel formant regions overlap considerably due to coarticulation.

Kuwabara [3] found a renormalization technique based on the theory of Lindblom and Studdert–Kennedy [4] which disambiguates Japanese vowels in continuous speech.

Hieronymus and Majurski [5] tried this technique on American English vowels and found that it did not work well. It has been speculated that the stress structure of English causes this method to fail. The presence of "robust" vowels as found in this study would cause this technique to fail, because the renormalization is applied uniformly to all vowels.

This is a report of an ongoing study of vowel properties and coarticulation in British English. The present approach is to study the speech of one talker at a time in detail to find the underlying mechanisms in coarticulation. Thus coarticulation can be studied without the confounding effects of regional accent, speaking styles, and formant ranges due to different talkers. Then speech data from other talkers will be studied and the pooling of the data explored to achieve speaker independent results later in this study.

It is postulated that some sort of hierarchical structure of linguistic factors modifies the effect of nearby phonemes such that the same vowel in the same phonetic context will have markedly different formant frequency trajectories in time. Some possibilities for factors which have been explored are sentential syllable stress, duration, and word identity. Originally it was thought that sentential stress would be the determining factor of vowel precision of production. Previous studies by us [5], [6] for American English have shown that sentential stress is not a determining factor, based on automatic stress labelling. The present study uses hand labelled stress and shows that, on the average, sententially stressed vowels are more precisely produced than their unstressed counterparts.

Method

The data was read by one male talker of a near RP dialect of British English in a sound isolation booth. The microphone was a Shure SM–10. The speech was digitized directly using a 16 bit a–to–d converter at 20 kHz sampling frequency with an anti–aliasing filter at 8 kHz. The talker was told to speak the sentence as if he was saying it in conversation and was prompted with the sentence on a computer screen. The speech was hand labelled by graduate phoneticians at a broad phonetic level with syllable stress marked using a PC based labelling workstation. The labellers were presented with a spectrogram and could play the segments. Subsequently the sentences were parsed by hand to provide loose bracketing of phrase boundaries, so that syntactic effects could be studied. Of the 660 sentences were designed for the CSTR/ATR database project to collect and label speech for speech technology studies. The other 460 sentences were Anglicized versions of the TIMIT compact sentences designed by the MIT Speech Group.

Each vowel formant is characterized by three values for each hand labelled vowel. The values are the first and second formant frequencies at points 10 %, 50 % and 90 % of the duration of the vowel. These values were chosen to minimize the effect of formant tracking errors. Formant tracks are obtained from a centroid based formant tracker developed by Crowe [8]. Except for low formant frequency values in the nasalized vowels the formant tracker seems to have a low error rate. These values are then fed into the APS system developed at CSTR by Watson [9] providing an interface to the S package to allow statistical studies of the data.

Discussion of the Data

Figure 1 shows a scatter plot of the first and second formant values measured at the temporal center for the long British English vowels extracted from 358 sentences with ellipses representing 66 percent of the data (/ae/ is a long vowel durationally in this data even though it is phonologically lax). Figure 2 shows data for short vowels. The normal range for a male talker is 200–1000 Hz for the first formant and 800–2300 Hz for the second formant. The minimum perceptible differences (DL) in formants were measured by Flanagan [ ] and found to be +/– 50 Hz for F1 and +/– 75 Hz for F2. Thus a measure of precision of production is how large the standard deviation of the data is relative to the DL. The cross hatched area in each vowel region is the ellipse for the sententially stressed vowels.

The formant regions for most vowels are as expected except that this talker has a very fronted /u/. The vowel /ɔ / is the highest back vowel for this talker with a median second formant of approximately 800 Hz. The long schwa is more precisely produced than the reduced vowel schwa (not plotted because of its large standard deviation) with the long form having a significantly lower second formant.

While the stressed vowels are more compact in the 66 percentile ellipses, there are a a considerable number of wide ranging outliers. Secondly there is a concentration of data points towards the outer edge of these ellipses. These are the "robust" vowels as will be shown.

The short vowels have more scatter and thus seem to be produced with less precision. Once again the stressed vowels are statistically more compact than the unstressed vowels. A superposition of these plots shows a considerable overlap between vowels in the tense–lax pairs. Duration plots show that the durations of the tense vowels pairs are statiscally longer than the vowels in their lax counterpart, but that there is considerable overlap in the distributions, especially for /i/ and /I/ and /I/ and fronted /u/ for this talker.

The presence of "robust" vowels is shown in Figure 3 which shows the stylized formant trajectories for /i/ in the environment of preceding semi–vowel /w/. The smaller font characters are the preceding context and the large character represent the following context.

The question to be answered is: why do some examples of the vowel /i/ have second formant "targets" above 2100 Hz. even in this environment? The primary stress vowels in this set are shown by a round circle and the secondary stressed vowels are highlighted by a square. As we

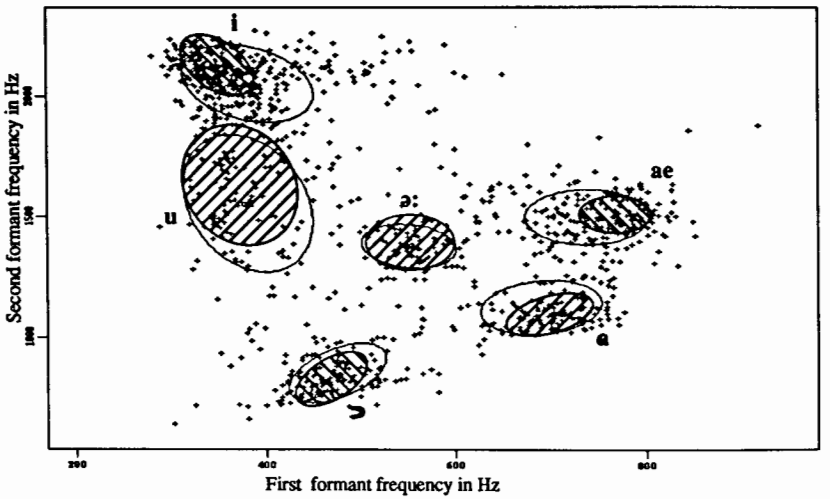


Figure 1 : First and Second Formants for British English Long Vowels

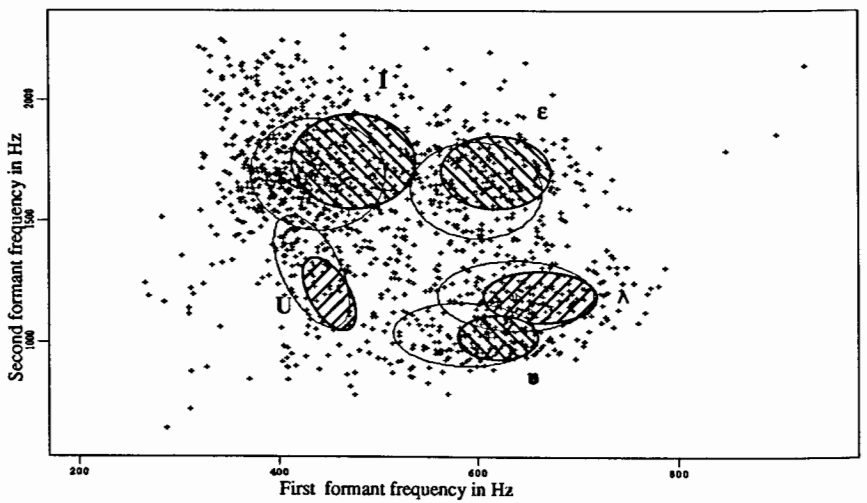


Figure 2 : First and Second Formants for British English Short Vowels

see, all of the stressed vowels and one unstressed vowel are the ones with high second formant. Consider the context /w i k/ in the figure. Two of the tracks are for the stressed vowel and appear in the robust region. One of these vowels in the same context (shown by the arrow) is not robust, and is unstressed. The /w i n/ context shows two unstressed examples, one of which is robust and one not. Thus stress does not seem to be a reliable correlate of "robustness" for vowel production. Similar plots of prevocalic /r/ show an even greater tendency for unstressed vowels to be robust but are more difficult to see at this scale.

The vowel /o/ was also examined for the presence of robust exemplars in the presence of /j/ the palatal glide, and they were found. There was a weak correlation between stress and robustness.

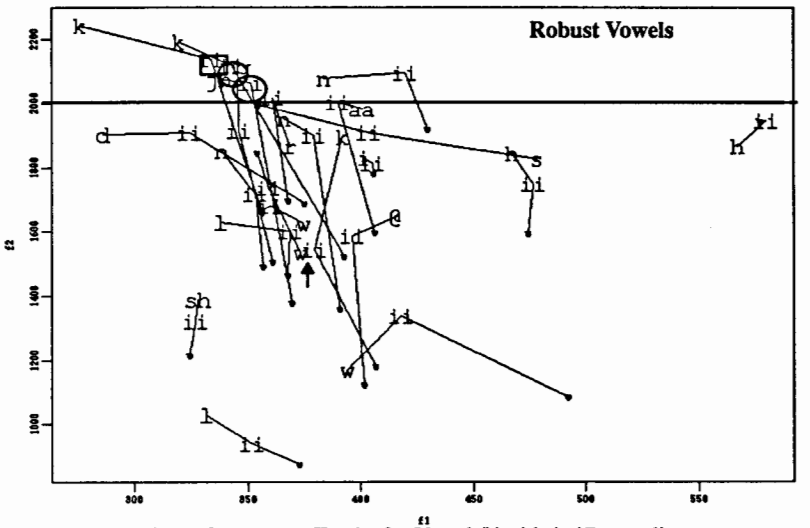


Figure 3: Formant Tracks for Vowel /i/ with /w/ Preceeding

Two possible factors, the type of word and duration, were examined to determine whether or not they determined robustness of the vowel. The longest, pre–pausal lengthened vowels are all robust. However for shorter vowels, duration is not a good correlate of robustness. Both function and content words were found to contain robust vowels. The word "between" was found to have a robust vowel on one occasion and a coarticulated vowel in another.

Acknowledgement

The programs written by Alan Crowe and Gordon Watson were used in this study. The speech labelling by Marie Alexander, Cathy Bennett, Delyth Davies, Imogen Cohen, Keith Edwards, Hazel Sydeserff, Ivan Uemlianin and Christina Barr were instrumental in this study. The collection and labelling project was supported by ATR, Japan. This work was supported by the Alvey Integrated Speech Technology Demonstrator project in collaboration with Marconi Speech and Information Systems and HUSAT, Loughborough University. The demonstrator is supported by the UK Department of Trade and Industry and the British Science and Engineering Research Council under grants D/29611, D/29628, D/29604 and F/10309.

References

1. Shearme, J.N. and Holmes, J.N. , Fourth Congress of Phonetic Sciences, Helsinki 1961 (Mouton and Co.), 233–240 (1962).
2. Peterson, G.E. and Barney, H.L., *JASA*, **24**, 175–184 (1952).
3. Kuwabara, H., *JASA*, **77**, 686 – 694 (1985).
4. Lindblom, B. E. F. and M. Studdert–Kennedy, *JASA*, **42**, 830 – 843 (1967).
5. Hieronymus, J. L. and W. J. Majurski, *Proc. ICASSP86*, Tokyo, Japan , 2787–2790 (1986)
6. Hieronymus, J. L., *Proc. ICASSP89*, Glasgow, Scotland , 608 – 611 (1989).
7. Cutler, A. and D.M. Carter, *Computer Speech and Language*, **2** (1988).
8. Crowe, A. S. and Jack, M. A., *Electonics Letters*, **23**, 1019–1020 (1988).
9. Watson, G. S., *Proc. Eurospeech* **89**, 300–303.

THE 'VOWEL-STICKINESS' PHENOMENON: THREE EXPERIMENTAL SOURCES OF EVIDENCE[1]

Bruce L. Derwing and Terrance M. Nearey

University of Alberta, Edmonton, Canada

ABSTRACT

Data are reported from three independent sources, involving active word-manipulation tasks, substitution-identification tasks and both active and passive syllable boundary tasks. All show a consistent tendency for glides in English to adhere most closely to vowels, followed by /r/, then by /l/, then by nasals, and last by obstruents. Cross-linguistic studies are now underway to test the universality of these findings, as well as formal modeling planned to account for these results.

1. BACKGROUND

We use the term 'vowel-stickiness' to refer to the tendency for some segments to adhere more closely to vowels than others [1]. Though much of the evidence for this phenomenon was conducted under the rubric of 'syllable structure' or 'intra-' or 'sub-syllabic units,' these terms imply a sharply delineated or 'hierarchical' view of syllables that is less well supported by the facts. Experimental evidence for the 'stickiness' notion comes from three distinct sources: production experiments and pattern-identification studies that were focussed on questions of the internal structure of syllables, plus a combination of production and judgment tasks that were directed at the question of syllable boundaries.

2. EXPERIMENTAL WORD GAMES (PRODUCTION TASKS)

Treiman [2,3] used a variety of experimental word games (notably word-blending) to explore the internal structure of English syllables, and Dow strengthened these findings, using primarily a unit-substitution (or deletion) task [4,5]. What all this work demonstrated was that there was more to a syllable than a simple linear sequence of (phonemic) segments. It also purported to show that well-defined 'units' were also involved (such as the onset, the rime, the nucleus/peak and the coda) and that the structure of syllables was not only hierarchical but also (at least for English) right-branching.

One disquieting fact emerged from this early work, however, to complicate the picture. Specifically, in one series of studies [6], Treiman found that the boundary between the nucleus and the coda was less than firm and, in fact, tended to shift in response to the sound class of the post-vocalic consonant involved. Specifically, subjects tended to break VCC syllables before the first consonant if that consonant was an obstruent, but after it if the consonant was a liquid (i.e., /l/ or /r/), whereas the two tendencies were of about equal strength if the first consonant was a nasal. Thus liquids (L) tended to stick with their original vowel in these tasks and

obstruents (O) to split away, with the nasals (N) holding an intermediate position. In terms of their general tendency towards vowel-stickiness, therefore, the order **L > N > O** was observed.

3. SUBSTITUTION-PATTERN IDENTIFICATION TASKS

In order to circumvent the slow and laborious production data-collection methods of these early production studies, we experimented with a new forced-choice judgment technique called the 'substitution-pattern identification task.' In this task, rather than asking subjects to actively replace some part of a syllable (such as the vowel, or an all-obstruent onset or post-vocalic coda) with a substitute segment or string, as Dow had done, subjects were trained instead merely to identify such a replacement. Thus, in a training session, subjects were orally presented with a dozen or so examples of a particular substitution pattern (e.g., replace the vowel by /I/; or delete the onset; or replace the coda by /ps/); then, in the testing phase, the subjects were asked to respond to new word pairs, merely by indicating whether the subsitutions involved were the same ('YES') or different ('NO') in kind to the particular pattern that they were trained on. Reinforcement items from the training set were also regularly interspersed among the test items, in order to remind subjects of the nature of the pattern that they were looking for (see [1,7,8] for details).

What was critical about the test items in this last study was that they all contained either pre- or post-vocalic sonorant consonants, and these were sometimes replaced along with the units in question and sometimes not. Thus, having been trained to replace an all-obstruent coda by /ps/ (as in /vIk/-/vIps/ or /f∧sk/-/f∧ps/), a subject might now be asked whether the nonsense-pair /r ɛlst/-/r ɛps/ illustrated the pattern (where all post-vocalic consonants were replaced) and, somewhere else on the test, also asked whether the pair /r ɛlst/-/r ɛlps/ did (where only the post-vocalic obstruents were replaced, leaving the sonorant - in this case /l/ - 'stuck to the vowel.') Using a slightly modified form of the d' statistic from signal detection theory, the relative tendency of the various sonorant consonants to adhere to vowels was then calculated, taking into account not only correct HITS (involving the nominally correct pattern, where all sonorants were treated as part of consonantal clusters) and MISSES (where such nominally correct substitutions were rejected), but also CORRECT REJECTIONS (where all but the nominally correct substitutions were rejected) and FALSE ALARMS (where nominally incorrect pairs were accepted, i.e., pairs that kept the vowel and associated sonorant stuck together). On the basis of a large body of experimental data for such a task, the following differential tendency was observed, adding the categories G (for the English glides /w,y/) and R (for English /r/) to the ones already discussed, and where data for O came from reinforcement items from the training session:[2] **G > R > L > N > O.** (Other tasks, such as onset deletion and vowel substitution, showed a similar tendency in this study, though the absolute differences were not in all cases statistically significant.)[3]

4. TESTS FOR SYLLABLE BOUNDARIES

Similar effects can also be extracted from the more recent work done on the problem of syllable boundaries by Treiman & Danis (T&D). Relying primarily on a production task of syllable inversion, T&D [10] investigated the problem of where common English disyllabic words were broken that contained only a single intervocalic consonant. Their results (largely confirmed by an associated forced-choice written task) showed that the position of the break depended on a number of factors, including (1) the quality (tense vs. lax) of the vowel in the first syllable, (2) the position

of stress (on first vowel or second vowel), (3) the way the medial consonant was spelled (i.e., with one letter, as in *melon*, or two, as in *gallon*) and (4), most interesting from our current standpoint, the quality of the consonant itself. Most notably, in the case of consonants with singlet spellings in words with initial stress on lax vowels (such as *melon*, *lemon* and *seven*), L showed the strongest tendency to be treated as part of the first syllable, and O the weakest, with N, once again, taking the intermediate position.[4]

Finally, in the attempt to extend this work to typologically diverse languages (see [11] in these proceedings for some initial results for Korean), Derwing sought to develop a simpler technique for syllable division that could be performed by subjects who were not necessarily literate, as well as administered to large groups of subjects simultaneously. The result was a so-called 'pause-break' task, in which subjects were asked to choose which of two or three alternative 'breakings' of a word sounded the 'most natural.' In the case of the English word *melon*, for example, the following three alternatives were offered (where ... indicates the location of the pause):
(a) /m ɛ...l ə ɹ/ (where /l/ is treated as the onset of the second syllable),
(b) /m ɛl... ə ɹ/ (where /l/ is the coda of the first syllable), or (c) /m ɛl...l ə ɹ/ (where /l/ is ambisyllabic). In the English pilot study, 95 speakers were presented with a word-set much like T&D's.[5] All four of T&D's main effects re-emerged, as well as a new factor of the morpheme division. Of chief interest to us here, however, is the now-familiar four-way distinction among R, L, N and O, which the table below displays for words like *herald*, *melon*, *lemon* and *seven*:

Sound Class	S1/Co	S2/On	Amb[6]
R	**.76**	.07	.18
L	**.62**	.19	.19
N	**.52**	.37	.12
O	.29	**.61**	.09

Once again we see the same familiar differential tendency towards 'vowel-stickiness' as before, in this case realized as a tendency for singlet-spelled consonants to stick together with a lax, stressed vowel as part of the first syllable of a word: **R > L > N > O.**

5. CONCLUSIONS

In sum, the 'vowel-stickiness' phenomenon now seems to be quite firmly established, as it has been shown to be manifested in a consistent way across three different methodologies originally conceived for quite different purposes: in productive word-blends, in substitution-pattern judgments, and now in both production and judgment tasks for syllable divisions. Two major questions now remain: (1) to ascertain whether the same pattern holds for other, typologically diverse languages; and, if so, (2) to find a satisfactory explanation for the phenomenon. (It is worthy of note that a tantalizingly similar ordering - variously referred to as the 'sonority' or, inversely, 'strength' hierarchy - has emerged from descriptive linguistics, based on the investigation of both synchronic and diachronic data.) Extensive cross-linguistic work is now underway in our laboratory in search of an answer to question (1), combined with theoretical modeling and testing efforts suitable to satisfy the needs of (2).[7]

NOTES

[1]The research reported here was supported in part by a research grant from the Social Sciences and Humanites Research Council of Canada (No. 410-88-0266), awarded to the first author.

[2]Note that L here now refers to English /l/ alone, as the distinct term R has been applied to /ɹ/.

[3]Using this same technique, the L > N portion of this hierarchy was re-confirmed in a later study [7] for post-vocalic sonorants, which also demonstrated the effect on 'stickiness' of both vowel and consonant quality, much along the lines suggested by Selkirk [9].

[4]In this study, both English /l/ and /ɹ/ were again treated as members of the same class ('liquids') and analyzed together.

[5]Except that the list was modified to include separate samples for both /l/ and /ɹ/, which, as already noted, were collapsed in T&D and treated together as 'liquids.' A few new words (*oily* vs. *doily, sailor* vs. *molar, foaming* vs. *moment*, etc.) were also added to check on the effect of morpheme boundaries.

[6]S1/Co = coda of first syllable, S1/On = onset of second syllable, Amb = both (ambisyllabic). Response proportions are shown for each, with majority responses in boldface.

[7]These include the construction of Markovian and neural network models of our own design, as well as alternatives proposed elsewhere (e.g., [12]).

REFERENCES

[1] DERWING, B.L., T.M. NEAREY & M.L. DOW (1987), "On the structure of the vowel nucleus: experimental evidence," presented at the Annual Meeting of the Linguistic Society of America, San Francisco.

[2] TREIMAN, R. (1983), "The structure of spoken syllables: evidence from novel word games," *Cognition 15*, 49-74.

[3] TREIMAN, R. (1988), "Distributional constraints and syllable structure in English," *Journal of Phonetics 16*, 221-229.

[4] DOW, M.L. (1987), "*On the psychological reality of sub-syllabic units*," Ph.D. dissertation, University of Alberta, Edmonton.

[5] DOW, M.L. & B.L. DERWING (1989), "Experimental evidence for syllable-internal structure," in R. Corrigan, F. Eckman & M. Noonan (Eds.), "*Linguistic categorization*," Amsterdam: John Benjamins, 81-92.

[6] TREIMAN, R. (1984), "On the status of final consonant clusters in English syllables," *Journal of Verbal Learning and Verbal Behavior 23*, 343-356.

[7] DERWING, B.L. & T.M. NEAREY (1990), "Real-time effects of some intrasyllabic collocational constraints in English," in "*Proceedings of the 1990 International Conference on Spoken Language Processing (Vol. 2)*," Kobe, Japan, 941-943.

[8] DERWING & NEAREY (forthcoming), "On the structure of the vowel nucleus: a substitution-pattern identification task".

[9] SELKIRK, E.O. (1982), "The syllable," in H. Van der Hulst & N. Smith (Eds.), "*The structure of phonological representations (Part II)*," Dordrecht: Foris, 337-383.

[10] TREIMAN, R. & C. DANIS (1988), "Syllabification of invervocalic consonants," *Journal of Memory, and Cognition 27*, 87-104.

[11] DERWING, B.L., S.W. CHO & H.S. WANG (1991), "A cross-linguistic experimental investigation of syllable structure: some preliminary results," in this volume.

[12] GOLDSMITH, J. & G. LARSEN (1990), "Local modeling and syllabification," presented at the CLS parasession on the syllable in phonetics and phonology, University of Chicago.

MARGINAL VOWELS IN HUNGARIAN

P. Siptár

Hungarian Academy of Sciences, Budapest, Hungary.

ABSTRACT

This paper suggests a variety of ways in which the number of categories needed for characterizing the surface phonetic vowel inventory of Hungarian (Table 1) can be reduced until eventually a minimal underlying system (Table 4) is reached. Four 'marginal vowels' (parenthesized in Table 1) are discussed in particular. [e], [ɛ:], and [ɔ:] are argued not to be necessary in the underlying system; on the other hand, nonround /ɑ/ turns out to be one of the most loaded Hungarian vowels: one that surfaces as [ɔ] in the regular case, due to an independently motivated rule of the language.

1. INTRODUCTION

A surface phonetic classification of the Hungarian vowel system is shown in Table 1. The system has fourteen 'full members' plus four additional candidates (parenthesized) whose phonological status will be considered in this paper (Section 2). The classification appearing in Table 1 involves five heights, three points of articulation along the sagittal axis, plus the rounded/unrounded distinction. Obviously, a number of phonetic details can be filtered out of this representation on grounds of predictability. 'Height 1' is conventionally labelled 'High'; the rest of the heights might be called Upper Mid, Lower Mid, Upper Low, and Lower Low, respectively. The difference between Upper Mid and Lower Mid might be taken to be a matter of Tense/Lax; but even that is predictable (redundant) on the basis of Long vs. Short (alternatively, VV vs. V in terms of timing slots). On the other hand, the two Lows may be simply taken to be the same height phonologically: the exact height of [(a) a:], as well as their centrality, is a matter of phonetic implementation since in the (morpho)phonological pattern of Hungarian [a:] behaves as a low back vowel (e.g. with respect to vowel harmony, long/short alternations, etc.). Hence, the simplified pattern in Table 2 emerges; this classification will serve as the general framework within which the phonological status of the four 'marginal vowels' will be discussed in Sections 2.1–2.3 below. In Section 3, some general conclusions will be drawn and further simplification of the system will be proposed.

Table 1

	FRONT				CENTRAL		BACK	
	UNROUNDED		ROUNDED		UNROUNDED		ROUNDED	
HEIGHT 1	i	i:	ü	ü:			u	u:
HEIGHT 2		e:		ö:				o:
HEIGHT 3	(e)		ő				o	
HEIGHT 4	ɛ	(ɛ)					ɔ	(ɔ:)
HEIGHT 5					(a)	a:		

Table 2

	[– back]				[+ back]			
	[– round]		[+ round]		[– round]		[+ round]	
[+ high, – low]	i	i:	ü	ü:			u	u:
[– high, – low]	(e)	e:	ö	ö:			o	o:
[– high, + low]	ɛ	(ɛ)			(a)	a:	ɔ	(ɔ:)

2. DATA AND DISCUSSION

2.1. Unrounded short [a]

This vowel appears on the surface (apart from regional dialects) in the following cases: (i) In nonfinal closed syllables it is the normal (colloquial) realization of /a:/ as in *általános* [altɔla:noš] 'general', *vásárváros* [va:šarva:roš] 'market town'; in certain phonetic contexts with vacillation (where the postlexical shortening rule concerned is optional / rate-dependent): [at:ɛkinthetö:] ~ [a:t:ɛkinthetö:] *áttekinthető* 'perspicuous'. (ii) Also with [a] ~ [a:] free variation in words like *spájz* 'larder', *Svájc* 'Switzerland', *Mozart* (here, however, 'free variation' means inter-speaker variability rather than intra-speaker vacillation). (iii) On the other hand, [a] ~ [ɔ] (inter-speaker) variation is found in words like *gavott* 'gavotte', *hardver* '(computer) hardware', *Csajkovszkij* 'Tchaikovsky', and in *halló* [halo:] 'hullo' as used in phone calls (where classical minimal pairs can also be found for both [ɔ] and [a:]: *haló* [ɔ] 'dying' vs. *halló* [a] 'hullo' vs. *háló* [a:] 'net').

The question, then, is what the phonological status of all these [a]'s should be. (From now on, I use the symbol /ɑ/ to refer to the underlying a-type – short back low – vowel with no roundness specification intended; the choice of symbol is motivated by considerations of clarity, i.e. I wanted a symbol that is distinct from both a and ɔ.) There are a number of convincing arguments to the effect that /ɑ/ behaves morphophonologically as a nonround vowel (cf. the length alternation /a:/ ~ /ɑ:/ and the vowel harmony alternation /ɛ/ ~ /ɑ/; in both cases an intermediate nonround low back vowel is derived that surfaces via an *a*→[ɔ] realization rule). Since the rounding of /ɑ/ is phonologically irrelevant (nondistinctive) and phonetically rather moderate as opposed to mid and especially high back vowels (though this does not weigh much in phonology), it is at least possible to claim that /ɑ/ is in general (i.e. not only in the alternating cases) underlyingly nonround. It was pointed out in Section 1 above that the centrality of [a:] and the fact that in terms of tongue height it is lower than [ɛ] or [ɔ] are just as redundant phonologically as the surface roundness of [ɔ] is. Hence, the /ɑ/ ~ /a:/ alternation will fit the rest of the pattern where alternants only differ in length (cf. 2.2. on /ɛ/ ~ /e:/).

Now if we accept this reasoning, the following can be said about the three groups of surface [a]'s exemplified above: (i) In addition to the morphophonological rule /a:/→/ɑ/ (*nyár* ~ *nyarat* 'summer' nom./acc.), followed by rounding adjustment /ɑ/→[ɔ], there is also a surface (postlexical) shortening rule that will of course apply (much) later than rounding adjustment and whose output will therefore remain unrounded. (ii) For speakers who say [špajz] etc., underlying nonround /ɑ/ will be a (lexical) exception to rounding adjustment in these words; for other speakers, the lexical representation will be /špa:jz/ to which shortening or rounding adjustment is inapplicable. (iii) The word *halló* and other similar items (the exact range of which varies from speaker to speaker) are exceptional in that they will be (optionally or categorically) exempt from rounding adjustment /ɑ/→[ɔ]. Alternatively, in terms of underspecification theory, garden-variety /ɑ/ will be underlyingly unspecified for rounding whereas the vowel in *halló* etc., as well as *spájz* etc. for [a] speakers, will be specified as [–round]; rounding adjustment would then be a "fill-in rule" in that it cannot *change* feature specifications but only fill in blanks; the desired result then follows without recourse to any exception feature.

In sum: If these conjectures are on the right track, nonround /ɑ/ is not marginal: in fact, it is one of the most loaded members of the Hungarian vowel system; what *is* marginal is the range of cases where it surfaces unaltered.

2.2. Short mid [e]

The case of this vowel is in some respects similar to that of [a], in others it is quite different. On the surface it appears with regional/cultural restrictions (i.e. in certain regional varieties): its use is much wider than that of – dialectal! – [a], but does not include standard Hungarian in the strict sense. (The postlexical shortening of /e:/ as in the second syllable of *keménység* 'hardness' results in a vowel tenser than [e], just like that of /o:/ and /ö:/; that is, as was pointed out in Section 1 above, [e] and [e:], [o] and [o:], [ö] and [ö:] differ not only in length but also in tenseness.)

If, in standard Budapest Hungarian, [e] does not appear even to the limited extent that [a] does, why do we mention it here? The reason is that Hungarian morphophonology works *as if* there was an /e/ in the system. The nonround member of the alternation *o* ~ *ő* ~ *e* (at the level of the immediate output of the rule) is mid, whereas the front member of the alterna-

tion *á* ~ *é* and the long member of *e* ~ *é* (*kefe* ~ *kefét* 'brush' nom./acc.) are low (at the same level), hence an *e*/*é*-adjustment (redundancy) rule is needed to convert such derived *e*'s into a low, and derived *é*'s into a mid (and tense) vowel. (Alternatively, Structure Preservation might produce the same effects without an explicit adjustment rule.) These facts, however, are still not sufficient to justify an underlying /e/, unless the ambiguous behaviour of [ɛ] in vowel harmony could be explained by positing mid /e/ along with low /ɛ/. In particular, Hungarian vowels fall into three harmonic classes as follows: back-harmonic /a: ɑ o: o u: u/, front-harmonic /ö: õ ü: ü/, and neutral /i: i e:/. Surface [ɛ] is ambiguous in that it behaves sometimes as front harmonic and sometimes as neutral (see [2] for details). It might be a good idea to recognize /e/ as a neutral vowel and /ɛ/ as a front-harmonic one. In fact, all five-vowel solutions implicitly involve this idea. Abondolo ([1]:29ff), for instance, has the following system:

	I	E	A	O	U
mid	−	+	−	+	−
back	−	−	+	+	+
rounded	−	−	−	+	+

plus a (morpheme-sized) 'front prosody'. Cf. also van der Hulst's similar solution couched in autosegmental terms ([2]:279ff). Considerations of space prevent us from exploring the full implications of this type of solution; we will rest content with observing that, although certain seemingly irregular classes of words (e.g. back-harmonic monosyllabic stems containing a neutral vowel) can be accounted for nicely in terms of such systems, positing two underlying sources for surface [ɛ] raises more problems than it solves. Hence, we will assume that the system has only one nonhigh front unrounded short vowel. For typographical convenience, we will henceforth refer to this item as /e/ – whether it is underlyingly mid (hence, exactly parallel to its long cognate /e:/) or low (hence, identical with its surface representation [ɛ]) will turn out to be irrelevant (see Section 3 below).

2.3. Long low [ɔ:] and [ɛ:]

Along with the surface shortening rule mentioned in the previous sections, there are surface lengthening rules as well. 'Pause-substituting' (i.e. hesitational or phrase-final) and emphatic lengthenings do not convert the short vowels into their long counterparts; rather, they either leave vowel quality unaffected or modify it in another direction (e.g. emphatic *ooolyan* 'so much' with an *o* opener than usual, whereas long /o:/ is closer/tenser than /o/). Other types of surface lengthening will produce [i:] out of /i/, [o:] out of /o/, etc. For instance, names of letters and sounds are usually quoted in a lengthened version as in *Ezt rövid* [i:]-*vel kell írni* 'This is spelt with short I', *A magyarban nincs rövid* [o:]-*ra végződő szó* 'There are no word-final short O's in Hungarian', etc. However, such (surface) lengthening of [ɔ] and [ɛ] will produce [ɔ:] and [ɛ:], rather than [a:] and [e:]. (This can be explained simply by assuming that such lengthening takes place at a point where the adjustment rules mentioned above have already applied.) For instance, the length of the initial vowels in *erre* [ɛ:rɛ] 'this way' and *arra* [ɔ:rɔ] 'that way' can be derived by compensatory lengthening although, on a strictly taxonomical view, these should be independent (micro)phonemes, cf. the minimal pairs *erre* 'this way'/*ere* 'his vein': [ɛ:rɛ/ɛrɛ] and *arra* 'that way'/*ara* 'bride': [ɔ:rɔ/ɔrɔ].

The names of the letters/sounds *a* and *e* exhibit a curiously intricate pattern. The basic case can be observed in contexts like *nagy* [ɔ:]-*val írjuk* 'it is spelt with capital A', *kétféle* [ɛ:]-*vel beszél* 'he distinguishes two types of E in his speech', etc. (Minimal pairs can be found again: *a-féle* [ɔ:fe:lɛ] 'of the type A' vs. *afféle* [ɔfe:lɛ] 'sort of', *e-be* [ɛ:bɛ] 'into E' vs. *ebe* [ɛbɛ] 'his dog', *a-hoz* [ɔ:hoz] 'to A' vs. *ahhoz* [ɔhoz] 'to that', *e-szer* [ɛ:sɛr] 'E times' vs. *eszer* [ɛsɛr] 'Social-Revolutionary', and so on.) On the other hand, the musical notes A and E are called [a:] and [e:], and the word *ábécé* [a:be:tse:] 'alphabet' itself makes it likely that the name of the letter A used to be pronounced [a:] (latinate influence?). Letters used for identification exhibit an even more chaotic pattern: the bus *7/a* is [he:t ɔ:] but a school class *7/a* is [he:t a:] (although *7/e* is [ɛ:] rather than [e:]); *A épület* 'building A' can be either [ɔ:] or [a:] but *E épület* can only be [ɛ:]; in geometry, *a pont* 'point A' is either [a:] or [ɔ:] but *e pont* is always [ɛ:], etc. Abbreviations, if they are pronounced as a sequence of letters, contain [a:] and [e:] if A or E is initial (*AB* 'abortion committee', *EKG* 'electrocardiogram') but [ɔ:] and [ɛ:] if final (*MTA* 'Hungarian Academy of Sciences', *BSE* 'Budapest Sports Club'). Those abbreviations that are read out as words (*USA* 'United States', *ELTE* 'Eötvös Loránd University') behave as normal words do: they end in short [ɔ]/[ɛ] which regularly undergoes Low Vowel Lengthening ([uša:bɔn] 'in the US', [ɛlte:rõl] 'from ELTE'), hence they are uninteresting for our present purposes.

What is much more interesting though is that [ɔ:] and [ɛ:] *never* undergo LVL: [ɛmte:ɔ:vɔl], not [ɛmte:a:vɔl] if the nominative is [ɛmte:ɔ:]. (See also the examples listed earlier in this paragraph.)

Now, are [ɔ:] and [ɛ:] to be regarded as independent (micro)phonemes or as rule-generated realizations of [ɔ]/[ɛ]? Cases like *arra* can be explained by (lexically conditioned) compensatory lengthening, despite the (surface) minimal pairs. But if the name of the letter E is underlyingly a short /e/, how can its surface lengthening block the application of a morphophonological rule like LVL (cf. *e-nek* [ɛ:nɛk] 'for E' ≠ *ének* [e:nɛk] 'song')? Such bleeding interaction undoubtedly runs counter to all current assumptions concerning the way phonological systems are organized. However, the phenomena discussed in this section are both peripheral and variable: therefore, the alternative approach (positing underlying /ɔ:/, /ɛ:/) will be discarded here and it will be assumed that some exception device takes care of the offending cases.

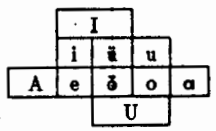
Table 3

			Back	
High	i	ü	−	u
	e	ö	ɑ	o
		Round		Round

3. CONCLUSION

It was argued above that (i) both [ɔ] and [a] go back to underlying /ɑ/ whose roundness need not be specified; (ii) all instances of surface [ɛ] should be derived from a single underlying segment, /e/, whose lowness need not be specified; and (iii) most, if not all, instances of [ɔ:] and [ɛ:] can be accounted for as due to surface lengthening of [ɔ:] (from /ɑ/) and [ɛ:] (from /e/), respectively. All these observations add up to an even more simplified underlying vowel system, given in Table 3. Notice that the short:long opposition is assumed to be encoded in V:VV on the skeletal tier (a move that would not be possible if the quality differences between [ɔ] and [a:], respectively [ɛ:] and [e:], were regarded as underlyingly valid distinctions); notice further that [low] is made superfluous as a classificatory feature (of course, it continues to figure as a phonetic feature that the rules of phonetic implementation need to refer to). Finally, notice that Table 3 uses the unary features High, Back, and Round rather than the binary features of Table 2; hence, /e/ is neither mid nor low – it is simply nonhigh; /ɑ/ is neither rounded nor unrounded – it is simply not characterized by the feature Round; and finally, neutral vowels are not necessarily defined as front; they simply share the property of not being characterized by the feature Back with the front-harmonic (front rounded) vowels. An alternative possibility (and one more in keeping with most current phonological theories) is recognising the three unary features (or particles, or elements) A I U for 'aperture', 'palatality', and 'labiality', respectively; this gives us the vowel system shown in Table 4. Although this version loses some of the advantges (listed above) of that in Table 3, it is nevertheless superior in one respect: unlike the system in Table 3, it does not leave any existing vowel of Hungarian completely unspecified (leaving the possibility of empty V for epenthetic vowels that acquire all their properties from the environment) and conversely, it does not define a vowel (cf. the high back unrounded slot in Table 3) that is nonexistent in Hungarian.

Table 4



4. REFERENCES

[1] ABONDOLO, D.M. (1988), "*Hungarian inflectional morphology*", Budapest: Akadémiai Kiadó.

[2] HULST, H. van der (1985), "Vowel harmony in Hungarian: a comparison of segmental and autosegmental analyses", in: H. HULST–N. SMITH (eds.), "*Advances in nonlinear phonology*", Dordrecht: Foris, 267–303.

[3] SIPTÁR, P. (1990), "Issues in Hungarian phonology", in: B. HURCH (ed.), "*Natural phonology*", Workshop at the Annual Meeting of the Societas Linguistica Europaea, Bern, 113–116.

LOCAL PROMINENCE OF ACOUSTIC AND PSYCHOACOUSTIC FUNCTIONS AND PERCEIVED STRESS IN FRENCH

P. Mertens

K.U.Leuven, Linguistics Department, Leuven, Belgium.

Abstract

Syllable duration, pitch, loudness, pause length, pitch change, and local difference values for the first 3 parameters, were studied for their ability to predict perceived stress as measured in a listening task. The best cues were duration increase relative to preceding and following syllable(s), followed by nucleus duration.

1. Introduction

Syllabic stress is a linguistic attribute realized in various ways, with or without prominence. It can not be observed directly. A measure of perceived prominence has to be established in order to classify the syllables. A brief review of terminology will clarify this point.

(1) A syllable is prominent when it stands out from its context due to a local difference for some prosodic parameter. Prominence is continuous (not categorical) and contributions of multiple parameters can interact.

(2) Stress is an abstract linguistic category, which can be realized by several types of prominence, in a way which is language-specific.

(a) In French, an intra-syllabic pitch glide of a given interval suffices to signal stress. Prominence by duration or loudness will be functionally redundant although very common.

(b) For static syllables prominence will results from an inter-syllabic change of a parameter.

(c) Finally, stress can result from tone level itself, on the basis of tone distribution [3,6].

(3) Word stress (lexical stress) indicates the syllable in a word which can receive stress.

(4) French has two stress types: final (word stress position) and initial stress (emphatic), with a different distribution.

In a listening task, the stress judgment will be based on a mixture of heterogeneous factors: acoustic, structural, lexical. Subjects may focus on an isolated factor, or on many; they find it very difficult to separately rate prosodic parameters. The test can show how untrained subjects judge stress, and whether they agree. Given the continuous nature of prominence, a stress score, the number of listeners that perceived a syllable as stressed [1,2], allows for a classification in min. 3 categories: stressed, unstressed, ambiguous.

Because of space limitations, previous studies on stress perception and stress cues can not be reviewed here.

2. Method.

Six extracts (277 syll.) were selected from a corpus [3] in such a way that the test contained at least 2 occurences of each stressed tone. A male and a female speaker each provided 3 extracts. The mean length of 46 syll/test was suggested by [2] where it was found that the proportion of syllables judged stressed decreases as the length of the carrier sentence increases. For lengths above 40 syll. the ratings are simular to those for continuous speech. The passages were very different in

Table 1. Agreement between 20 raters, and prosodic complexity, (judged by phonetician) for tests 1 to 6.

TEST(S)	1	2	3	4	5	6	1-6
P(A)	.7667	.7705	.8333	.6939	.8590	.7756	.7867
P(E)	.6704	.6691	.7086	.6194	.6195	.6009	.6558
Kappa	.2922	.3064	.4281	.1958	.5429	.4379	.3804
Complexity	medium	medium	low	high	low	medium	

terms of prosodic complexity (table 1), which can be defined in terms of (1) rate of speech (without pauses), (2) proportion of stressed syllables, (3) of emphatic stresses, (4) of pauses, (5) of glides, and (6) rhythmic structure.

2.1. Perceptual experiment.

The 20 untrained subjects heard each passage once (with 25s silence) and 6 times (with 6s intervals) during which they had to indicate the stressed syllables on the test sheet.

Each syllable was judged either stressed or unstressed; so, it was assigned to 1 out of 2 categories. The nominal scale calls for a non-parametric test: the kappa statistic [9] was used. P(A), the proportion of times that the raters agree, and P(E), the proportion of estimated chance agreement, are determined. The kappa coefficient is the ratio of P(A) to the maximum proportion of times that raters could agree, both corrected for chance agreement. A kappa 1 indicates complete agreement, a 0 indicates no agreement other than chance. Since only 2 categories are used here, chance agreement is high, and kappa rather low (table 2). The pooled data (P(A)=.7867, kappa=.38) show a moderate agreement among the listeners, although significantly different from 0. The relation with prosodic complexity is obvious.

2.2. Acoustic measurements.

For each syllable, 5 primary attributes are obtained, using an interactive analysis program [3,5]: nucleus DURATION, PITCH peak, LOUDNESS peak, intra-syllabic GLIDE, PAUSE duration. The segmentation into syllabic nuclei [4] provides boundaries necessary for the parameter extraction and pitch contour stylization. PITCH is the peak and GLIDE the interval of the stylized contour, positive or negative according to slope. Pitch values are expressed in semitones (ST): the melodic (in mel) and harmonic (in ST) scales are almost identical in the F0-range of speech [10]. The results were hand-corrected where necessary.

The measurement of LOUDNESS [10,8] (in soneG) accounts for frequency dependence, critical bands, frequency masking, level, but ignores the effect of stimulus duration. Level values (dB SPL) for each critical band were obtained from the power spectrum (512pt FFT, 40ms), by summation of the components in the band range, and dB-conversion.

Prominence estimates were calculated for duration, pitch and loudness. Prominence is defined as the difference between the parameter value for a syllable and the parameter mean of the context, either left (L) or right (R), with length 1 and 2 syll., giving 4 relative values: resp. DL1, DL2, DR1 and DR2 for duration, PL1, PL2, PR1, PR2 for pitch, and LL1, LL2, LR1 and LR2 for loudness. This allows for a continuous scaling of prominence. A similar measure combining left and right contexts with length 1 was used in [7].

3. Results

Scatter diagrams were made for the 17 attributes, with stress SCORE as the dependent variable. Some results were predictable: PITCH varies randomly with stress score

Table 2. Correlation between stress SCORE and parameters (above line) prominence measure (below line), for the pooled data.

DURATION	PITCH	LOUDNESS	GLIDE	PAUSE
.477	.203	.299	.070	.296
DL1 DL2 DR1 DR2	PL1 PL2 PR1 PR2	LL1 LL2 LR1 LR2		
.49 .49 .41 .41	.47 .44 .45 .48	.31 .30 .43 .36		

Table 3. Mean values for 7 variables cross-tabulated with ranges for SCORE. N is the number of syllables in a group.

SCORE	N	DUR	DL1	DR1	PL1	PR1	LL1	LR1
0-20	277	88	0	0	0.1	0.1	-0.2	0.2
0- 3	194	72	-26	-19	-1.2	-0.9	-1.2	-1.0
4-11	50	109	45	25	1.6	1.7	2.1	1.6
12-20	33	156	80	77	3.6	4.8	2.5	5.4

(because of speaker's range, declination line, etc.) and so does LOUDNESS. There are too few cases of glides and pauses to find a relation with SCORE.

Although no clear linear relation was found, Pearson correlation coefficient r was used to estimate the amount of information that could be gained from each variable (table 2). r varies considerably from one passage to another: for DURATION, from .63 to .18. Test 4 (with high complexity) gives very poor correlation for all attributes and is to a large extent responsible for the low r in the pooled data.

DURATION is the only primary parameter with relatively high r: this can be explained by minimal syllabic duration, small variability for unstressed syllables and large for the stressed.

The best prominence estimates are DL1 and DL2, indicating that syllables with high SCOREs are generally longer than the preceding one(s). DL2 and DR2 give results close to DL1 and DR1. LOUDNESS, LL1 and especially LR1 score quite good (r=.5) in some tests, but not on the average.

Depending on the method used and the number of variables taken into account, multiple regression gives a correlation of .60 to .88 with the stress SCOREs.

Stress score can be used to classify the syllables in 3 groups: not prominent, ambiguous, and prominent (table 3), showing clear differences between groups. The choice of the ranges depended on the number of elements in each group.

Labeling according to the transcription by a phonetician gives a further classification (table 4). Group means show that the stressed are twice as long as unstressed; they are prominent by duration (DL1,DR1) and, in the case of low stressed, also by loudness (LL1,LR1). PL1, PR1 and GLIDE reflect the tones used (H,L,HL,LH,L-,H+). The values for emphatic stress are very close to those for unstressed syllables. The parameters do not reflect the evident phonatory effort of emphatic stress.

Predictions by the intonation model are observed in the data: (1) syllables with extra-low tone (L-) can be short and weak because their stressed status is already indicated by tone level, (2) glides (HL,LH) lack loudness prominence because stress is already signaled by the glide.

4. Conclusion

A listening task provided ratings of perceptual prominence for 277 syllables. The relative agreement between the raters indicates the perceptual reality of prominence. The importance of acoustic parameters as well as of four prominence measures were studied. The stress scores by the listeners are best predicted by durational prominence relative to the preceding 1 or 2 syllables, and by syllabic duration itself. When the transcription of intonation by a phonetician is used for syllable classification, the same order of importance for the studied parameters is found. Predictions by the intonation model on the relative importance of individual prosodic parameters depending on the tone used, are confirmed by the data.

Table 4. Parameter means for syllables classified according to transcription by phonetician. The mean score by the untrained listeners is shown under SCORE. N is the number of elements in a category.

	N	DUR	DL1	DR1	PL1	PR1	LL1	LR1	GLIDE	SCORE
EMPHATIC	15	72	-19	-12	2.8	2.5	-1.3	1.4	0.0	6.3
STRESSED	74	137	70	65	2.6	2.9	1.6	2.7	0.0	10.6
H	28	135	64	53	4.3	4.7	1.4	3.5	0.3	11.3
HL	3	186	120	115	5.0	2.6	-0.3	-1.3	-4.3	11.0
L	25	139	74	77	1.8	2.2	3.7	4.3	-0.6	9.1
LH	4	228	168	133	4.2	3.7	2.0	0.7	7.2	13.2
H+	3	103	56	56	10.6	9.6	4.0	2.6	0.0	17.3
L-	11	100	32	33	-3.0	-1.9	-2.7	-0.6	-1.2	9.4
UNSTRESSED	188	71	-26	-24	-1.4	-1.1	-0.8	-0.9	-0.1	1.8
h	13	61	1	-46	-0.3	1.6	0.3	1.0	-1.0	4.6
l	168	71	-29	-23	-1.5	-1.4	-1.0	-1.0	0.0	1.6
l-	7	79	-20	-7	-1.1	0.1	-0.2	-1.2	0.4	2.4
POOLED	277	88	0	0	0	0.1	-0.2	0.2	0	4.4

5. References

[1] Allen, G.D. (1972a) The location of rhythmic stress beats in English: an experimental study (Part I.), Lang. & Speech 15(1), 72-100.

[2] McDowall, J.J. (1974) The reliability of ratings by linguistically untrained subjects in response to stress in speech, J. Psycholing. Res. 3, 247-259

[3] Mertens, P. (1987a) L'intonation du français. De la description linguistique à la reconnaissance automatique. Unpubl. Ph.D.

[4] Mertens, P. (1987b) Automatic segmentation of speech into syllables, Proc. Eur. Conf. on Speech Techn., II, 9-12.

[5] Mertens, P. (1989) Automatic recognition of intonation in French and Dutch, Eurospeech 89, I, 46-50

[6] Mertens, P. (1990) Chapitre IV. "Intonation", in Blanche-Benveniste, C. et al. (1990), Le français parlé, Paris: Ed. du CNRS, 157-176.

[7] Gaitenby, J.H. & Mermelstein, P. (1977) Acoustic correlates of perceived prominence in unknown utterances, SRSR-49, 201-216.

[8] Paulus, E. & Zwicker, E. (1972) Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln, Acustica 27(5), 253-266.

[9] Siegel, S. & Castellan, N.J. (1988) Nonparametric Statistics, NY: McGraw-Hill.

[10] Zwicker, E. & Feldtkeller, R. (1981) Psychoacoustique. L'oreille, recepteur d'information, Paris: Masson.

WORD STRESS IN GEORGIAN

P. McCoy

University of California, Los Angeles, USA

ABSTRACT

Word stress in Modern Georgian, the language spoken in the Soviet Republic of Georgia in the USSR, is known to be weak in nature, in fact is not certain that there is stress in Georgian. An experiment was conducted to test words in isolation, in phrases and in complete texts to see if there were any common denominators. The parameters examined here were pitch and duration. For each phonological word, FO measurements found a single peak for the whole word; correlations between words in text and in isolation were fairly consistent, though not uniform. Greater duration fell either in the syllable with the FO peak or in the initial syllable. The results indicate that although stress in Georgian is weak, it is clearly a word level phenomenon.

1. INTRODUCTION

One of the more confusing questions for the student of Georgian is the placement of stress. Although this may be elementary for languages with fixed stress on some syllable of the Phonological word, for example Czech, where it is always on the first syllable, or for languages with mobile stress where for the most par it has to be learned, the question of stress in Georgian is one that is almost avoided. Part of the reason for this lies in the fact that it is not certain that there is stress in Georgian. Or if so, there is no consensus as to its location. In all the varying opinions about Georgian there is agreement on one point, that stress is weakly dynamic and has melodic tone. Because it is weak, it tends to defy both description and analysis. This paper takes as its point of departure two questions of import for stress in Georgian: (1) Is there a word level stress (or is it phrase or sentence level)? (2) If there is, then how is it implemented in Georgian?

The structure of the paper will be as follows. I will first review literature on the subject as it is useful to be fully aware of the variety of views there are available in stress in Georgian. Having done this, I will proceed to look at some data from a study examining minimal pairs of words representing two environments -- within the flow of continuous speech, here a read text, and the same words read in isolation.

2. LITERATURE

Starting with the most impressionistic, we have two descriptions: 1) "Die Betonung gleicht dem geglatteten Meer nach dem Sturm." [1]; 2) "...wie murmelndes Wasser lauft die georgische Rede hin." [2]. These would imply that there is a significant lack of perceptual cues with which to identify stress. This may well be true at the most impressionistic level where not much attention is focussed on the physical aspects of perceptual cues, but at a slightly more conscious level, there do seem to be enough cues to generate varying opinions on the nature of stress in Georgian.

As regards duration, a common indicator for stress in language, some sources say that vowels are of equal duration throughout the word, irrespective of the length of the word. These sources seem to focus more on the melodic nature of Georgian. Others however indicate that in addition to the melodic structure, duration may have a place in determining the place of stress. Tschenkeli [3] in his grammar indicates a

	I. Stress Placement: pnul	apnul	init	II. Primary vs. Secondary Stress: pnul	apnul	init	III. Intonation Stronger: yes	no	no comment
Vogt 1939			✓	‖		|	✓		
Tschenkeli 1958			✓		‖	|	✓		
Cikobava 1967		✓		—	No comment —		✓		
Marr and Briere 1931	✓✗✓			‖	*	‖	|		✓
Rudenko 1940	✓✗✓			‖	*	‖	|		✓
Robins and Waterson 1952	✓✗✓				|	‖		✓	
Zgenti 1964		✓		—	No comment —		✓		
Shimomiya 1978			✓		‖	|			✓

function of duration and Japaridze, a Georgian phonetician who has done work on perception, comments on the perception of Czech by Georgians. He proposes an element of duration as Georgians hear Czech (a language which has constant phonetic stress on the initial syllable, phonemically long and short vowels) as stressed in syllables with long vowels.

Table 1 [3]-[10] gives a summary of Georgian stress as described by various views proposed in grammars and articles. What is interesting is that in addition to the differences expressed among the works, there also seems to be a lack clarity within a given description.

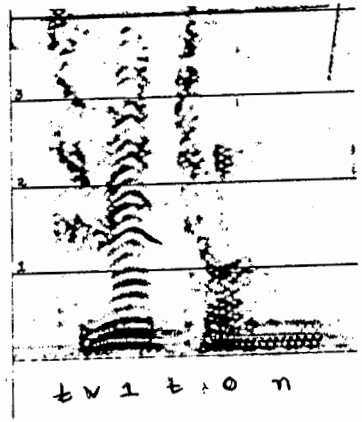
3. PROCEDURE

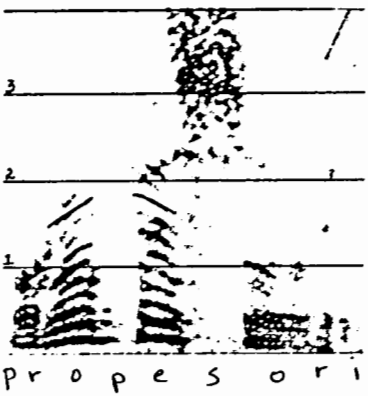
The speaker was male speaker of the literary norm as judged by colleagues at Tbilisi State University. He pronounced the sets of words in isolation, and in paragraphs, presented to him in a random order. A practice session was conducted in order that he be familiar with the words and his task. The speaker was cautioned to read at a set pace and to observe a fixed distance form the microphone. Each word was read thrice in each environment for a total of six tokens per word across environments.

Recordings were made in a soundbooth using a Marantz superscope cassette PDM 350 and a superscope condenser microphone. Broad and narrow band spectrograms were made of the tokens on a Kay Digital spectrograph. Fundamental frequency F0 was measured by tracing the 10th harmonic in the narrow band spectrograms. Duration measurements were taken form the broad band spectrograms.

4. RESULTS

Correlations of duration and pitch for words in isolation and in a text were fairly consistent, though not uniform. The greater duration measurements fell either in the syllable with the F0 peak or in the initial syllable. There was only one peak in a word and the peak, taking the duration of any word as a whole, seemed to be a third of the way into the word from the onset of the word.





From these results one can infer that stress, though weak is a word level phonomenon. Further, the acoustic tiggers for stress would seem to be a combination of duration and rising F0.

5. REFERENCES

[1] SCHUCHARDT, H. (1895), "Uber das Georgische", Wien: Selbstverlag des Verfassers (from Shimomiya, p.15).

[2] DIRR, A. (1928), "Einfuhrung in das Studium der kaukasischen Sprachen", Leipzig.

[3] TSCHENKELI, K. (1958), "Einfuhrung in die georgische Sprache", Bd. 1, Zurich.

[4] SHIMOMIYA, T. (1978), "Zur Typologie des Georgischen (Vergleichen mit dem Indogermanischen). Mit einem Exkurs zur Sprachbundtheorie", Gakushuin-Forschungsreihe 4., Gakushuin.

[5] CIKOBAVA, A. (1967), "Uzlovye voprosy istoricheskoe fonetiki iberijsko-kavkazskix jazykov", Tezisy dokladov, str.3-5, Tbisili.

[6] VOGT, H. (1939), "Grammaire de la langue Georgienne" Oslo.

[7] RUDENKO, B. (1940), "Grammatika gruzinskogo jazyka", Moscow, Leningrad.

[8] MARR, N. and BRIERE (1931), "La langue Georgienne", Paris.

[9] ROBINS, R. and N. WATERSON (1952), "Notes on the phonetics of the Georgian word", BSOAS, XIV, 55-72.

[10] ZHGENTI, S. (1964), "Rhythmical stress and intonation structure", Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung, 2-4, 357-368.

STRESS CLASH AVOIDANCE IN DUTCH: INVERSION OF STRESS PATTERN IN COMPLEX NOUNS?

Vincent J. van Heuven

Dept. Linguistics/Phonetics Laboratory
Leyden University, The Netherlands

ABSTRACT

We tested the phonetic basis of a recent claim made by metrical phonologists that the stress pattern of di-syllabic Dutch words with initial stress is inverted to final stress in order to avoid "stress clash" when such words are embedded (as the right-hand element) in a compound noun. In one experiment speakers produced crucial words both as simplex nouns and embedded in compounds; listeners were then asked where they perceived the stress in the targets after these had been excised from their spoken context. In a second experiment we presented diphone-synthesised versions of the crucial word types with systematically varied stress patterns; listeners had to rate the acceptability of the range of patterns in various rhythmic contexts. Results indicate that listeners perceive no stress shift in naturally produced word tokens, and that they always disallow versions of such words with inverted stress patterns.

1. INTRODUCTION

Compound adjectives in Dutch and English, such as red hot, have final stress when used predicatively: the 'poker is red 'hot (a single quote preceding a syllable marks strong stress). In attributive position, however, the final stress on these words is retracted: a 'red hot 'poker. If the stress had not been retracted, the result would have been two strong stresses abutting one another, a situation called "stress clash": a red 'hot 'poker. It is generally claimed that an immediate succession of two strong stresses on the same prosodic level violates a basic rhythmic principle underlying languages such as Dutch and English. These languages have a strong preference for a so called alternating stress pattern, i.e., a regular alternation of strong (stressed) and weak (unstressed) syllables. Native speakers of Dutch and English can easily be convinced that stress retraction occurs in compound adjectives. In the older literature we find numerous claims to the same effect [2,3,4]. Moreover, laboratory experiments have shown that the inversion of stress pattern in Dutch compound adjectives is clearly audible and has robust acoustic correlates [1].

In the past few years Dutch phonologists have studied another class of rhythmic stress adjustment phenomena, viz. the behaviour of stress patterns in polysyllabic nouns embedded in compounds (cf. [6,7]). When a word like 'harnas (armour), with lexical stress on the first syllable, is embedded in a compound noun, a situation of stress clash may arise, as in 'borst'harnas (breast armour). The authors concerned [6,7] claim that stress clash is resolved in these cases by inverting the stress pattern of the embedded word, yielding 'borsthar'nas, which would have the same stress pattern as 'scheepskom'pas (ship's compass), of which the embedded noun kom'pas (compass) has lexical stress in final position. Moreover, stress

pattern inversion is claimed to be applicable only when the embedded noun has initial stress on a closed syllable (a so called non-branching rhyme). Therefore no stress adjustment is said to occur when the lexically stressed first syllable of the embedded noun is open as in, e.g., 'premie (premium) - 'jaar'premie (annual premium).

Curiously enough, the older literature contains no allusions to this type of stress adjustment at all, and ever since the claims were made, phoneticians have expressed their doubts whether these are indeed cases of stress adjustment. In the present study we tried to settle this issue in a series of experiments.

2. EXPERIMENT I: PERCEIVED STRESS IN NATURAL SPEECH

2.1. Method

The basic stimulus material consisted of three types of di-syllabic Dutch nouns, each category filled with five exemplars:

1. initial stress on an open syllable ('premie-type)
2. initial stress on a closed syllable ('harnas-type)
3. final stress (kom'pas-type)

These 15 words were used as simplex words as well as embedded word-finally in tri-syllabic compound nouns, e.g., jaarpremie, borstharnas, and scheepskompas. The resulting set of 30 words were recorded four times onto audio tape by two male speakers of Dutch, who pronounced the target words twice in a fixed carrier phrase Heb jij een [TARGET] ontdekt? (Have you a [TARGET] discovered?) with accent on the target and two more times in Heb JIJ een [target] ontdekt? (with a contrastive accent on jij).

The 120 di-syllabic target word tokens were excised from their spoken contexts using a digital wave form editor, and presented twice (in different random orders) to 18 Dutch listeners. These were asked for each stimulus word to indicate along a scale from -5 to +5 what stress pattern they perceived. In this scale "0" meant that the stress levels of the two syllables were exactly equal. "-5" was to be chosen if the subject felt that the initial syllable was much less stressed than the final syllable. "+5" had to be responded when the subject perceived much more stress on the initial syllable than on the final syllable. Intermediate values stood for less extreme differences in the distribution of stress over the two syllables.

2.2. Results and conclusions

Table I contains the results.

Table I: Mean perceived stress distribution (see text) broken down by accentedness of target, type of word (simplex vs. embedded in compound), and lexical stress type (each mean is based on 360 judgments nominally).

target	accented		unaccented	
	simpl.	emb.	simpl.	emb.
'premie	3.7	1.6	1.9	0.9
'harnas	3.7	1.7	1.5	0.5
kom'pas	-2.5	0.8	-0.5	-0.4

The perceived stress distribution clearly differs for words with initial stress ('premie-type and 'harnas-type) and those with final stress (kom'pas-type), $F(3,4194)=957.1$ $p<<.001$. The difference between initial stress and final stress is larger for simplex words than for the same words incorporated in a compound, $F(1,4195)=55.8$, $p<.001$ (this corresponds to the difference between primary versus secondary stress on the word level). The stress patterns are perceived as more extreme in accented simplex words than elsewhere. Crucially, however, none of the differences between the 'premie-type and the 'harnas-type are ever significant, but these two types always differ significantly from the kom'pas-type (Scheffé procedure, $p<.05$).

So far, these results do not support the claims made by metric-

al phonologists, who predicted that the stress pattern of harnas would resemble that of 'premie in simplex words but that of kom'pas in compounds. One may argue somewhat perversely, however, that our spreakers may have behaved a-typically, and that an (even) more proficient speaker would have displayed the predicted stress shift after all. In order to resolve this possibility we ran a control experiment with an ideal (synthetic) speaker who produced the desired stress shifts exactly the way we wanted.

3. EXERIMENT II: PREFERRED STRESS PATTERN IN SYNTHETIC SPEECH

3.1. Method

The lexical material underlying the stimuli were three word pairs:

(jaar)premie: initial stress, 1st syll. open
(borst)harnas: initial stress, 1st syll. closed
(scheeps)kompas: final stress

These words were embedded in final position in compounds; the resulting set of eight words were then synthesized from diphones (using the PB30 diphone set; for details cf. van Rijnsoever, 1988) in the same two carrier phrases (i.e., once with and once without accent on the target) that were used in experiment I. Each utterance was given the same pitch pattern with a standard declination and a with a 6 semitone rise-fall on the accented syllable. The duration of final two syllables in the targets was systematically varied in five steps, so as to create a continuum from stress on the penult syllable, via level stress, to stress on the final syllable (note that 80% of the original recording speed is the standard synthesis output rate):

	penult	final
rising	48%	112%
	64%	96%
level	80%	80%
	96%	64%
falling	112%	48%

The resulting set of 3 (lexical words) * 2 (simplex/embedded) * 2 (yes/no accent on target) * 5 (temporal stress patterns) = 60 stimulus types were presented to 19 native Dutch listeners in two different random orders, who had to indicate the acceptability of each item along a scale from 0 (unnacceptable stress pattern) to 7 (completely acceptable stress pattern).

3.2. Results

From the acceptability scores of the five temporally different versions of a stimulus type we derived its preferred stress pattern for each individual listener. To this efect we devised an index such that negative values indicate stronger preference for initial stress (i.e., a relatively long first syllable), and positive values stronger preference for final stress (i.e., a relatively long second syllable); an index of 0 would indicate that perfectly even stress is preferred. Table II summarizes the results.

Table II: Mean preferred stress pattern broken down by accent type (yes/no accent on target), word type (simplex vs. embedded in compound), and lexical stress type.

target	accented		unaccented	
	simpl.	emb.	simpl.	emb.
'premie	-.22	-.08	-.52	.03
'harnas	-.19	.02	-.24	-.02
kom'pas	.03	.29	.36	.40

We notice that the effects are stronger for unaccented than for accented targets. Words with initial lexical stress are always towards the negative end of the scale, while words with final stress appear at the positive end of the scale. When the simplex words are embedded into compounds, there is a general preference for a stronger (more stressed, longer) final syllable. This effect is especially clear when the targets are accented, and somewhat instable for unaccented targets. Crucially, however, there is not the slightest preference for stronger final stress when 'harnas is embedded, even though rhythmic inversion was predicted there. Moreover, counter to the linguists' prediction, there is no systematic difference between 'premie and 'harnas.

4. CONCLUSION AND DISCUSSION

Our experiments have failed to support the predictions of metrical phonologists to the effect that embedding an initially stressed word in a compound noun would lead to an inversion of stress pattern. The stress pattern, and the temporal organisation associated with it, of an embedded noun with initial stress remains completely distinct from the stress pattern of an embedded word with final stress. We therefore take the view that these phonological predictions are wrong, and suggest that the principle of stress clash avoidance be restricted to the class of compound adjectives (the "stress retraction"-cases in §1). Notice that compound adjectives receive their stress pattern through the phrasal stress rule, i.e., by a process that is intrinsically above the level of the word. Apparently, there is no stress clash when two lexical stresses become adjacent in a compound noun, i.e., no stress clash is felt at the word level.

The duration of the first syllable in any di-syllabic word gets relatively shorter if this word is the final element of a compound (cf. table II). Three general (non-language-specific) low-level duration rules account for this phenomenon: (i) A syllable with main stress is longer than other syllables. When a word is embedded in a compound, it loses its main stress, i.e., the lexically stressed syllable loses its pitch movement, and gets shortened. (ii) Longer words are spoken faster than shorter words, therefore the syllables of the di-syllabic words will generally be shortened when they are embedded in a longer compound. (iii) A word-final syllable is lengthened so as to mark off the word (final lengthening).

Since the result, a shortened syllable at the onset of the embedded word, is compatible with the desired stress pattern of kom'pas, the shortening is not picked up for this type of word. When a long, open initial syllable is shortened (as in 'premie), the decrement in duration will be too small to reach the listener's awareness. But if a short, closed syllable gets shortened by the same amount, the effect may be above threshold and the linguist will be tempted to interpret this as a shift in stress.

We take the view, of course, that the effects of such low-level duration rules should not be mistaken for stress effects; or else we would have to interprete the same effect as a stress shift in one case ('harnas) and as a subliminal duration shift in others ('premie, kom'pas).

NOTE

Experiments 1 and 2 were run by my students Ellen L. Bish and Ruben van de Vijver, respectively.

5. REFERENCES

[1] HEUVEN, V.J. van (1987). Stress patterns in Dutch (compound) adjectives: acoustic measurements and perception data, Phonetica, 44, 1-12.

[2] JONES, D. (1918 [1964]). An outline of English phonetics, Cambridge: Heffer.

[3] KRUISINGA, E. (1918 [1964]). An introduction to the study of English sounds, Groningen: Wolters-Noordhoff.

[4] KURATH, H. (1963). A phonology and prosody of Modern English, Heidelberg: Winter.

[5] RIJNSOEVER, P.A. van (1988). From text to speech: user manual for diphone speech program DS, Handleiding no. 88, Eindhoven:IPO.

[6] VISCH, E.A.M. (1989). A metrical theory of rhythmic stress phenomena, doct. diss, Utrecht Un.

[7] ZONNEVELD, W., TROMMELEN, M. (1989). Klemtoon en metrische fonologie [Stress and metrical phonology], Muiderberg: Coutinho.