PHONETICS LABORATORY
DEPARTMENT OF GENERAL LINGUISTICS
LUND UNIVERSITY

WORKING PAPERS

19 · 1980

PAPERS GIVEN AT THE SEVENTH
SWEDISH-DANISH SYMPOSIUM
LUND APRIL 25—26 1980

PREVIOUS NUMBERS AND CONTENTS

WP 9	1973	PHONETIC SYMPOSIUM	Postgraduate students Stockholm – Lund
		SIDNEY WOOD	Speech tempo
WP 10	1975	ROBERT BANNERT ANNE-CHRISTINE BREDVAD-JENSEN	Temporal organization of Swedish tonal accents: The effect of vowel duration
		GÖSTA BRUCE	Stockholm accents in focus
		EVA GÅRDING OSAMU FUJIMURA HAJIME HIROSE ZYUN'ICI SIMADA	Laryngeal control of Swedish word accents
		KURT JOHANSSON	Perceptual characteristics of vowels
		ANDERS LÖFQVIST	Some phonetic correlates of emphatic stress in Swedish
		BERTIL MALMBERG	Niveaux, choix et systèmes approximatifs dans le langage
		KERSTIN NAUCLÉR	Some thoughts on reading and writing
		THORE PETTERSSON	In favour of the archiponeme
		EVA WIGFORSS	Foreign accent and bilingualism
		SIDNEY WOOD	What is the difference between English and Swedish dental stops
WP 11	1975	MONA LINDAU	Vowel feature
			A phonetic explanation to reduced vowel harmony systems
		SIDNEY WOOD	The weakness of the tongue-arching model of vowel articulation
			Tense and lax vowels – degree of constriction or pharyngeal volume?
WP 12	1975	ROBERT BANNERT	The significance of vowel features in the perception of complementary length in Central Bavarian
			Temporal organization and perception of vowel-consonant sequences in Central Bavarian
		GÖSTA BRUCE	Swedish accents in sentence perspective
		EVA GÅRDING	The influence of tempo on rhythmic and tonal patterns in three Swedish dialects

		PETER KITZING HANS-ERIK RUNDQVIST EWA TALO	Fundamental frequency of the voice in continuous speech. - Preliminary report on a device for determining mean and distribution of frequencies
		ANDERS LÖFQVIST	On the control of aspiration in Swedish
		MAGNÚS PÉTURSSON	Étude glottographique de quelques consonnes islandaises
WP 13	1976	ANDERS LÖFQVIST	Closure duration and aspiration for Swedish stops
			Oral air pressure in the production of Swedish stops
WP 14	1976	GÖRAN SONESSON	Au sujet des fondements de l'analyse phrastique par Per Aage Brandt. Compte rendu
		THORE PETTERSSON	Polsk fonologi i generativ tappning
WP 15	1977	MILAN BÍLÝ	Coreference rules described in terms of functional sentence perspective (FSP)
		EVA GÅRDING	Vergleichende Studien zur Prosodie schwedischer Dialekte
		MAGNÚS PÉTURSSON	L'aspiration des occlusives après [s]
		HANS RANDLER	Identification test concerning isolated disyllabic Swedish accent 1 and accent 2 words
		SIDNEY WOOD	A radiographic analysis of constriction locations for vowels
		ROBERT BANNERT ANNAE-CHRISTINE BREDVAD-JENSEN	Temporal organization of Swedish tonal accents: The effect of vowel duration in the Gotland dialect
WP 16	1978		STUDIES IN GENERAL LINGUISTICS dedicated to BERTIL MALMBERG by students and colleagues on the occasion of his 65th birthday 22nd April 1978
		MILAN BÍLÝ	Some thoughts about functional sentence perspective, empathy, and reflexives
		EVA GÅRDING KRISTINA LINDELL	Tones in Northern Kammu: a phonetic investigation
		D.J. HACKMAN	The communicative competence of foreigners in Swedish: listener attitude and contextual appropriacy

		KERSTIN HADDING KERSTIN NAUCLÉR	Permissible and not permissible variations in pitch contours
		OLAV HAMMER	Historical Linguistics and generative phonology
		KENNETH HYLTENSTAM	A framework for the study of interlanguage
		EVA LARSSON	Effet communicatif de la dislocation d'un NP en français
		THORE PETTERSSON	Reading Chomsky
		CHRISTER PLATZACK	The Swedish prepositions på and i followed by measure phrases
		SVEN PLATZACK	The sign and its substance. A coloured view
		BARBARA PROHOVNIK	Linguistic aspects of bilingual aphasia
		CHRISTOPHER STROUD	The concept of semilingualism
		SIDNEY WOOD	Tongue retraction is not so useful after all
WP 17	1978		The PROSODY OF NORDIC LANGUAGES Symposium 14-16 June 1978 Abstracts
WP 18	1978	KENNETH HYLTENSTAM	Variation in interlanguage syntax

Nos. 3, 7, 8 are out of print. A revised edition of No. 8 has been published separately as No. XI in Travaux de l'Institut de Linguistique de Lund (Gleerups, Lund), 1977.

Correspondence concerning this series should be addressed to:

Sidney Wood
Lund University
Department of Linguistics and Phonetics
Helgonabacken 12
S-223 62 LUND

PHONETICS LABORATORY
DEPARTMENT OF GENERAL LINGUISTICS
LUND UNIVERSITY

WORKING PAPERS

19 · 1980

PAPERS GIVEN AT THE SEVENTH
SWEDISH-DANISH SYMPOSIUM
LUND APRIL 25—26 1980

This volume of Working Papers contains summaries of papers given at the seventh Swedish-Danish Phonetics Symposium in Lund on April 25-26, 1980. Contributors to the symposium are mainly advanced and graduate students of phonetic sciences.

We dedicate this volume to

Eva Gårding

on the ocasion of her sixtieth birthday 14 July 1980. It was on her initiative that the first Symposium was held in Lund 1973.

Friends and colleagues at
the Department of Linguistics
and Phonetics, Lund.

LIST OF PARTICIPANTS

Institute of Phonetics, University of Copenhagen

Niels Dyhr
Eli Fischer-Jørgensen
Ved Kumari Ghai
Mette Greiffenberg
Peter Molbæk Hansen
Birgit Hutters
John Jørgensen
Kirsten Morsing
Niels Reinholt Petersen

Institute of Linguistics, University of Stockholm

Eva Holmberg*
Wendy Linker

Department of Speech Communication, Royal Institute of Technology, Stockholm

Rolf Carlsson
Björn Granström
Sharon Hunnicutt
Inger Karlsson
Lennart Nord*
Karl-Erik Spens
Ulf Stålhammar

Department of Linguistics, Uppsala University

Christoffer Ekströmer
Olle Engstrand*
Lars Fant*
Lennart Nordstrand
Sven Öhman

Department of Linguistics and Phonetics, Lund University

Robert Bannert
Anne-Christine Bredvad-Jensen
Barbro Bruce
Gösta Bruce
Eva Gårding
Kurt Johansson
Boris Larnert
Per Lindblad
Knut Lindell
Eva Magnusson
Bertil Malmberg
Ulrika Nettelbladt Dalby
Barbara Prohovnik
Paul Touati
Ursula Willstedt
Sidney Wood

* absent contributor

SUMMARIES OF PAPERS

MODEL PREDICTIONS OF VOWEL DISSIMILARITY*

Rolf Carlson & Björn Granström
Institutionen för talöverföring, Tekniska Högskolan, Stockholm

Several models of the peripheral auditory system have been studied. As test material we have used perceptual data on psychoacoustic and phonetic dissimilarity (Carlson, Granström, & Klatt, 1979; Klatt, 1979). The predictive value of the models depends on type of stimuli and perceptual task. Type of metric used in the dissimilarity calculation is of small importance for the correlation between predicted and perceived dissimilarity. We find a strong support for the view that some kind of peak picking mechanism is involved in speech perception. A model is presented that includes such a feature.

* An expanded version of this paper is published in STL-QPSR 3-4/1980 (Inst. för talöverföring, Tekniska Högskolan, Stockholm).

References

Carlson R., Granström B., and Klatt D. 1979. Vowel perception: the relative salience of selected acoustic manipulations. Speech Communication Papers. Eds. Wolf and Klatt. Acoustical Society of America. Cambridge, MA 1979, 9-14.

Klatt D. Perceptual comparisons among a set of vowels similar to /æ/: Some differences between psychophysical distance and phonetic distance. Paper NN1 presented at the 98th Meeting of the Acoustical Society of America, Nov. 1979.

A PILOT INVESTIGATION OF THE F_o PATTERN IN AMERICAN ENGLISH

Niels Dyhr
Institute of Phonetics, University of Copenhagen

Abstract

Fourteen declarative sentences were recorded by one native speaker of American English. The F_o pattern (the F_o movement within the stress group) was investigated, and a model was proposed. This model was compared to one based on identical material, but with a British speaker.

1. Introduction

In British English and "East Coast American English" the stressed syllables have a higher fundamental frequency than the unstressed ones (Fry, 1958, Lieberman, 1960). Bolinger (1970) mentions, however, that in other types of American English a pattern may be found in which the fundamental frequency is lower in the stressed syllables than in the unstressed ones. The purpose of the pilot experiment reported below was to examine instrumentally this opposite F_o pattern.

2. Procedure

The test material consisted of 14 declarative sentences, each consisting of a test word (either a natural word or a nonsense - but possible - word) embedded in a carrier phrase. To avoid the influence of intrinsic F_o differences among vowels, the material was constructed in such a way that each test sentence contained either low or high vowels throughout.

The test material contained the following stress combinations: Stress on the first, second, third, and fourth syllable in the test word, followed by zero, one, and two unstressed syllables.

The material was recorded ten times by a twenty-two year old American male speaker, born and raised in California.

Tracings of the F_o movements were made, and superimposed on one another. Average F_o curves were drawn by hand. A quantitative measure for the central tendency of the F_o level in the vowels was obtained by measuring the average F_o curves at a point two thirds from the vowel start, cf. Rossi (1971).

3. Results

It was found that in the natural words all the stressed syllables had a lower fundamental frequency than the unstressed ones. For the nonsense words the pattern was the opposite. This seemed to be due to the fact that the speaker, who had had no phonetic training, was unable to pronounce the nonsense words without emphasizing them. The change of pattern did not occur at the word boundary but at the stress group[1] boundary - i.e., the F_o pattern did not change before the stressed syllable in the nonsense word, even when the stressed syllable was the last one in that word. This seems to indicate that the F_o pattern is controlled, not by the word but by the stress group.

The results for the natural words can be described by the model shown in figure 1.

The model based on the data from the American speaker can be compared to one based on an identical test material, but with a British speaker, figure 2. It is evident that, apart from the opposite F_o pattern, a definite sentence nucleus (focus) is found only in the British model, which also has a greater F_o variation than the American model.

It must be pointed out that, considering the very restricted material, the present result should be viewed with care. A preliminary examination of data obtained for another American (Wisconsin) speaker, however, seems to be in good agreement with the tendencies outlined in the present paper.

References

Bolinger D.L. 1970. Relative Height. Intonation, 137-153.

Fry D.B. 1958. Experiments in the Perception of Stress. Language and Speech 1, 126-151.

Lieberman P. 1960. Some Acoustic Correlates of Word Stress in American English. JASA 32, 451-454.

Rossi M. 1971. Le seuil glissando ou seuil de perception des variations tonales pour les sons de la parole. Phonetica 23, 1-33.

---

1) The stress group is defined as a stressed syllable plus the following unstressed ones.

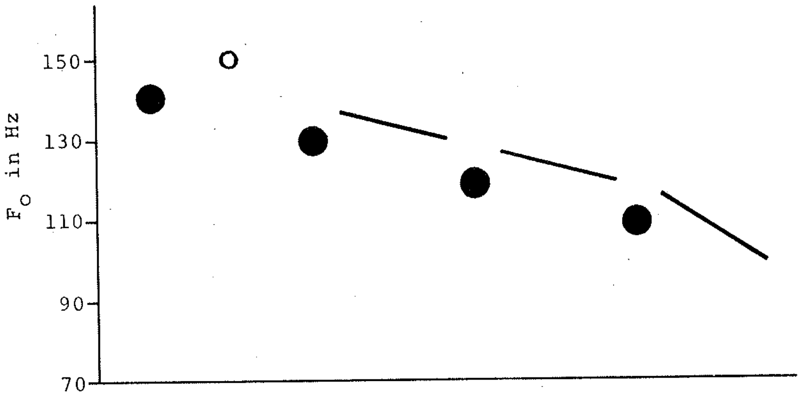


Figure 1. American model based on sentences with natural words. ● indicates stressed syllable, o unstressed syllable, —— a series of unstressed syllables.

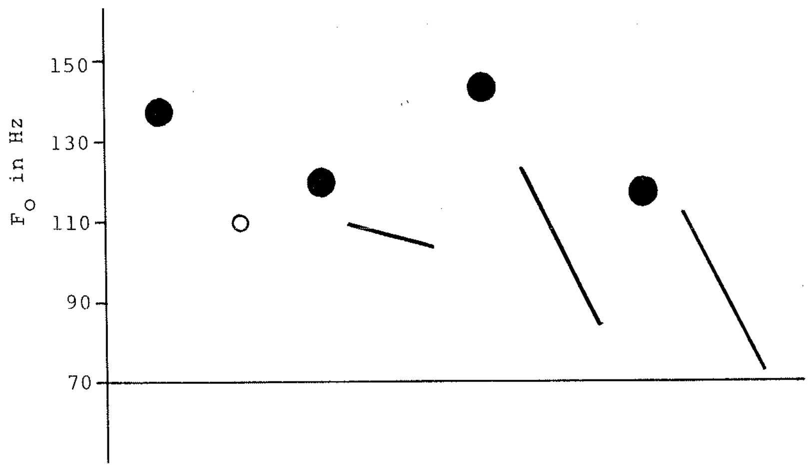


Figure 2. British model based on material identical with that of the American model. See further legend of figure 1.

ARTICULATORY COORDINATION IN SELECTED VCV UTTERANCES: ACOUSTIC-AUDITORY CONSIDERATIONS

Olle Engstrand
Institute of Linguistics, Uppsala

Introduction

The basic argument of this paper is that the predictive power of articulatory models largely depends on the extent to which they are supported by an explicit theory of the acoustic-auditory properties of speech sounds. More specifically, the purpose is to demonstrate a case where the choice of motor strategy is determined by acoustic-auditory criteria rather than by constraints on the motor system.

The speech sample consists of cineradiographic recordings of a few utterances of the vowel-consonant-vowel (VCV) type, namely ipi, ipu, upi, upu, and ipa. The experimental task is to study the trajectories of the tongue body movements from the first vowel (V_1) to the second (V_2) across the intervening consonant. For the sake of the discussion, we adopt as our null-hypothesis that these trajectories are approximately linear. From an articulatory point of view, this is a reasonable assumption since there are no obvious anatomical restrictions on the mutual independence between tongue and lip movements.

As our alternative hypothesis, we simply take the contrary, i.e., the tongue body trajectories from V_1 to V_2 will deviate from a straight course. If this turns out to be true, further qualifications as to the nature of the deviation will have to be made.

We assume, as a matter of fact, that the null-hypothesis will have to be rejected. This assumption is based on the following two grounds:

1. An optimal articulatory strategy will be such that the crucial acoustic-auditory effects intended by the speaker are kept (a) essentially invariant and (b) free from conflicting noise.

Strong evidence in favor of acoustically invariant properties of stop sounds has recently been offered by Blumstein and Stevens (1979). They showed that properties such as 'diffuse-rising', 'diffuse-falling', and 'compact', residing in the short-time release spectra, are not appreciably affected by changes in the immediate vowel context.

The characteristic (manner) feature of Swedish stops is taken to be a short burst of fricative noise with an abrupt onset. As is well known, however, the production of a voiceless stop is often accompanied by numerous acoustic events. This immediately necessitates a justification of the choice of one particular property, or event, as the 'object of the speaker's intention' in the sense of von Wright (1971, p. 89). To this end, we can try to show that all acoustic events that tend to materialize in connection with the noise burst can be plausibly regarded as causally related to the burst in a certain way (cf. Öhman et al. 1979). Then, in terms of production, the closing of the mouth in combination with an egressive airflow generated in the lungs can be looked upon as adjustments made in preparation for the 'executive' opening gesture through which the intended burst is brought about. Those adjustments make the burst possible, so to speak, not the other way around.

Consequently, the fricative burst is taken to be the primary, intended effect of the articulatory movements in question. The acoustic consequences of preparatory (and postcipatory) movements, e.g. silent interval and formant transitions, will then be regarded as side-effects, secondary to the burst.

2. In order to bring about the primary acoustic properties of a p sound, i.e. a short noise burst characterized by a diffuse-falling spectrum and an abrupt onset, some aero-dynamic conditions have to be met with.

According to Stevens (1971), the pressure drop P_d at a constriction is proportional to the density of the air ρ, the volume velocity U, and the cross-sectional area A of the constriction by

$$P_d = k(\rho U^2/2A^2),$$

where k is a constant.

Now consider a pi or pu sequence. We assume (Stevens 1971) that the cross-sectional area A at the tongue-palate constriction is in the order of 0.3 cm^2 for the high vowels i and u, and, further, that the volume velocity U rises to about 1500 $cm^3 sec^{-1}$ as an immediate consequence of the p release. Then, if both these conditions are present simultaneously, i.e. if the tongue fully anticipates the vowel at the release, the results of Stevens' calculations show that a considerable turbulence noise source will be generated at the tongue-palate constriction. Consequently, the (primary) p burst will be (a) perceptually masked by the secondary noise, and, moreover, (b) physically weakened, since, by the above equation, the lowered volume velocity U which results from the presence of a secondary constriction will reduce the pressure drop P_d across the mouth orifice. Also, (c) the secondary noise will contain frequency components not compatible with an ideal p spectrum.

These being negative consequences of full vowel anticipation by the tongue at the stop release in pi and pu, we assume that some articulatory measure is taken to avoid them. For instance, according to the above formula, a great enough increase in A will, all other things being equal, lead to a reduction of the pressure drop across the tongue-palate constriction. This means that the tongue should not be in a high vowel position at the moment of release. In other words, the tongue body movement from a high V_1 to a high V_2 across p has to deviate from its straight course so that a wide enough air outlet at the release moment is provided for. In the special case of the symmetrical ipi and upu sequences we would, consequently, expect a tongue lowering gesture to coincide with the p, whereafter the tongue body would resume its high vowel position.

Concerning the timing of the initiation of the tongue lowering gesture, we draw on a few pilot experiments with speech synthesis recently carried out in our laboratory[1)]. Those experiments indicate that, if the voice source is maintained slightly too long at a p implosion, listeners tend to hear the sequence bp, for instance ibpi for ipi. We therefore expect the glottis to begin its opening gesture before the labial closure in order to avoid the undesired impression of a voiced consonant. The concomitant rise in volume velocity will then create a turbulence noise if the high vowel constriction is still present. We therefore assume that the tongue lowering gesture will begin before the implosion so that turbulence noise is avoided.

These considerations lead us to reject the null-hypothesis - i.e. that the tongue body movement trajectory follows a straight line - and to accept the alternative hypothesis with the above temporal specifications. We will now compare these predictions with experimental evidence.

Experimental methods[2)]

The results presented in this report are drawn from a larger body of data including simultaneous high-speed cineradiographic and electromyographic recordings made in an attempt to cover various aspects of speech production. Here, the relevant cineradiographic observations will be singled out for inspection. No EMG data will be reported on in this paper.

The subjects were two Central Standard Swedish speakers, one female (LE) and one male (BG), both in the beginning of their thirties. During the recording session the subjects read various VCV utterances for two minutes each using systematically varying rate and stress on each utterance. The utterances studied here were produced at a fairly slow rate with the sentence accent on the VCV part which was embedded in the carrier phrase 'säja p___p igen' ('say p___p again'). The grave word accent characteristic of Swedish compounds was used. In combination with the sentence accent, this gives roughly equal prominence to both syllables.

The observed X-ray frames were sampled from a film of the mid-sagittal aspect of the subjects' vocal tracts taken with a 16 mm cine camera at a speed of 60 frames/sec. Small radio-opaque pellets were attached to three positions along the midline of the tongue dorsum and to one position each on the lower teeth and the upper and lower lips. A reference pellet was attached to the upper teeth. A contrast medium was applied to the lips of both subjects and to the upper tongue surface of subject LE. For the purpose of simultaneous and immediately subsequent EMG-recording, both subjects had needle electrodes implanted in a few muscles (lips, jaw, and tongue). Both subjects reported, after a few minutes practice, that the presence of the pellets and the needles was not felt to disturb their speech.

The subjects were seated comfortably with their heads positioned towards a hollow head-support. No other head-holder was used. The subjects were asked not to move during the run. Their steady position was continuously monitored, and no appreciable movement was detected.

The film analysis was carried out in different ways. Several frames were traced in order to obtain the outlines of the mandible, tongue, lips, and palate at certain points in time. For all test utterances, except those rejected because of subject errors, the x and y pellet coordinates relative to the reference pellet were fed into a computer, and their movements were plotted as a function of time. Those diagrams will be published in a forthcoming report.

Results[3)]

The predictions made in the introductory section are fully corroborated by the data. A consistent cross-subject and cross-utterance feature of the data is the striking non-linearity of the trajectories of tongue body movement from V_1 to V_2 across the intervening consonant (see Fig. 1). The tracings evidence a slight movement away from the extreme V_1 tongue position before the p implosion. This movement goes in the direction of the approximately neutral vocal tract which, except for the lips, characterizes the moment of p release. Inspection of consecutive frames indicate that a movement towards V_2 has started at the moment of release but that most part of the trajectory still remains to be completed at that time. In Fig. 2, the strikingly similar vocal tract shapes at the p release of ipi, ipu, and ipa are shown. A comparison with the fully developed shapes associated with i, a, and u, which can be seen in Fig. 3, reveals that the second vowel anticipation at the release is a very weak one.

In ipi and upu, there is a tongue lowering gesture which coincides with the consonant, i.e., the tongue body does not remain in the i and u

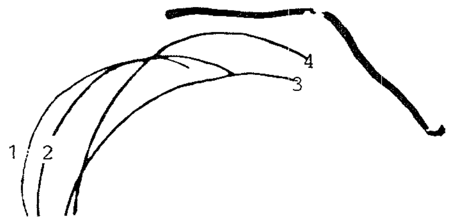


Fig. 1. Tongue contours for utterance upi. 1 = u, 2 = moment of lip closure for p, 3 = moment of p release, 4 = i. Subj. BG.

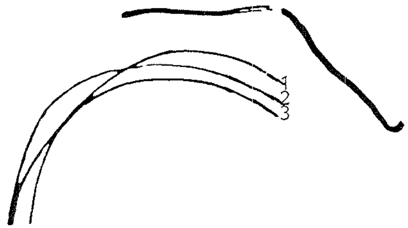


Fig. 2. Tongue contours at the moment of p release for (1) ipi, (2) ipu, and (3) ipa. Subj. BG.

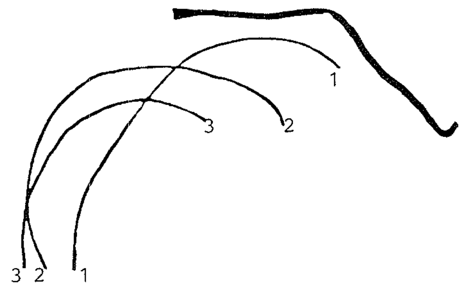


Fig. 3. Fully developed second vowel tongue contours for (1) ipi, (2) ipu, and (3) ipa. Subj. BG.

Fig. 4. Tongue contours for utterance upu. u_1 and u_2 have essentially similar contours, while the tongue body has been lowered at the moment of p release. Subj. LE.

positions even though the second vowel requires roughly the same position as the first one. The sequence upu is exemplified in Fig. 4. The 'trajectories' are, in other words, not straight in the symmetrical case either. Similar phenomena have been observed in American English subjects (Gay 1978).

Conclusion

The results referred to provide a case where one pattern of articulatory coordination (the 'alternative hypothesis') is preferred on its acoustic-auditory merits over another, different pattern (the 'null-hypothesis'). What makes this choice of motor strategy particularly interesting is that it cannot be readily explained on mere anatomical or physiological grounds. This shows the necessity of paying adequate attention to the real object of articulatory action, i.e. the bringing about of specific acoustic-auditory effects.

Notes

1. With Lennart Nordstrand.
2. Further details about experimental methods and procedures will be given in a forthcoming publication.
3. Due to space limitations, only a few representative examples of the data can be shown here.

References

Blumstein, S.E. and K.N. Stevens. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. J. Acoust. Soc. Am. 66(4), 1001-1017

Gay, T. 1978. Articulatory units: Segments or syllables? In Syllables and segments. Eds. Bell and Hooper. North-Holland, Amsterdam

McAllister, R. 1978. Temporal asymmetry in labial co-articulation. Papers from the Institute of Linguistics, University of Stockholm 35

Stevens, K.N. 1971. Airflow and turbulence noise for fricative and stop consonants: Static considerations. J. Acoust. Soc. Am. 50(4), 1180-1192

von Wright, G.H. 1971. Explanation and understanding. Routledge and Kegan Paul, London

Öhman, S., S. Zetterlund, L. Nordstrand and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. In Proceedings of the 9th International Congress of Phonetic Sciences held in Copenhagen Aug. 6-11 1979. Vol. II

SYLLABLE STRUCTURE AS A FACTOR IN SOME DANISH SOUND CHANGES

Peter Molbæk Hansen
Institute of Phonetics, University of Copenhagen

Introduction

Basbøll (1973) has argued that the syllable - considered at a rather low level of abstraction - must be recognized as an important unit in phonology. Basbøll has shown, among other things, that the classification of Danish consonants according to their distributional behaviour - in particular along the lines suggested by Sigurd (1965) - may be meaningfully and fruitfully compared with their phonetic classification in terms of distinctive features.

It is tempting now to apply Basbøll's insight to the dynamics of phonology, in order to see whether the hierarchic description of syllable structure will in any way contribute to our understanding of sound change.

I have attempted to relate some of the recent sound changes in Standard Danish - so well documented by Brink and Lund (1975) - to syllable structure conditions expressed in terms of "order classes" and the corresponding distinctive feature hierarchy.

Sound changes in 19. century Standard Danish

A close inspection of the findings of Brink and Lund (1975) allows us to reconstruct with reasonable confidence the sound system of Copenhagen Standard Danish (henceforth CSD) of the late eighteenth century, and also to consider the structure of syllables in terms of a distinctive feature hierarchy a la Basbøll. Fig. 1 shows the structure of peak + coda of CSD syllables shortly before 1800. The order classes are numbered (arbitrarily) from 1 to 5. The hierarchy deviates from that of Basbøll in two respects: 1) the feature framework is that of Jakobson, Fant and Halle (1952), cf. the existence of the feature "vocalic". I do not intend to argue here for or against certain features or

Order class:	5	4	3	2	1
	V	:	l	m	s
		i̯	ð	n	f
		u̯	ʁ	ŋ	b
		ɨ̯		v	d
				ɣ	g
kål 'cabbage'	ɔ	:ˀ	l		
rød 'red'	ø	:ˀ	ð		
hær 'army'	ɛ	:ˀ	ʁ		
negl 'nail'	ɑ	i̯ˀ	l		
tid 'time'	i	:ˀ	ð		
arbejd! 'work!'	ɑ	i̯ˀ	ð		
sejr 'victory'	ɑ	i̯ˀ	ʁ		
skovl 'shovel'	ɒ	u̯ˀ	l		
ud 'out'	u	:ˀ	ð		
sur 'sour'	u	:ˀ	ʁ		
hagl 'hail'	ɑ	ɨ̯ˀ	l		
lån 'loan'	ɔ	:ˀ		n	
bog 'book'	ɔ	:ˀ		ɣ	
støv 'dust'	ø	:ˀ		v	
hegn 'fence'	ɑ	i̯ˀ		n	
sagn 'tale'	ɑ	ɨ̯ˀ		n	
havn 'harbour'	ɑ	u̯ˀ		n	
rødm! 'blush!'	ø		ðˀ	m	
elm 'elm'	ɛ		lˀ	m	

+syllabic (class 5)
-consonantal (classes 5–4)
+vocalic (classes 5–3)
+sonorant (classes 5–2)

Fig. 1. Segments and order classes before 1800.

feature definitions, but the substitution in generative phonology of the feature "vocalic" by the feature "syllabic" seems to me misguided: the feature "vocalic" is in my opinion a useful inherent feature. 2) vowel length is interpreted analytically, i.e. the symbol ":" is an abbreviation for a series of -consonantal, +vocalic segments (ɑ̯, e̯, i̯, etc.; e.g. a word like vane 'habit' is interpreted as /vɑɑ̯ne/). This is not to deny that vowel length may in certain respects function as a prosodic or even as an inherent feature; I simply assume that length and gemination are equivalent in the sense that long vowels behave as single segments in certain respects and as a sequence of two segments (which happen to be identical as far as their inherent feature composition is concerned) in other respects, e.g. in their function in the syllable. I follow Basbøll (1973), however, in considering ɣ a dorsal fricative occurring after long vowels and (before 1800 probably only in over-distinct spelling pronunciations) after l and r; I thus take ɣ and i̯ (the latter occurring after short vowels only) to be (phonetically) distinct, although this distinction is not (explicitly) recognized by Brink and Lund, who use the symbol ɣ for both sounds.

If we assume that the order classes and feature hierarchy of fig. 1 correctly depicts the structure of possible codas shortly before 1800, the following language specific coda structure conditions may be set up:

1) the coda may never contain more than two sonorant segments.

2) if a coda contains two sonorant segments, then these segments must not belong to the same order class; more specifically: the rightmost segment must belong to a lower order class than the leftmost segment, i.e. sonority must fall during the sonorant part of the rhyme of the syllable.

Now, Brink and Lund report that vocalized pronunciations of postvocalic /ʁ/, /ɣ/ and /v/ turn up - and eventually are generalized - in the speech of people born in the first half of the 19. century. These vocalized pronunciations - i.e. ɒ̯ instead of ʁ, i̯ or u̯ instead of ɣ, and u̯ instead of v - occur roughly in the following contexts: 1) before consonants, 2) before shwa, and 3) word finally. In the following the term "weak position" will be used as a common denominator for such contexts. These changes may be stated informally thus:

1) ʁ → ɒ̯ in weak position
e.g. gɔːʼʁ → gɔːʼɒ̯ 'farm'

2a) ɣ → i̯ in weak position after front vowels and l
e.g. søːʼɣ → søːʼi̯ 'search!'

2b) ɣ → u̯ in weak position after back vowels
e.g. bɔːʼɣ → bɔːʼu̯ 'book'

3) v → u̯ in weak position
e.g. gʁɑːʼv → gʁɑːʼu̯ 'grave'

These changes were not simultaneous: according to Brink and Lund, 1) must have started around 1800, 2a) and 2b) started shortly before 1840, i.e. the first occurrences of i̯ and u̯ instead of ɣ are found in the speech of people who were born between 1830 and 1840. 3) started around 1850, i.e. the first occurrences of u̯ instead of v are found in the speech of people who were born around 1850.

There is reason to believe, furthermore, that the segment ð had been reinterpreted as -consonantal, too (Brink and Lund mention that this segment was often pronounced as a fricative by their oldest informants; today ð is never pronounced with friction and is classified as -consonantal (as a vocoid) by Basbøll (1975). It is reasonable to hypothesize that around 1860 ð was classified together with ɒ̯, i̯, u̯, ɨ̯, and : (V̯) as -consonantal, +vocalic.

Thus, around 1860 the vocalized pronunciations of postvocalic /ʁ/, /ɣ/, /v/, and /ð/ was already common, at least after long vowels (= VV̯, cf. above), and this meant that many words in which a long vowel was followed by one of these vocalized segments had come into conflict with the syllable structure conditions mentioned above: If words like gård, ud, bog, søg!, grav were frequently and fashionably pronounced gɔ:'ɒ̯, u:'ð, bɔ:'u̯, sɑ:'i̯, gʁɑ:'u̯ around 1860, at least by young people, i.e. by people born later than, say, 1840, then these words contained syllable codas with sequences of sonorant segments of equal sonority, i.e. belonging to the same order class, viz. the one defined by the features -consonantal, +vocalic (order class 4 in fig. 1). It is very interesting to note, therefore, that Brink and Lund report a new change to have started around 1860, viz. a general shortening of long vowels before the segments ɒ̯, i̯, u̯, and ð, i.e. exactly the segments before which long vowels are in conflict with the hypothesized coda structure conditions (presupposing, still, that long vowels function as VV̯ in the syllable).

This vowel shortening is far from completed today, but in the speech of younger Copenhageners (especially of the higher social classes) pronunciations with long vowels in monosyllabic words like the ones mentioned above are definitely obsolete or even impossible.

<u>Interpretation of vowel shortening</u>

If this vowel shortening is seen as a sort of therapy the function of which is to reestablish the syllable structure conditions mentioned above, then we can visualize the new state of affairs (still not completely reached) as in fig. 2:

order class:	5	4	3	2	1
	V	:	l	m	s
		i̯		n	f
		u̯		ŋ	b
		ɒ̯		v	d
		ð			g

Fig. 2. Segments and order classes after the changes.

A comparison of figs. 1 and 2 will reveal that the order classes - as defined by distinctive feature configurations - are the same as before, but the inventories of segments in order classes 4 and 3 have been changed, order class 3 having been reduced to containing only the segment l, and order class 4 having been enriched by the segments ɒ̯ and ð. In the speech of young people the segment ɨ̯ (occurring only after certain short open vowels) has also disappeared, having developed into i̯ or u̯ in a way parallel to γ.

Words like bjerg 'mountain' and tørv 'turf' are of particular interest in this connection: it is uncertain whether these words were pronounced as true monosyllables before 1800, cf. old pronunciations like bjɛʁ'(ə) and tæʁ'(ə) attested by Brink and Lund and by old spellings. To-day they are often pronounced with a short (r-coloured) vowel, i.e. bjæu̯', tŒu̯', respectively, at least by young people. It seems natural to relate vowel shortening before -consonantal segments to the deletion of a -consonantal, +vocalic segment in the same position.

Needless to say, the proposed explanation of the above-mentioned changes is, at best, a de-post-facto explanation, but this is what most "explanations" of sound change are.

References

Basbøll, H. 1973. Notes on Danish Consonant Combinations. ARIPUC 7, p. 103-142.

Basbøll, H. 1975: On the Phonological Interpretation of the Falling Diphthongs in Danish. ARIPUC 9, p. 49-108.

Brink, L. and J. Lund, 1975: Dansk Rigsmål. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København. Gyldendal, Copenhagen.

Jakobson, R., G. Fant, and M. Halle, 1952: Preliminaries to Speech Analysis. MIT Press, Cambridge, Massachusetts.

Sigurd, B., 1965: Phonotactic Structures in Swedish. Lund.

LARYNGEAL AIRWAY RESISTANCE AS A FUNCTION OF PHONATION TYPE.

Eva Holmberg
Institutionen för lingvistik, Stockholm

Abstract

An indirect method to estimate laryngeal airway resistance was investigated. The ratio of the intraoral pressure (IOP) for the voiceless stop and the volume air flow (V_o) for the open vowel in the CV utterance /pa/ was suggested to give an estimate of the laryngeal airway resistance (R_{law}); $R_{law} = IOP/V_o$. The primary variable was phonation type. It was hypothesized that the R_{law} value would reflect the laryngeal airway resistance during 1) normal, 2) pressed and 3) breathy phonation. Other controlled variables were intensity and fundamental frequency. The results suggested that the R_{law} value was highest for pressed phonation and lowest for breathy phonation irrespective of intensity and fundamental frequency. Good discrimination in the R_{law} value between high and low intensity was also found. These results were interpreted to give evidence that the R_{law} value obtained with this method was a good index of actual resistance.

References

Hixon T.J. A Clinical Method for Estimating Laryngeal Airway Resistance during Vowel Production. Personal communication.

Netsell R. 1969. Subglottal and Intraoral Air Pressures during the Intervocalic Contrast of /t/ and /d/. Phonetica Vol 20 No 2-4.

NEW ENGLISH RULES FOR THE KTH TEXT-TO-SPEECH SYSTEM

Sharon Hunnicutt
Dept. of Speech Comm., Royal Inst. of Technology, Stockholm

Introduction

A set of English rules is presently being written for the speech synthesis system developed at the Royal Institute of Technology (KTH) in Stockholm. This system is constructed to be language-independent. Rules were first written for Swedish,[1] and an English rule system was first presented in 1975.[2] The focus of the current effort has been the development of a more complete set of grapheme-to-phoneme and lexical stress rules. A set of rules to convert expressions involving numbers to words has also been written for the KTH system, and a small lexicon has been added. (1) The material in this paper is an outgrowth of the process of constructing grapheme-to-phoneme and lexical stress rules for the existing formalism of the KTH system. Expressing the rules in this formalism provided the impetus for a study of the constraints and the opportunities presented by this system, and also led to a categorization of rules in terms of special contexts which signal likely exceptions.

The KTH system accepts unrestricted input text, and its first operation is to convert this text to phonemes. This conversion is accomplished either by a small lexicon or by two parallel sets of rules: a set of grapheme-to-phoneme and lexical stress rules, and a set of number-to-phoneme rules. The remainder of the English system contains phonological rules such as devoicing and flapping which are followed by prosodic rules to determine segment durations and fundamental frequency. The segments are expressed as parameters, and synthesized with an OVE III.[5]

An important feature of the KTH system is a special higher-level programming language, the structure of which is similar to that used in generative phonology.[6] The present effort represents the first large-scale attempt to have someone familiar with the rules of another language use this programming language to express their knowledge of these rules.

(1)The author has written a set of grapheme-to-phoneme and lexical stress rules for English, and has worked extensively with various modules in the text-to-speech system developed at the Massachusetts Institute of Technology (MIT). References concerning this work are given in notes 3 and 4 at the end.

The attempt appears to have been successful; the rules were written quickly, and the discipline of the new formalism provided an inspiring perspective on previous work. The categorization of rules mentioned above and some observations about the KTH formalism and the utility of the higher-level programming language are presented below.

Rule Types and Special Contexts

The types of rules needed to predict the grapheme-to-phoneme correspondence in English may be separated into two groups, basic rules, giving the normal pronunciation, and contextually-dependent rules. The KTH system contains approximately 310 grapheme-to-phoneme rules, 50 of which specify the basic, or most frequent, pronunciation of all single vowels and consonants and some consonant clusters and vowel digraphs.

Remaining rules are rather evenly divided into (a)rules for affixes and (b)rules for consonants and consonant clusters, and for vowels and vowel digraphs in special contexts. There are around 130 rules of each of these two types.

There is some question as to whether affixes in general should be recognized and converted by separate rules. Many affixes would be correctly pronounced by the rules for vowels and consonants, would be correctly analyzed by the stress rules, and are not used in any other rule contexts. On the other hand, the morpheme boundaries they define may be useful in syllabification, and it is possible that they signal some prosodic effects such as reduced duration or less F0 excursion.

Special contexts, in which less frequent grapheme-to-phoneme correspondences occur, are seen to be specified by only about a dozen categories. Furthermore, these categories frequently predict special pronunciations for both consonants and vowels. These categories are shown in Figure 1; the same, or similar, contexts for vowels and consonants are found opposite each other. Examples of graphemes receiving correct pronunciation by rules in these categories are also shown.

Most special contexts can be defined in terms of morpheme boundaries. Some contexts express the notion of morph-initial (1) or morph-final (4), while others specify the first (2), last (5,6,10,11) or only (3) consonant(s) or vowel in a morph. Other special contexts can be defined in terms of suffixes (5,8,10). Vocalic inflectional suffixes (5) signal word-final contexts and the end of free roots (6). Two types of "laxing"

VOWELS	CONSONANTS
1) MORPH-INITIAL eulogy, ewe	1) MORPH-INITIAL knee, white, wrote, pneumatic, years
2) FIRST SYLLABLE thiamine, iodine	
3) SINGLE-SYLLABLE MORPH pie, sour, table, raced, striate, vocal, me	
4) WORD-FINAL / MORPH-FINAL bonnie, toe, potato, bake, ma	4) WORD-FINAL / MORPH-FINAL sing, tic, inch, arguable, pariah
5) PRECEDING C - VOCALIC INFLECTIONAL SUFFIX baking, miles, loner, themes	5) PRECEDING VOCALIC INFLECTIONAL SUFFIX wreathes, antiques, apples, acres
6) PRECEDING C- MORPH-FINAL "e" bake, mile, alone, theme	6) PRECEDING C - MORPH-FINAL "e" wreathe, antique, apple, acre, orange, cheese
7) PRECEDING LIQUID / C - LIQUID air, earth, ward, shoulder, doll, or, ...	7) PRECEDING / FOLLOWING A LIQUID quadrille, orle, place
8) IN THE CONTEXT V - C - V_{high} - V alienate, ameliorate, usual	8) IN THE CONTEXT V - C - V_{high} - V revision, dispersion, Russian, racial, ..ation
9) PRECEDING ANOTHER VOWEL Ptolemaic, meteor, myopic, experience	9) PRECEDING ANOTHER CONSONANT chrome, pick, comprehension
10) PRECEDING C - SPECIAL "LAXING" SUFFIXES ratify, utility, conic, edible	
11) OTHERS ..ook, wa..., ..aste, ..ind, ..igh	11) OTHERS -voiced "s" rules

Figure 1 - SPECIAL CONTEXTS

Note: "C" indicates a single consonant, and "V" a single vowel.

suffixes occur (10), and seven suffixes are included in the more general context specified in (8) which is used to signal palatalization of some preceding consonants and the occurence of a long vowel preceding a single consonant in this position. The most prolific exception-generating contexts are those in which a liquid occurs; thirty such rules are included.

<u>Aspects of the Formalism: A Comparison</u>

Most of the differences in the statement of the KTH rules and the MIT rules stem from the type of rule cycle used in grapheme-to-phoneme conversion. Application of the MIT rules is accomplished in three passes: affix removal, consonant conversion in the remnant (assumed to be a monomorphemic root), and conversion of the remainder, i.e., the vowels and affixes. Suffixes are removed by moving inwards from the right word boundary, and other rules are applied by moving from left to right through the word. In each of the passes, the word is scanned, and the appropriate ordered set of rules for that pass is tried until a match in contexts is found.

Application of the KTH rules is accomplished in one pass through the set of rules. If a rule context matches anywhere in the word, moving left to right, the conversion is made, and the next rule context is compared. This method appears to be much more efficient, and does not require the program code needed in the MIT method to direct the various passes with the appropriate set of rules. In fact, no new code was written for the English system at all: the code existent for the Swedish system serves for the English rules as well.

The major difference between the multi-pass method and this one-pass procedure is in the manner of processing and ordering affixes. Recognition and removal of all affixes as a first step in the MIT algorithm corresponds to less than ten rules in the KTH system which recognize vocalic inflectional suffixes and insert a morph boundary marked with the feature "inflectional." The effect of not recognizing all affixes before consonant conversion appears to be rather small: initial consonant clusters after unrecognized prefixes have been observed to be mispronounced in a few cases in the KTH system. However, the opposite effect may be observed in the MIT system: strings incorrectly recognized as prefixes before application of the consonant rules also lead to mistaken pronunciations.

There is a significant difference in the ordering of suffix rules in the two algorithms. Suffixes in the MIT algorithm are recognized first and converted later (in any order). In the one-pass system, however, suffixes must be listed in the order of their probable occurrence from the right-hand side of the word so that their word-final or morph-final position is verified. A short study was undertaken for the purpose of determining the proper order.

There are several other differences in the processing of affixes. Because all consonants are converted before the recognition of most affixes in the KTH algorithm, those consonants in affixes are also converted. The KTH set therefore contains a few rules which are necessary in order to recognize suffixes containing consonants with multiple pronunciations, e.g., the suffix ic in electric or electricity. Suffixes whose final letter may undergo a spelling change are also listed in two rules. The feature of compatability of parts-of-speech in a compound suffix which is found in the MIT algorithm, has not been implemented in the KTH system. This feature is well-developed, but is not frequently needed, and would require additional code and a table of parts of speech for suffixes.

A number of other differences in the two sets of rules are due to the objective of expressing all rules in the KTH system in the higher-level programming language. The most important difference is in the lexical stress rules, which, in the MIT system, are embedded in code. The KTH rules are expressed in the rule language, and are applied using the same formalism as that used for the grapheme-to-phoneme rules. A rule cycle has not been implemented, but the effect of the cycle has, for the most part, been captured in the rules.

Special stress effects due to suffixation are accomplished in two ways. Stress-carrying suffixes are pre-stressed in the suffix rules by noting primary or secondary stress as a feature of the appropriate vowel. This stress may be adjusted later by the stress rules themselves. Suffixes which have no effect on the stress cycle are preceded by a suffix boundary marker with the feature "minus stress cycle." This feature is also assigned to word boundary symbols such as "space," and "period," and becomes part of the right context in many stress rules.

Unlike the MIT system, the KTH rules provide no device with which to retain graphemes after their conversion to phonemes. The retention of graphemes in the MIT system provides for the specification of either letters or phonemes in both left and right contexts. As a consequence, a substantial subset of rules differ in specification of context. The KTH rules have not yet been tested on a large set of data, but it is believed that this difference gives neither set of rules an advantage worthy of note.

In addition, the KTH programming language allows each phoneme and punctuation mark to be expressed in terms of distinctive features. This type of specification makes the rules more "transparent" than those in the MIT program where variables are used. The facility of specifying optional elements in this programming language has also allowed rules to be expressed more succinctly in several cases.

The experience gained in writing English rules for the KTH system emphasizes the utility of the higher level programming language in which the rules are written. Future development for other languages is very much recommended.

References

1. R. Carlson and B. Granström. 1976. A Text-to-Speech System Based Entirely on Rules. Conf. Record, IEEE International Conf. on Acoustics, Speech, and Signal Processing, pp. 686-688, Philadelphia, Pa.
2. R. Carlson and B. Granström. 1975. Text-to-Speech Conversion by Ordered Rules. Presented at the Eighth International Congress of Phonetic Sciences, Leeds.
3. S. Hunnicutt. 1976. Phonological Rules for a Text-to-Speech System. AJCL Microfiche 57.
4. J. Allen and S. Hunnicutt; R. Carlson and B. Granström. 1979. MITalk-79: The 1979 MIT Text-to-Speech System. Speech Communication Papers, ASA-50, pp. 507-510, The Acoustical Society of America, New York.
5. J. Liljencrants. 1968. The OVE III Speech Synthesizer. IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, pp. 137-140.
6. R. Carlson and B. Granström. 1975. A Phonetically Oriented Programming Language for Rule Description of Speech. Speech Communication, Vol. 2, Almqvist & Wiksell, Stockholm.

THE TERMS INTENSIVE/EXTENSIVE IN HJELMSLEV'S THEORY OF LANGUAGE

John Jørgensen
Institute of Phonetics, University of Copenhagen

1 Introduction

In works on phonological (and phonetic) theory of more recent date the notion of markedness is often referred to as an important phenomenon. In general, different writers do not agree in detail on what this notion is taken to mean - nevertheless, certain characteristics are most often connected to one of the series of the oppositional pairs as opposed to the other. The marked members of such oppositional pairs are by most investigators said to be 1) more complex, 2) absent in positions where only one of the members occurs, 3) less frequent - both in a given text of a given language and in the phoneme systems of the languages of the world. Examples are numerous and probably well known to the readers, so I shall limit myself to mention only an extreme one: If a language has two series of stops: one produced with egressive airstream (A), and another produced with ingressive airstream (B), then the (B)-series is considered to be the marked one. Members of this series are more complex, absent in positions of neutralization, and found only in a few of the world's languages. There are other instances, however, where it is more difficult to decide whether a given segment is to be characterized as marked or unmarked. As an example it could be mentioned that /s/ is generally described as unmarked compared to /c/, in spite of the fact that the former is the more complex of the two segments - at least from a phonetic point of view. Segments may be categorized as marked or unmarked depending upon the priority given to the criteria chosen for categorization. For example, should the criteria for categorization be language specific or universal, should they be based primarily on form or on substance?

2 First appearance of the notion intensive/extensive

If form is chosen as the primary and most reliable aspect of language, the name of Louis Hjelmslev comes inevitably to mind. In his theory of language (glossematics) substance is almost completely neglected. Only formal phenomena (i.e. relations between the elements of language) are considered. In Hjelmslev's terminology the terms marked/unmarked do not exist. The glossematic distinction which to some degree corresponds to marked/unmarked is termed intensive/extensive.

To get a deeper understanding of this notion it seems relevant to go back to the works where Hjelmslev introduced it, viz. the book on case (1935) and his paper on linguistic relations in general (1933)[1]. It should be emphasized that in these early works the notion intensive/extensive was used as a device in the description of grammatical relations, and it was introduced in the preglossematic period.

According to Hjelmslev, the description of case is always related to one primary parameter or dimension, viz. directionality, which can be represented schematically as follows:

+	0	-
closeness	rest or neutralness	remoteness

or in graphical form as in fig. 1:

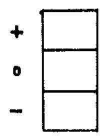


Figure 1. Graphical representation of the directionality dimension.

The term neutral (0 in the figure) may, however, have the following two interpretations: 1) a given case is termed 0 because it is neither + nor -, or 2) a case is termed 0 because it is indifferently + or - or 0. Thus we are faced with a new sort of opposition, the so-called participative opposition. The "normal" exclusive opposition (known from logics) is of the form: A vs. anything but A. The participative opposition has the following form: A vs. anything else (including A).

---

1) This paper was not published until 1973.

Linguistic oppositions are, according to Hjelmslev, most often participative and can be represented graphically as follows:

Figure 2. Graphical representation of the participative opposition. X is intensive and Y is extensive.

In such oppositions x is called the intensive member (characterized as precise and well specified), y is called the extensive member (vague and unspecified). In the field of grammar, examples of participative oppositions are numerous:

	y		x	
tense	present	vs.	preterit	(Danish)
mood	indicative	vs.	conjunctive	(German)
number	singular	vs.	plural	(German)

In the field of semantics examples of participation are easily found, too:

	y		x	
adjective	big	vs.	little	(English)
adjective	old	vs.	young	(English)
noun	man	vs.	woman	(English)

3 Phonological applications

An explicit application of the intensive/extensive distinction in the field of phonology is not found in the preglossematic period. When it is used in the glossematic period, the point of view has been changed in such a way that the formal relations implied by the distinction are used in "defining" (i.e. constituting) the single units (e.g. the units corresponding - more or less - to phonemes). The terms intensive/extensive can now only be used in instances of what is generally called neutralization, since it is only in such instances that it can be proved that the extensive member covers the whole zone. As a phonological example stop consonants in German may be pointed at. Because of the neutralization in final position, p t k are described as extensive, as opposed to b d g, which are intensive.

Basing our assumptions on the presentation in Hjelmslev (1937, 1948, 1951), we now proceed to the application of the theory on real data. Parts of the French and Danish consonant systems are taken as examples.

3.1 The French consonant system

The first step of the procedure is to establish the categories. On the basis of position in the syllable (including ability to enter into clusters) the following 4 categories are set up: 1) only initial, not in clusters, 2) initial and final, not in clusters, 3) initial and final, always vowel adjacent, 4) the rest. We may focus on category 4 and, leaving out (for the sake of simplicity) the problems of the horizontal dimension, we find the following configuration (Hjelmslev, 1948):

α	p	t	f	k	s
A	b	d	v	g	z

Figure 3. Graphical representation of "category 4" (see text). (Part of the French consonant system).

According to Hjelmslev, the lower series (indicated by the capital A) is the extensive one. If neutralizations are taken into account, however, it turns out that the t/d and k/g oppositions are neutralized under dominance of "liaison" and realized as t and k, respectively. This implies (contrary to what Hjelmslev indicates) that t and k are extensive. The opposition f/v may equally be neutralized (at least in the word 'neuf') under the dominance of "liaison" - implying (in agreement with what is indicated by Hjelmslev) that v is extensive. The analysis of the s/z opposition is simplified by Hjelmslev, since for all instances of final, latent s, he represents it as z in the "underlying form". Thereby the neutralization of the s/z opposition under dominance of "liaison" is eliminated. However, this procedure is problematic for adjectives ending in a final latent s, e.g. 'las', since such words in their "underlying form" should differ from the corresponding feminine form only by the absence of a final schwa. The last consonant pair to be considered is p/b for which pair there are no instances of neutralization. There are, it is true, a few instances of "liaison" with latent p but none with a latent b. Consequently they should be characterized as contensive. If the critical remarks and alternative propositions given above concerning the consonant system are accepted, we may suggest the following modified system:

p		t	v	k	z	A
b		d	f	g	s	α

Figure 4. Revised arrangement of "category 4". The contensive b/p opposition is shown separately.

3.2 The Danish consonant system

Another kind of problem turns up when an attempt is made to apply the intensive/extensive distinction to the Danish consonant system - such an analysis is made by Hjelmslev (1951).

At first sight it might seem evident that the Danish stop consonants fulfil the requirements for being categorized according to the intensive/extensive distinction because of the neutralization in final position (and before schwa). But the distinction becomes inapplicable, because Hjelmslev - in accordance with his principle of "greatest possible reduction of the inventory", which is of primary importance to him - reduces the Danish stop consonant system from 6 to 3 "units" (plus 3 "units" consisting of stop + h). There may still be said to be neutralization - not between p t k and b d g but between h and 0.

4 Concluding remarks

In conclusion the following points can be made:
1) An analysis based on the intensive/extensive distinction may give results different from those achieved by a "normal" markedness analysis. 2) The benefit of employing the intensive/ extensive distinction may be minimized by the priority given to other procedures of analysis, e.g. reduction of the inventory.

Finally, it may be hypothesized that one of the reasons for the many problems involved in the analysis of the intensive/ extensive parameter in the field of phonology is that phonological oppositions are more often exclusive than participative. In the field of grammar the situation may well be the opposite.

References

Hjelmslev Louis 1935. La catégorie des cas. Acta Jutlandica VII.1. Aarhus.

Hjelmslev Louis 1933. Corrélations morphématiques. Essais linguistiques II. Travaux de Cercle Linguistique de Copenhague XIV, 1973.

Hjelmslev Louis 1937. On the principles of phonematics. Essais linguistiques II. Travaux de Cercle Linguistique de Copenhague XIV, 1973.

Hjelmslev Louis 1948. Le système d'expression du français moderne. Bulletin du Cercle Linguistique de Copenhague 1941-1966.

Hjelmslev Louis 1951. Grundtræk af det danske udtrykssystem med særligt henblik på stødet. Selskab for nordisk filologis årsberetning for 1948.

INVERSE FILTERING

Inger Karlsson
Institutionen för talöverföring, Tekniska Högskolan, Stockholm

Definition: Inverse filtering means that the formants of the speech sound are damped out by filters so that only the voice source signal remains.

My interest in this area stems from my so far unsuccessful attempts to synthesise a female voice. In an earlier study (Karlsson, 1979), I varied all the different parameters that separate male and female voices according to published data, but obtained no definitely human and female voice. As very little data about the voice source is available, my suspicions are that much of the secret is hidden there. The published studies of the voice source mainly pertain to descriptions of methods.

At present considerable research into the voice source and different methods of inverse filtering is being done at the Inst. of Talöverföring: G. Fant (1979) is constructing hard ware filters to be used on speech recorded with an FM tape recorder. In an FM recording the phase of the signal is retained and frequencies down to DC are recorded, the same signal can accordingly be filtered with many different filter settings. G. Fant is using his results to get a mathematical description of the voice source, especially in word endings, stress and phoneme boundaries. J. Sundberg and J. Gauffin (1978) are investigating the voice source of singers and untrained, normal subjects with the aid of the mask for measuring the volume velocity waveform at the mouth constructed by M. Rothenberg (1973). Finally J. Liljencrants has written a computer program for inverse filtering described by him as "an OVE III with formant anti filters replacing formant filters and the speech signal as voice source." This method, as well as G. Fant's, allows the same utterance to be inverse filtered many times and, furthermore, it is possible to alter the frequencies of the anti filters in the computer program every millisecond. I have just started to use this program but have so far no results to present.

In a short study from which I am going to mention some results, I have been using the same experimental set up of the Rothenberg mask as used by Sundberg and Gauffin (1978). Five females and one male participated in the experiment and were asked to repeat the syllable /pɑ:/ with different F_0: high, medium, and low within their normal register and with different voice levels: weak, moderate, and loud (not whisper or scream). Figure 1 shows one speaker's voice pulses for different voice levels at the same F_0. The material is too small to allow for a description of different voice types but I have tried to see how the different parameters defined by Sundberg and Gauffin and by Fant (see figure 2) vary with F_0 and voice level. The only manifest correlation between parameters I have found is between sound pressure level and the derivative at the instant of closing: the peak flow A_G, divided by the offset time T_d. The correlation coefficient between these two parameters varies for the different subjects between 0.83 and 0.96. The two parameters for one subject are depicted in figure 3. Most of the subjects also show a correlation between F_0 and the "base band" formant, $1/T_G$ according to figure 2.

This is only a preliminary study, which I have made principally to get to know the technique. I will now work mainly on getting descriptions of different voices.

References

Fant G. (1979). Glottal source and excitation analysis. STL-QPSR 1/1979, 85-107.

Fant G. (1979). Vocal source analysis - a progress report. STL-QPSR 3-4/1979, 15-24.

Karlsson I. (1979). Försök till utvärdering av akustiska skillnader mellan mans- och kvinnoröster. Könsroller i språket. Rapport nr 75 från avd. för forskning och utbildning i modern svenska (FUMS) 1979, Inst. för nordiska språk vid Uppsala Universitet.

Rothenberg M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J. Acoust.Soc.Am. 53, 1632-1645.

Sundberg J. and J. Gauffin (1978). Waveform and spectrum of the glottal voice source. STL-QPSR 2-3/1978, 35-50.

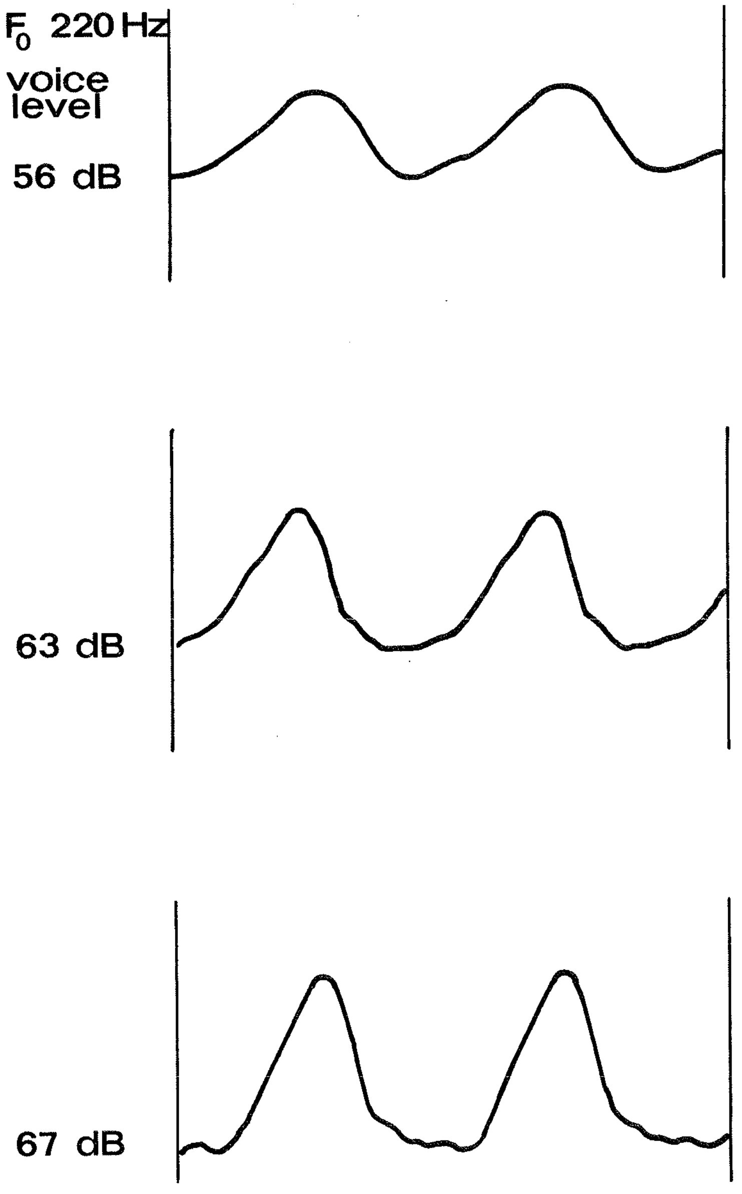


Figure 1. Voice source pulses of the same F_0 for one subject. The voice level is for the top curve 56 dB, middle curve 63 dB, and bottom curve 67 dB.

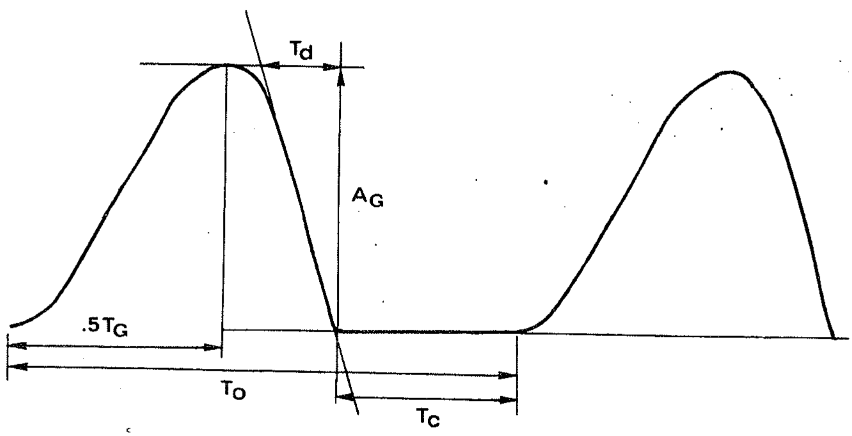


Figure 2. Different parameters used in descriptions of the voice source pulses: T_0 = the duration of the pulse. T_G = two times the pulse rise time, T_C = the closed interval of the pulse, T_d = offset time; A_G = the peak flow through the glottis.

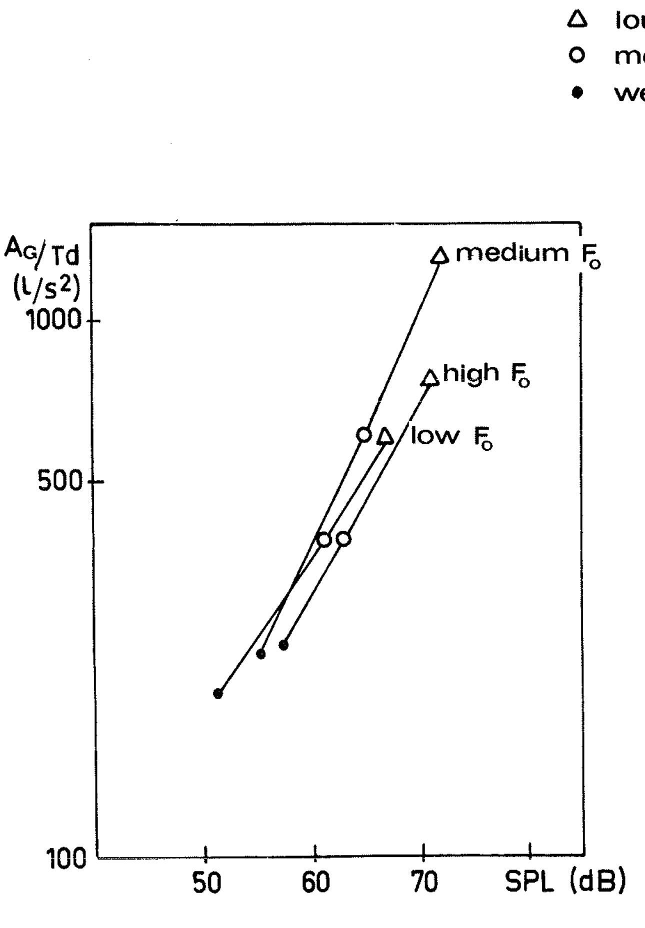


Figure 3. The relation between sound pressure level, SPL, and the derivative of the air flow at the instance of closing for one subject.

RHYME OR REASON?
ON LANGUAGE DISTURBED CHILDREN'S RHYMING

Eva Magnusson
Department of linguistics and phonetics, Lund

Rhyming is something that most children at one time of their development are fascinated by. What I have in mind are rhymes like "Humpty Dumpty sat on the wall.

Humpty Dumpty had a great fall.
All the king's horses,
All the king's men
couldn't put Humpty together again".

Learnt conventional rhymes like these are not the only kind that children use and enjoy. They also seem to take advantage of the discovery that they can produce an unlimited number of rhymes of their own invention, an ability they make frequent use of in games, dialogues, teasing etc.

Language disturbed children, on the other hand, are remarkably insensitive to rhymes. This is a fact I have often observed when working as a speech pathologist with such children. In a nursery rhyme or rhymed story, normal children easily fill in the left out rhyming words while language disturbed children have difficulties in doing so, even if they are familiar with the story and it has been read to them several times. If they suggest a word, it is in most cases more or less appropriate in the semantic context but it is most likely not a rhyming word. Why then is rhyming difficult for these children? What is it that they cannot do?

What we do in rhyming is to separate the prevocalic element(s) of the stressed syllable from the rest of the syllable or the word and to use what is left of the syllable or the word as a model when producing new rhymes. To do this requires

- ability to segment within the syllable, to segment phonemically,
- ability to identify segments as vowels and consonants, in order to be able to make the delimitations in the correct places,
- ability to identify the stressed syllable,

- knowledge of the segmental order in the sequence.

It has been suggested by Moskowitz (1971) and Waterson (1971) among others, that children early in language acquisition use the syllable or a larger unit like the word as their basic phonological unit. Data suggest that this may be the case for at least some language disturbed children as well and that the disability in rhyming found in these children may correlate with their prolonged use of the syllable. They would thus not be able to participate successfully in an activity like rhyming which requires an ability to segment within the syllable.

Some support for this is given by Savin (1972) who claims that children, "normal middle-class children", do not learn to segment phonemically until after the age of five. By the age of five, most children have acquired the main part of the phonological rules of the language as evidenced by their speech which is by then easily intelligible. Five-year-old language delayed children do not have the same control of the phonological rules. This leads to the question whether there is a correlation between rhyming and level of phonological development.

Some of the forms produced by language disturbed children differ from the normal forms. When this is the case, on which forms do the children make their rhyming operations, on their own produced forms or on the normal forms?

Subjects

The subjects are 28 children, aged 3;9 to 6;6 years, with the diagnosis retardatio loquendi idiopathica. The diagnosis means, among other things, that there is no easily identified etiology for the disorder, that psycho-motor and social development is roughly normal, and that there is no diagnosed neurological dysfunctions. Hearing is normal as shown by tone audiometry. In this group, I have studied the children´s speech production, their ability to make auditive discrimination in their own and in other people´s speech as well as their performance on a rhyming task.

Procedure

Eight sets of pictures with three pictures in each were used (see table 1). Two of the pictures in each triplet represent words that rhyme, and the third is used as a distractor. The distractors consist of words which have the same prevocalic and sometimes the same vocalic segments as one of the rhyming words, or which have a strong semantic association with one of the rhymes, as in the triplet gran-kran-bada (fir-tap-bathe).

Table 1. Test material

åtta råtta äta	(eight rat eat)	pil bil boll	(arrow car ball)
hår får fot	(hair sheep foot)	kran gran bada	(tap fir bathe)
såg tåg tår	(saw train toes)	klocka docka flicka	(watch doll girl)
sol stol skor	(sun chair shoes)	hatt katt kam	(hat cat comb)

The principle of rhyming was demonstrated to the children. Those, who did not seem to understand the meaning of the word "rhyme", were told that their task was to select the two pictures out of three that "sound alike at the end". The test triplets were then introduced in conversation. I named the pictures and tried to discourage the children from naming or repeating the words. They were then asked to respond by selecting what they thought were the two rhyming pictures in each of the eight triplets.

Pretesting

Before starting the main study, I tried out the test material on four children with normal speech, aged four to six years, in order to ensure that children of this age could perform the kind of task required. None of the four children had difficulties in understanding the task and they picked out the rhyme pairs without any hesitation. Some of them also produced new rhymes spontaneously, rhymes both with and without a semantic content.

Results and discussion

If the children had merely made chance choices, the expected distribution of correct answers per individual would have been as shown in fig.1. One child would have made no correct answers, four children would have made one correct answer, eight children two correct answers etc. The distribution of the children's correct choices observed in this study differs markedly from the expected distribution as can be seen in fig. 1. The children who made six or more correct answers have apparently used a rhyming strategy, since no child making exclusively chance choices was expected to make more than five

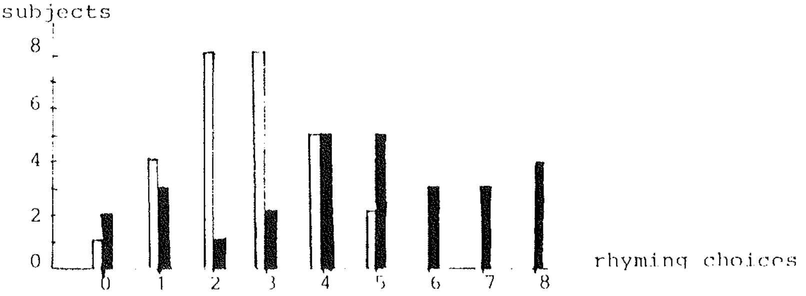


Fig.1. Expected distribution of random choices (white column) and distribution of actual choices (filled column).

correct answers. It is however a matter for discussion which criterion ought to be used to categorize a child as a rhymer. If a criterion of five correct choices is used, this would be equivalent to a chance level of 10 percent and thus allow a certain amount of uncertainty. Ten children are classed as good rhymers if a criterion of six or more correct answers is used and fifteen if the criterion is five or more correct answers.

Rhyming and level of phonological development

My next question is concerned with the relationship between rhyming and level of phonological development. Phonological developmental level or in this case rather degree of deviance, was assessed in the following way. A rating system with numerical values was used. The children's speech production was analysed in terms of linguistic simplification. Each process was assigned a numerical value in relation to its propagation in the system, to the number of possible contexts where it is actually applied and to the frequency of application in possible contexts for each child. Processes fequently used early in children's language acquisition were assigned a low figure and processes prevalent later in development were assigned a high figure. This gives a system where each child's degree of deviance is indicated by a figure. These values must however be treated with caution since they are based on ratings and they are employed here not as an exact measure of deviance but only as a rough estimate of the phonological level.

Fig.2. shows the relation between rhyming and degree of phonological deviance. As can be seen, the range of variation is considerable. Some children with a high degree of deviance can rhyme and some with a nearly normal speech can not. The correlation for the whole group is -0.31.

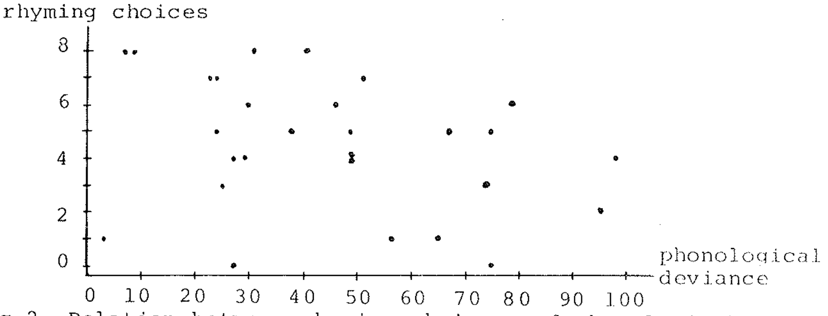


Fig.2. Relation between rhyming choices and phonological deviance.

In the rhyming group, where the criterion is either five or six correct choices, the tendency is toward a negative correlation, -0.60 for the group with six or more correct choices. The tendency is that the good rhymers have a lower degree of deviance or a more normal speech than the poor rhymers.

In the non-rhyming group, there is no such relation as the correlation is -0.01 for the group with five or less correct choices. Several hypotheses are possible. The non-rhyming group consists of children who are not able to rhyme, or of children who are not able to handle rhyming tasks of this particular kind or of a combination of both.

In order to test this, some of the children were excluded from the non-rhyming group, namely those children whose degree of phonological deviance was the same or lower than the mean value for the rhyming group. Even so, the correlation was -0.20. A possible interpretation is that there is more variation in phonological development in the non-rhyming group than in the rhyming group.

These results are in agreement with other findings. Studies dealing with phonetic segmentation and early reading acquisition indicate that all normally speaking children are not able to segment phonemically or to rhyme (Liberman et als,1977, Savin, 1972). On the other hand, normal speech is not necessary for an understanding of the rhyming principle (Curtiss,1977).

Error analysis and representational form

When the children did not choose the two rhyming words as a pair, were their choices totally random or were they made according to some other principle?

One of the original hypotheses was that language disturbed children make semantic choices. It appeared, however, that this was hard to test. The children rarely chose pairs which have an obvious semantic association by adult standards, but it can not be excluded that some of their choices were made on semantic grounds nevertheless. Since they were not asked to motivate their choices, it is difficult to decide whether a semantic strategy was used or not.

In their erroneous choices some children prefer pairs that have identical initial consonants. More children choose for-får as a pair than fot-hår, tår-tåg is a more likely choice than tår-såg. This tendency is even stronger, if it is assumed that the children compared their own produced forms and not the normal forms that they heard. One boy's performance may illustrate this. He made no correct rhyming choices and in three cases he indicated all words in the triplets as rhymes. On the other hand, if his results are analysed with the assumption that he made his choices on the basis of identical initial consonants in his produced forms, this accounts for six of his choices. Furthermore, in one case, he said that all the words were different, which they were in his production.

Can children who choose words with identical initial consonants segment within the syllable? One possibility is that they compare syllables as wholes and that the initial resemblance is sufficient for their decision which might then be based on similarity of syllables and not on identity of parts of syllables. Another possibility is that they are able to segment within the syllable but have insufficient knowledge of the order in the sequence.

Choices based on identical initial consonants result in a rhyming pair of words even if it is not the kind of rhyme that the children in this study were instructed to make. The children in this case made alliteration instead of end or full rhymes. Historically, alliteration is an older kind of rhyme and one of the kinds that occur in old Icelandic poetry as in the Poetic Edda (Oldberg, 1945, Hallberg, 1970). End rhymes did not appear until later and it has been suggested that they originate from the older kind of rhymes such as alliteration. The same kind of development may be hypothesized for children, so that alliteration is mastered before end rhymes. This hypothetical ordering is supported by the observation that alliteration choices were more frequent among the good rhymers than in the non end rhyming group implying that children first acquire an ability to segment within the syllable and only later become aware of the sequential ordering.

In conculsion it can be said that children who are able to rhyme have a lower degree of phonological deviation or a more normal speech. But rhyming is also possible for individuals with deviant speech production. A more normal speech, a better knowledge of the phonological rules of the language, does not necessarily involve an ability to rhyme, to segment phonemically. Something else and more is needed than the control of phonological rules as it is shown by speech production. One possibility is that rhyming has a closer connection with perceptual than with productive ability and that perceptual competence is more developed in the deviant speakers who rhyme than in those who do not. An alternative explanation is that rhyming has to do with such vaguely defined notions as linguistic awareness or metalinguistic ability and if so the control of phonological rules in perception and production is of minor interest.

References

Curtiss S. 1977. Genie. A psycholinguistic study of a modern-day "wild-child". Academic Press, New York

Hallberg P. 1975. Litterär teori och stilistik. Akademiförlaget, Göteborg

Liberman I.Y., D. Shankweiler, A.M.Liberman, C.Fowler, and F. W. Fischer. 1977. Phonetic segmentation and recoding in the beginning reader. in Towards a psychology of reading:

Proceedings of the CUNY conference. Eds. Reber and Scarborough. Hillsdale, New Jersey

Moskowitz A. I. 1971. The acquisition of phonology. Dissertation, University of California, Berkeley

Oldberg R. 1945. En bok om rim. C.W.K.Gleerups förlag, Lund

Savin H.B. 1972. What the child knows about speech when he starts to read. in Language by ear and by eye. Eds. Kavanagh and Mattingly. The M.I.T. Press

Waterson N. 1971. Child phonology: a prosodic view. Journal of Linguistics 7, 170-221

SYNTAGMATIC AND PARADIGMATIC RELATIONS IN DYSPHONOLOGY

Nettelbladt Dalby Ulrika
Institutionen för lingvistik, fonetik

Theoretical background

Research in child phonology received its first impetus from the work of Roman Jakobson (1941/1968).

His theory on the orderliness and universality of phonological development provided a fruitful framework within which the rather amorphous data from earlier studies of children's language could be interpreted.

Jakobson explicitly stated that the route of phonological development is in general the same for all children, with a successive unfolding of phonemic contrasts from the most maximal contrast into finer ones.

This has led researchers to concentrate on the order in which children acquire particular contrasts, and their results have provided norms of phonological development, useful for clinical practice.

In the 70's interest turned from universal aspects of phonological development to differences in the phonological development of individual children.

Instead of stating the order in which the child acquires phonemic contrasts, Ferguson (1976), describes different strategies used by different children in developing a phonological system. By means of this type of description researchers try to get a deeper insight into the underlying processes which guide language development.

Another change in focus concerns the linguistic entities which are considered basic and relevant for analysis.

Moskowitz (1971) and Menn (1978) have modified the theory of Jakobson by introducing other and larger entities as fundamental in early phonology.

According to Moskowitz the child develops from a stage where intonation frames form the basic organizational units of phonology, to a syllabic stage characterized by reduplications or single open syllables.

Menn introduces yet another concept, the canonical word form, to characterize the same stage. At the end of the babbling stage the child formalizes certain segmental combinations into recurrent canonical word forms. Two factors are seen to interact in the emergence of new and more complex word forms: the child's increasing control over articulation, and his/her need to use linguistic contrasts in order to be understood.

Not until later, around 2 - 2½ years of age, according to both Moskowitz and Menn, does the child reach a stage where phonemic contrasts are of primary importance.

These new developments in the theory of child phonology have proven very useful for describing the phonology of language disabled children.

Aims of the investigation

The starting point for the investigation reported on here has been clinical practice with language disabled children. In almost all cases of language disability there is a concomitant dysphonology of varying degree and quality.

With some dysphonologic children traditional methods of analysis and treatment have been found insufficient, and thus the question of whether the dysphonology in such children might be qualitatively different from other types of dysphonology arose.

The first problem for the investigation was: Can different subgroups of dysphonologic children be distinguished? The second problem was: Do dysphonologic children develop in similar or different ways compared to normal children?

Subjects and material

Ten South-Swedish children with dysphonologies of various degrees were chosen as subjects (mean age 5,5, range 4,3-7,0 six boys and four girls).

(A preliminary investigation of 32 children preceeded the present one.)

Single words were elicited by a picture-naming test and running speech was elicited by asking the children to tell a story about sequences of pictures.

The tapes were transcribed according to narrow phonetic and child-language-modified transcription systems (Bush et al, 1973).

Follow-up studies have been made for all children at ½-year intervals for 2 to 3 years.

Analysis

Traditional clinical practice analyzes dysphonology in terms of consonant substitution patterns. Among others Lorentz (1976) argues that this approach to dysphonology is insufficient. Other types of simplifications (e.g., harmony restrictions) also have to be taken into account to reach a deeper under-standing of dysphonology.

In the present work the following aspects are considered:

a) Distribution of phonemic contrasts, substitution patterns of consonants and vowels and presence of diphthongization.

b) Idiosyncratic vowel and consonant harmony restrictions and other types of phonotactic restrictions on permitted word forms, e.g. deletion of unstressed syllables and reduction of clusters.

c) Some prosodic aspects relevant for early phonological development are also taken into account: development of stress patterns and of word accents.

Results

In discussing the results of the investigation I find it useful to use the dichotomy of paradigmatic versus syntagmatic relations. Two main types of dysphonology are found:

1. Paradigmatic group. This is characterized by paradigmatic substitutions only. 5 children belong to this group, (3 boys, 2 girls, mean age 5,8, range 4,10 - 7,0).

Within this group no syntagmatic restrictions are found except for consonant cluster reductions. The phonological simplifications are exclusively of a substitution type.

Formalized as rules, these could be described as context-free, i.e., the substitutions are clearly predictable irrespective of word context. Productions of specific sound: or words are fairly consistent.

The substitutions mostly affect consonants. Certain consonant contrasts are cancelled in the child's production.

The most typical simplifications in the paradigmatic group are the following: reduction of consonant clusters, dentalization of velars and stopping of fricatives.

2. Syntagmatic-paradigmatic group. Simplifications both in the syntagmatic and the paradigmatic dimensions are characteristic of this group. The dysphonology in this group is considered more serious than in the first group. 5 children belong to this group (3 boys, 2 girls, mean age 5,4, range 4,3 - 6,7).

Several phonemic contrasts are collapsed in the child's production by substitutions, some of which are similar to those of the paradigmatic groups, e.g. dentalization.

The effect of the substitutions is variable, however, and due among other things to strong harmony conditions or to the use of a restricted number of canonical word forms. Formalized as rules these restrictions could be labelled context-sensitive rules or alternatively as strong surface phonetic constraints, (Shibatani, 1971).

Although the production of individual lexical items often varies segmentally, canonical forms are stable within the same recording.

Two subgroups can be differentiated within the syntagmatic-paradigmatic group:

a) children with the most extreme type of dysphonology show very strict canonical word forms, e.g. reduplications only. A typical prosodic feature is equal stress assignment.

b) children with a less extreme type have strong harmony conditions on vowels or consonants. Most typical is an anticipatory, non-contiguous consonant assimilation. Equal stress is replaced by an overgeneralization of word accent 2 (e.g. "`fågel", bird, "`springer", runs) in this group, except for words with late stress (e.g., "ba'nan", banana), where the initial syllable is deleted. In general polysyllabic words are strongly reduced.

Discussion

The second type of dysphonology closely resembles the phonology of young children 1 - 2 years of age. (Detailed studies of early phonology are presented by Compton & Streeter, 1977, for example.)

Instead of subdividing the children into two different groups, it might be advantageous to see both types of dysphonology as points on a common line of development, where the more serious type of dysphonology would represent an early stage of phonological development.

For two reasons, however, it seems preferable to view the two groups as distinct from each other.

The follow-up studies show that three of the five children with syntagmatic restrictions retain their syntagmatic dependence, in spite of some development along the paradigmatic dimension. Newly acquired contrasts also become involved in assimilatory processes, and when trying to pronounce 'tongue-twisters' these children easily relapse into earlier patterns of strong word form restrictions.

Another reason to differentiate the two groups is the implications for clinical work. The syntagmatic-paradigmatic group clearly represents a more serious type of dysphonology with a poor prognosis. In all of these cases we find concomitant dysgrammatism. In four of the five syntagmatic-paradigmatic cases we also find symptoms of minimal brain dysfunction which might explain the severity of their language disability.

The present investigation shows that strong syntagmatic restrictions need to be identified as risk-symptoms indicating a poor prognosis, while exclusively paradigmatic simplifications usually disappear in time.

Distinguishing among these two groups demands a revision of current logopedic treatment.

References

Bush, C, Edwards, M, Luckau, J, Stoel, C, Macken, M & Petersen, J, 1973: On specifying a system for transcribing consonants in child language. Unpublished. Child language project, Committee on Linguistics. Stanford University.

Compton, A & Streeter, M. 1977: Studies of early child phonology: data collection and preliminary analyses. Papers & Reports on Child Language

Development, 13. Stanford University.

Ferguson, C. 1976: Learning to pronounce: the earliest stages of phonological development in the child. Papers & Reports on Child Language Development, 11. Stanford University.

Jakobson, R. 1941/1968 (English translation): Child language, aphasia and phonological universals. Mouton.

Lorentz, J. 1976: An analysis of some deviant phonological rules of English. In Morehead, D & Morehead, A: Normal and deficient child language. University Park Press.

Menn, L. 1978: Pattern, control and contrast in beginning speech. A case study in the development of word form and word function. Indiana University Linguistics Club.

Moskowitz, A. 1971: Acquisition of phonology. Unpublished PhD thesis. Berkeley.

Shibatani, M. 1973: The role of surface phonetic constraints in generative phonology. Language 49: 87 - 106.

ANALYSIS AND PREDICTION OF DIFFERENCE LIMEN DATA FOR FORMANT FREQUENCIES. THE TESTING OF TWO DISTANCE MEASURES

Lennart Nord
Dept. of Speech Communication, KTH, Stockholm

Introduction [1]

The aim of this study is to relate perceptual distances drawn from difference limen data to a physical distance measure, taking into account some facts about the hearing mechanism. The distance measure and its correlation to the perceptual distance can be regarded as a way of testing the underlying perception model.

We used sound stimuli of the type Flanagan (1955) used in his study of difference limen for formant frequencies. Six groups of four-formant synthetic vowels were produced with the series synthesizer OVE III. In each group, F1 or F2 was systematically shifted in seven steps of 10 or 25 Hz above and below the reference vowels (see Table 1).

Table 1. Reference sounds for the stimulus sets

no	1	2	3	4	5	6	
F_4	3550	./.	./.	./.	./.	./.	(Hz)
F_3	2500	./.	./.	./.	./.	./.	
F_2	1500	./.	./.	1000	1500	2000	
F_1	300	500	700	500	./.	./.	

In an AB test, listeners judged whether they could discriminate between the reference sound and a test sound.

A number of distance measures based on a more or less refined processing of static speech sound spectra have been presented during the last years. The correlation between perceptual data and distance measures that are solely based on peripheral auditory processing may give poor results in identification tests which involve higher levels of processing. Hopefully, a better correlation will be reached if the perceptual data are taken from discrimination tests where less phonetic processing will play a role for the listener's judgement. This was the reason for choosing this particular set of stimuli.

Distance Measures

In the present study, distance measures based on two types of auditory representations of the sound stimuli were tested.

Plomp (1970) formulated a distance measure in a study of the timbre of complex tones. Basically, the distance between two steady state complex tones, having the same loudness, pitch and duration was defined as the distance between two points in an m-dimensional space where the m coordinates correspond to the intensity differences in m 1/3-octave bands (cf. the critical band theory of hearing, Zwicker, 1961). Plomp found a good correlation between the distance measure and dissimilarity indices for musical instrument tones, vowels, and pipe organ stops. In Lindblom (1978), the measure is tested on speech-like material and good results were likewise obtained.

Schroeder et al (1980) proposed a spectral distance measure based on the masking properties of auditory perception. The auditory spectrum is transformed into loudness per critical band (sone/Bark) as compared to Plomp's model which does not include masking effects, but represents the spectrum as intensity per critical band (dB/Bark). For further details, see Plomp (1970) and Schroeder et al (1980).

Lindblom and Bladon (1980) tested the "Schroeder et el" model in a discrimination test with synthetic vowels. Carlson and Granström (1979) examined a number of auditory models and tested them with a variety of different speech materials. As one of their auditory representations of the signal spectra, they adopted the "Schroeder et al" model but introduced the equal loudness curves of hearing in the transformation from intensity to loudness.

In the present study, the Carlson-Granström version of the sone/Bark representation of spectra was used.[2]

At least two ways of calculating the difference between two transformed spectra seem reasonable. Either a simple summation of the differences between spectra at every sample point (in every critical band) or a geometric summation, that is, the square root of the sum of the squared differences. The first distance is usually referred to as the city block distance, the second gives a Euclidean distance.

The correlation between the discrimination test data and the distance measures was evaluated for the six groups of vowel-like stimuli. A linear regression was assumed, but due to the usual s-shape of the discrimination curve sample points with discrimination values below 5 % and above 95 % were excluded from the regression analysis.

Results

The results from the DL test confirmed as a whole the results obtained by Flanagan (1955) with a DL value of 3-5 % of the first and second formant frequency.

However, when looking at details of the discrimination curves for the six sets of vowel stimuli, some deviations appeared. Asymmetries of the curves reported by Flanagan, that is, cases in which the listeners were more sensitive to a shift of a formant in one direction than in the other, did not show up in the expected way in the present study. As one example, the stimulus set with the reference formant values 500, 2000, 2500, 3550 Hz, gave asymmetric discrimination data results in the opposite direction compared to the Flanagan study. Our listeners were more sensitive to an F2-lowering than to a raising.

The explanation for the asymmetry obtained in the Flanagan study seems intuitively correct. When two formants approach each other, an increase of the F2F3 complex would effect the listener's response towards being more sensitive to an F2 raising than to an F2 lowering. A closer look at the local intensity change when F2 is raised 25 and 50 Hz, however, will not show any drastic increase. Informal tests with native English and American listeners (though with a knowledge of Swedish) showed the same asymmetries as for our Swedish listeners, which seems to rule out an explanation in terms of different phonemic border within the stimulus set. Presently, we have no explanation for the different asymmetries.

The correlation between the perceptual data and the distance measures proved to be high (r=0.8-0.9) for both models. The intensity per critical band representation gave a somewhat but not signficiantly better fit than the loudness per critical band representation. A distance metric using the Euclidean distance did not differ appreciably from the city block distance.

A notable result was that also the discrimination test data for the stimulus sets with the marked asymmetry discussed above, correlated well with the distance measure, thus giving some support to the auditory models that were tested.

However, any distance measure based on models which try to capture the peripheral auditory processing will not be able to predict listeners' responses as soon as the decision involves a phonetically based judgement, or, in other words, a more centrally located processing.

In this study it was thus not possible to achieve a single physical distance value as a transformed version of the Difference Limen for formant frequencies, independent of formant number, vowel spectrum etc.

For a discussion of the relative merits of different auditory models, see Carlson & Granström (1979).

Footnotes

[1] Part of this work has been reported elsewhere. For a more detailed presentation, see Nord & Sventelius (1979).

[2] The use of the computer programs developed by Carlson & Granström for the auditory models, distance measure calculations, and the repression analysis is thankfully acknowledged.

References

Bladon R.A.W. and B.E.F. Lindblom, 1980, Modelling the judgement of vowel quality differences. Submitted for publication.

Carlson R. and B. Granström. 1979. Model predictions of vowel dissimilarity. STL-QPSR 3-4/1979, 84-104.

Flanagan J.L.. 1955. A difference limen for vowel formant frequency J.Acoust.Soc.Am., 613-617.

Lindblom, B.E.F.. 1978. Phonetic aspects of linguistic explanation. Studia Linguistica XXXII, I-II, 137-153.

Nord L. and E. Sventelius. 1979. Analysis and prediction of difference limen data for formant frequencies. STL-QPSR 3-4/1979; also in PERILUS, Report I: Experiments in Speech Perception (eds. B. Lindblom & J. Lubker), 25-37.

Plomp R.. 1979. Timbre as a multidimensional attribute of complex tones. In Frequency Analysis and Periodicity Detection in Hearing (eds. R. Plomp & G.F. Smoorenburg). Sijthoff, Lieden, 397-414.

Schroeder M.R., B.S. Atal and J.L. Hall. 1979. Objective measure of certain speech signal degradations based on masking. In Frontiers of Speech Communication Research (eds. B. Lindblom & S. Öhman). Academic Press, London, 217-229.

Zwicker E.. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J.Acoust.Soc. Am. 284.

ASSIMILATION OF Fo

Niels Reinholt Petersen
Institute of Phonetics, University of Copenhagen

1 Introduction

The observed fundamental frequency course of an utterance can be viewed as the result of contributions from several simultaneous components. In Advanced Standard Copenhagen (ASC) Danish four such components have to be considered (see Thorsen, 1979): (1) a sentence component, which gives the intonation contour of the sentence, (2) a stress group component, which supplies the Fo movements of the stress groups (a stress group in ASC Danish is constituted by a stressed syllable plus the following unstressed ones), (3) - in words with 'stød' - a stød component, and (4) a segmental (or microprosodic) component, which gives the Fo variation attributable to the segments constituting the utterance, such as inherent Fo differences and coarticulatory effects on Fo across segment boundaries.

Since it is a consequence of inherent properties of the speech production system and cannot be voluntarily controlled by the speaker, the segmentally determined Fo variation cannot - unlike the variation accounted for by components 1, 2, and 3 - carry linguistically relevant prosodic information. On the contrary, if sufficiently large the segmentally determined Fo variation could be expected to interfere with and possibly distort the contributions of the linguistically relevant components. The present paper is intended to examine the interaction between the segmental component (with the emphasis on the inherent Fo differences between vowels) and one of the linguistically relevant components, viz. the stress group component.

The Fo pattern of the ASC stress group can be described as a relatively low stressed syllable followed by a high falling tail of unstressed syllables (Thorsen, 1979). The Fo rise from the stressed to the first post-tonic syllable varies between 3 and 0.5 semitones. These values are very similar to the inherent Fo differences between high and low vowels, which have been found by Reinholt Petersen (1978, 1979) to vary between 1 and 3 semitones.

In theory this might imply that the Fo contour of an utterance could be distorted in such a manner that the perceived stress pattern would differ from that intended by the speaker as a function of the qualities of the vowels in the utterance. Now, such things do not happen; people normally hear stress patterns as intended by the speaker. One explanation for this could be that the listener perceptually compensates for inherent Fo differences, and thus reconstructs the intended Fo contour. However, the inherent Fo differences are considerably larger in stressed than in unstressed syllables (Reinholt Petersen, 1979), and this means that the perceptual system would have to know the stress pattern in order to be able to select the appropriate correction factor for the reconstruction of the intended stress group contour. Another possibility, which is the one under investigation here, could be hypothesized, namely that the compensation takes place in the speech production system as a coarticulatory assimilation of Fo between syllables which more or less smoothes out the distortion and thereby preserves the intended Fo contour.

2 Experiment I

In experiment I the following questions were considered: (1) Can Fo in one syllable be influenced by high vs. low vowel (i.e. a vowel with high vs. low inherent Fo) in adjacent syllables? (2) Is the effect, if any, directional (i.e. is it the preceding or the following vowel that has the stronger effect)? (3) Are there differences between syllables in different positions in the stress group as to how strongly Fo is influenced?

2.1 Method

The material consisted of mVmVmVmV nonsense words in which the vowels i and ɑ and u and ɑ alternated in such a manner that it was possible to see how Fo in these vowels was influenced by i vs. ɑ and u vs. ɑ in preceding and following syllables in the stress group positions 1st pretonic (or more correctly, in the last unstressed syllable in the preceding stress group), stressed, 1st post-tonic, and 2nd post-tonic. The words - embedded in frame sentences - were read by two female (KM, SI) and two male (PA, NR) ASC speakers. Six repetitions of each word were obtained for each speaker.

2.2 Results

Fig. 1 shows the average Fo in semitones over all subjects as a function of high vs. low vowel in the preceding and the following syllables in the four stress group positions under investigation. It is seen from fig. 1 that Fo is fairly consistently higher after i and u than after a. The strongest effect is found in the first post-tonic syllable, where it approaches one semitone. The following vowel seems to have no consistent effect on Fo. Thus the present data may be taken as evidence for a progressive assimilation of Fo which tends to preserve the intended Fo contour, and particularly the Fo rise from the stressed to the first post-tonic syllable, although - it must be emphasized - the inherent difference of about two semitones in the

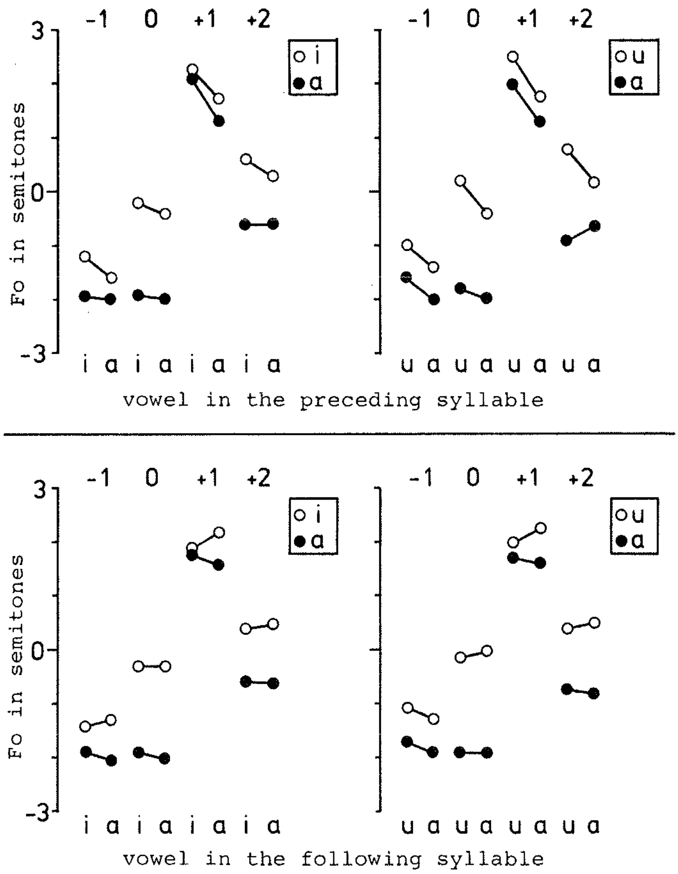


Figure 1. Effect of high vs. low vowel in preceding (upper graph) and following (lower graph) syllable on Fo in 1st pretonic (-1), stressed (0), 1st post-tonic (+1), and 2nd post-tonic (+2) syllable.

stressed syllable is compensated by less than one semitone in the first post-tonic.

3 Experiment II

The aim of experiment II was to see (1) whether an intrinsic Fo difference in the stressed syllable ascribed not to high vs. low vowel but to the initial consonant would influence Fo in the first post-tonic, and (2) if the effect of high vs. low vowel in the stressed syllable on Fo in the first post-tonic would be the same whether the consonant intervening between the two syllables was voiced (as in experiment I) or voiceless.

3.1 Method

The material consisted of nonsense words representing all 'CVCV combinations of C = m and f and V = i and ɑ. The words were inserted in carrier phrases and read by the same subjects as in experiment I. Six repetitions of each word were obtained per subject.

3.2 Results

A prerequisite for considering at all question (1) above was that initial f vs. m had an effect on Fo in the same syllable. In the present material Fo was found to be 0.9 semitones higher after f than after m in stressed syllables. This figure is comparable in magnitude to the Fo difference between i and ɑ, which was found to be 1.3 semitones in this material. On this basis the effects of i vs. ɑ and of f vs. m in the stressed syllable on Fo in the first post-tonic might also be expected to be of comparable magnitudes. This is not the case, however. From fig. 2 it appears that whereas Fo in the first post-tonic is higher

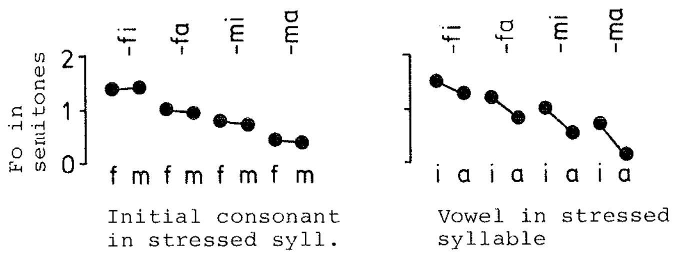


Figure 2. Effect of initial f vs. m (left) and i vs. ɑ (right) in the stressed syllable on Fo in first post-tonic -fi, -fɑ, -mi, and -mɑ syllables.

after i than after ɑ in the stressed syllable, the initial consonant in that syllable has only a very slight effect, if any at all, on Fo in the first post-tonic.

The effect of i vs. ɑ in the stressed syllable on Fo in the first post-tonic is present whether the consonant intervening between the two syllables is voiced or voiceless, although the effect seems to be slightly smaller across f than across m.

4 Discussion

As shown in the experiments reported above the effect of high vs. low vowel in the stressed syllable on Fo in the first post-tonic approaches one semitone. If an effect of that magnitude were to be accounted for by coarticulatory assimilation of tongue height in the first post-tonic vowel, a radical quality shift should be expected in that vowel as a function of the tongue height of the vowel in the stressed syllable. This applies whether the inherent Fo differences between vowels are to be explained by an acoustic source/tract coupling hypothesis or by a physiologically based tongue pull hypothesis (a review of these hypotheses is given eg. in Ohala, 1973). However, no such change is immediately audible, and measurements of F1 and F2 in ɑ in first post-tonic syllables of the material of experiment I show only a moderate effect on the formant frequencies from vowels in adjacent syllables (see fig. 3). Furthermore, the formant

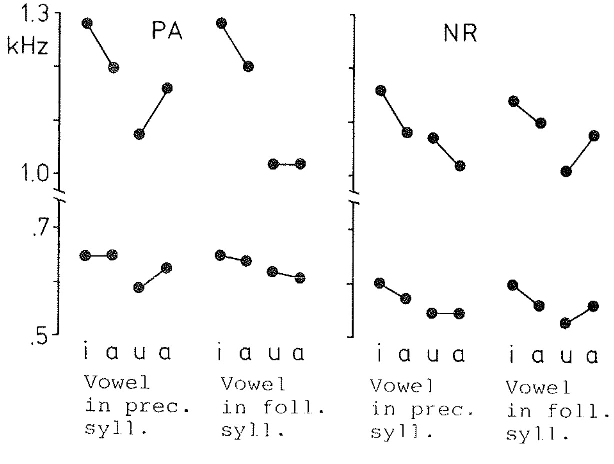


Figure 3. F1 and F2 in ɑ in first post-tonic syllables in the material of experiment I as a function of high vs. low vowel in the preceding and following syllables. Subjects PA and NR.

frequencies (particularly the second formant frequency) seem to be rather systematically influenced by the vowels both in the preceding and the following syllables.

Thus, there seems to be a discrepancy between the effect of neighbouring vowels on Fo and the effect of neighbouring vowels on the position of the tongue body, both as regards magnitude and direction. (This statement presupposes, of course, that formant frequency changes in the present material can be ascribed to tongue body position changes, which can be taken to be the case in the i/ɑ words, but not with safety in the u/ɑ words, where lip rounding is involved).

The discrepancy may be explained, however, if - under the tongue pull hypothesis - the relation between tongue height and vocal cord tension can be described in terms of a spring-mass system, where the driving force is the tongue body, the spring represents the connection between the tongue body and the laryngeal structures, and the mass represents vocal cord tension. If this description is tenable, changes in vocal cord tension, and hence in Fo, will occur with a delay relative to the changes in tongue height which produce them.

References

Ohala J.J. 1973. Explanations for the intrinsic pitch of vowels. Monthly Internal Memorandum, Phonology Laboratory University of California, Berkeley, 9-26

Reinholt Petersen N. 1978. Intrinsic fundamental frequency of Danish vowels. Journal of Phonetics 6, 177-189

Reinholt Petersen N. 1979. Variation in the inherent Fo level differences between vowels as a function of position in the utterance and in the stress group. Annual Report of the Institute of Phonetics University of Copenhagen 13, 27-57

Thorsen N. 1979. Lexical stress, emphasis for contrast, and sentence intonation in Advanced Standard Copenhagen Danish. Proceedings of the 9th International Congress of Phonetic Sciences. Vol 2, 417-423. Copenhagen

ETUDE COMPARATIVE DES VARIATIONS DE LA FREQUENCE FONDAMENTALE EN SUEDOIS ET EN FRANÇAIS.

Paul Touati
Institut de Phonétique,Lund.

Cette communication présente les premiers résultats d'une étude comparative des variations de la fréquence fondamentale dans des phrases simples du suédois et du français.Dix phrases suédoises(répétées cinq fois par un locuteur suédois)et dix phrases françaises(elles aussi répétées cinq fois par un locuteur français)ont servi de corpus.Le dépouillement des 100 tracés de la fréquence fondamentale a permis de dégager des contours-types.Le choix de l'étude comparative des variations de la fondamentale n'est pas linguistiquement aléatoire. Il se base sur la corrélation existante entre accent,entité linguistique et la fréquence fondamentale,marque substantielle de cet accent[1].Cette étude comparative se situe cependant dans une perspective pédagogique.Elle vise,dans un premier temps,à prédire des erreurs de prononciation(au niveau prosodique)de français parlant suédois et de suédois parlant français.

1.Procédure expérimentale

Notre but étant de réaliser une analyse comparative,nous avons donc chercher à mettre au point un corpus de phrases susceptibles d'être comparées.Nous avons selectionné des phrases très similaires de par leur signifié,de par leur composition segmentale et surtout de par leur accentuation (les unités lexicales suédoises sont accentuées,comme les françaises,sur la syllabe finale).L'énoncé de base du corpus est:"C'est+SN"(en suédois:"Det är+SN")comme p ex dans"C'est un café" = "Det är ett kafé".A partir de cet énoncé de base nous avons procédé à une double extention,d'une part une extention syllabique de l'unité lexicale(café devenant canapé, puis monographie etc..)et d'autre part à une extention lexicale(la phrase de base ayant été ainsi transformée,par cette adjonction,en deux groupes prosodiques,un groupe initial et et un groupe final)."C'est un café" devient donc "C'est un café de Paris".Les phrases de l'extention syllabique composent l'ensemble E1,celles de l'extention lexicale l'ensemble E2.Les deux locuteurs masculins viennent du sud de leur pays respectif(Grenoble et Lund). Sile locuteur français parle un français standard,le locuteur suédois parle,en revanche,le scannien qui est une variante dialectale dont les caractéristiques prosodiques ont été spécifiées,en relation avec d'autres variantes,dans un modèle intonatif proposé par Bruce G. et Gårding E.(1978).L'enregistrement a été réalisé en chambre sourde à l'institut de Phonétique de Lund.Chaque locuteur a donc lu cinq fois les dix phrases présentées sur fiches.Chaque fois l'ordre de présentation a été différent et décidé de manière aléatoire.Les tracés ont été obtenus àl'aide d'un "Fundamental Frequency Meter"(Type FFM 650).Après dépouillement de tous les tracés,certains contours-types ont été retenus.

2.Résultats

Nous commencerons avec l'examen des contours du suédois. Les phrases de l'ensemble E1 se caractérisent par une légère montée(10 Hz) à partir d'une fréquence d'attaque extrêmement régulière(105 Hz)qui permet au locuteur scannien d'atteindre un plateau sur lequel il se maintient tout au long de la réalisation des syllabes atones.Arrivé à la syllabe accentuée(qui dans ce cas porte l'accent de mot et de phrase) le locuteur réalise une montée rapide(de 30 à 40 Hz)qui lui permet d'atteindre le sommet intonatif de l'énoncé.Cette montée est alors suivie d'une descente très marquée(50 à 60 Hz)(fig.1b par exemple).La stratégie intonative du locuteur suédois pour cet ensemble de phrases est relativement simple:elle consiste à se maintenir pendant les syllabes atones sur un plateau et à produire,sous la syllabe accentuée, une variation tonale bien marquée(montée-descente).Les phrases de l'ensemble E2 se subdivisent en deux groupes prosodiques.Le groupe initial reproduit en grande partie les variations de E1.Le groupe final,plus court,emploit les syllabes atones comme préparation de la montée tonale de la syllabe accentuée pour atteindre un sommet intonatif(plus bas que celui du groupe initial).La syllabe accentuée est alors réalisée par une rapide chute tonale(fig.1c).

Dans les phrases francaises de l'ensemble E1,il semble évident qu'il faille séparer les énoncés courts du type "C'est un café" (fig.2a) des énoncés longs du type "C'est une manipulation" (fig.2b).En effet dans le cas de l'énoncé court,le contour-type est de forme triangulaire avec un sommet intonatif sous la syllabe prétonique.On observe en revanche que plus l'énoncé devient long,plus la prétonique descend et plus une montée apparait clairement à l'initiale de l'unité lexicale.J'interpreterai cette différence de stratégie comme une plus grande intégration de l'unité lexicale à l'accentuation de phrase dans le cas des énoncés courts et comme une définition de cette unité en tant que groupe prosodique dans le cas des énoncés longs.L'ensemble E2 diffère de E1 dans la mesure où l'extention lexicale a opéré une division de l'énoncé en deux groupes prosodiques de niveau inférieur;ceci a pour résultat de donner à la syllabe finale du groupe initial un contour montant,celle du groupe final ayant un contour descendant.Les montées apparaissent aussi dans cette ensemble à l'initiale des unités lexicales(fig.2c).Ceci viendrait corroborer l'idée d'une stratégie chez le locuteur français,lequel à l'aide d'une invariable(la montée initiale) et d'une variable(la montée ou la descente finale) marque les limites des groupes prosodiques.

3.Analyse comparative et conclusion

Cette description des variations de la fréquence fondamentale de phrases similaires suédoises et francaises nous permet d'établir une comparaison qui met clairement à jour deux différences.La première est la présence en francais(versus absence en suédois) d'une montée initiale au début du mot prosodique[2].La deuxième différence nous semble plus importante.Elle concerne la réalisation de la syllabe accentuée finale.Celle-ci est réalisée en francais soit par une montée, soit par une chute tonale alors qu'en suédois,elle exige une variation tonale bien marquée[3] et double(montée-descente)(fig. 2a,2b,2c et 1a,1b,1c).Une petite matrice nous permettra de mieux visualiser ces différences phonétiques:

	Montée initiale	Variation marquée
Suédois	—	+
Francais	+	—

Ma conclusion prendra la forme de prédictions.Un locuteur français qui prononcerait les phrases suédoises risque de trop marquer l'initial du groupe prosodique par une montée tonale bien nette,la réalisation de la syllabe accentuée risque,en revanche,d'etre limitée à une seule montée ou chute tonale.Quant au locuteur suédois,il y a de grandes chances pour que sa prononciation des phrases françaises passe sous silence la montée initiale et manifeste l'accent final par une variation tonale trop marquée.Il conviendrait naturellement de vérifier ces prédictions par une analyse d'erreurs des productions tonales françaises d'un locuteur suédois(et vice versa).Une synthèse qui introduirait,sur la base des différences constatées,l'accent suédois dans des phrases francaises et l'accent français dans des phrases suédoises pourrait nous renseigner sur la valeur,pour la perception,de ces deux différences de production. Il suffirait pour cela d'employer ces différences comme variables dans une série de tests de perception.Ces deux procédures de vérification seront l'objet de recherches ultérieures.

Notes

1.Cf.Fonagy I.:"J'entend par accent l'entité linguistique(prosodique) ayant pour fonction principale la mise en relief d'une syllabe et dont la substance consiste dans un plus grand effort expiratoire et articulatoire...La courbe de fréquence fondamentale qui réagit sensiblement aux changements de pressions sous-glottique et supra glottique est le meilleur indicateur de l'effort."(Fonagy I. 1979).

2.S'il s'avère possible,dans une description du système prosodique du français,de se limiter aux seuls contours mélodiques finaux,comme le propose p ex Martin Ph.(1978),il me semble,dans le cadre d'une analyse comparative,plus judicieux de conserver comme éléments de comparaison tous les éléments qui établissent une différence entre les systèmes comparés.

3. Cette différence semble confirmer la réflexion que fait Fonagy(1979)lors qu'il écrit:"En tachant de déterminer la place de l'accent dans la phrase,on a l'impression que l'accent français est souvent moins marqué,plus"fuyant"que dans les autres langues romanes,dans les langues germaniques et slaves ou dans le hongrois."

Références

Bruce G.et Gårding E. 1978. A prosodic typology for Swedish dialects.Travaux de l'Institut de Linguistique de Lund XIII 219-228.

Fonagy I. 1979.L'accent français:accent probalitaire.Studia Phonética 15.Didier.Montréal,Paris,Bruxelles.123-233.

Martin Ph.1978.L'INTONATION:aspects linguistiques et reconnaissance de la parole.8ièmes JEP,G.A.L.F.,Aix-en-Provence.

Det är ett kafé

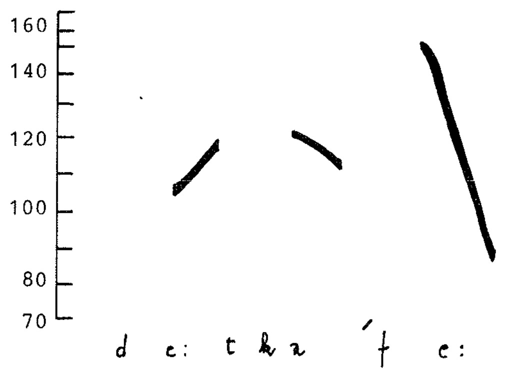


Fig.1a

Det är en manipulation

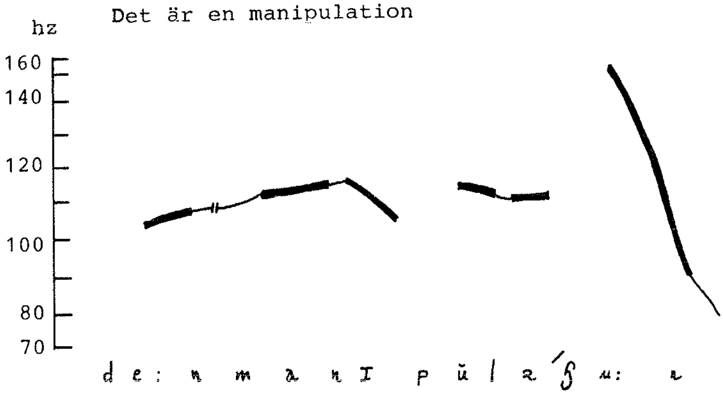


Fig.1b

Det är en manipulation från Marie

Fig.1c

C'est un café

Fig.2a

C'est une manipulation

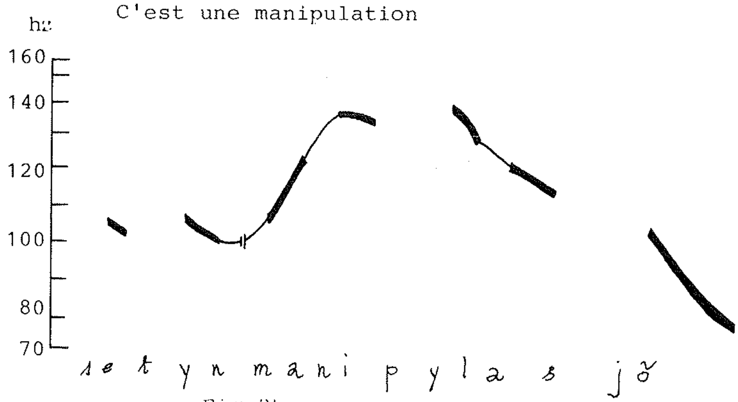


Fig.2b

C'est une manipulation de Marie

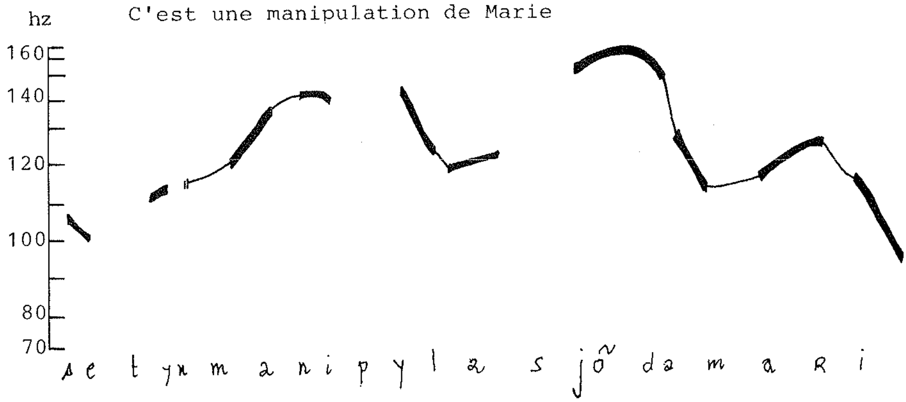


Fig.2c

AN INVESTIGATION OF PROSODY IN SOME DIALECTS OF SMÅLAND

Ursula Willstedt
Institutionen för lingvistik, Lund

Introduction

The framework on which this study is based is the model for describing and comparing the prosody of Swedish dialects developed by E. Gårding & G. Bruce, as part of the project "Swedish Prosody". Four geographical areas served as the basis for their study, represented by the dialects of Malmö (designated as 1A), Dalarna (1B), Stockholm (2A) and Gothenburg (2B). The principle has been to use a small number of variables to derive and generate the intonation patterns of the different dialects. The model consist of three parts:

1. A linguistic part with dialect-dependent representations of the manifestations of word accents and sentence accents (Fig.1).
2. An algorithmic part with dialect-independent rules which, together with the information in parts 1 and 3, generate the intonation patterns.
3. A set of conventions and prescriptions for the application of the rules of the algorithm.

The word accents are represented by a high and a low frequency point (the stars in figure 1.) These occur at different points in time depending on the dialect, but accent 1 (A1) always occurs earlier than accent 2 (A2). Sentence accent (SA) is manifested as a widening of the frequency range, either in the focused word itself (cf. Malmö and Dalarna) or after it (cf. Stockholm and Gothenburg). For a more detailed description of the model cf. Bruce & Gårding, 1978.

The purpose of the investigation that I will describe here was to analyze transitional dialects lying somewhere in between the prototypes for 1A and 2A. Småland is a suitable area for finding such dialects (cf. Fig. 2).

Materials

The test sentences I have used were a subset of those developed and studied by G. Bruce (1977). These were of three different types:

Man vill a´namma nåra ´längre ´nummer (accent 1 only)

'{we / one} want(s) to adopt some longer numbers'

Man vill `lämna nåra `långa `nunnor (accent 2)

'{we / one} want(s) to leave some long nuns'[1]

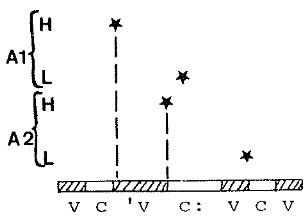
Man vill `lämna nåra `långanummer (compound with accent 2)

'{we / one} want(s) to leave some Långa-numbers'[1]

Informants

The informants who participated in the study were twelve high school (gymnasie) students studying in Växjö and Älmhult. All were female except for one informant from Växjö. Their home towns lie for the most part in southern Småland, as shown in Fig. 2.

Växjö dialect (Fig 3.)



Wide interval after A

SA H after A

Fig. 4.

Fig. 4 shows the data for the Växjö dialect as they are represented in the linguistic part of the model. If we compare the data for the manifestation of word accents in Växjö with those for Gothenburg (as in Fig. 1), we find that the positions of the high points for accent 1 and accent 2 are strikingly similar. The low points occur approximately one segment later in the Växjö dialect. In other words, it seems as though the word accent fall is somewhat less steep in Växjö than in Gothenburg. But after normalizing and comparing the data for additional informants (to be presented in a later report) from both Växjö and Gothenburg, I found that there was no evidence for such a relationship between the two dialects' word accent manifestations. On the other hand, the high point of accent 2 in Gothenburg and the other dialects in the 2B area seems to occur somewhat earlier in the stressed vowel than the model predicts. Consequently, for area 2B (Gothenburg), I will be using a high point for accent 2 which lies approximately 10% (of the vowel duration) from the end of the vowel segment. The timing of the other turning points, both for word accent and sentence accent manifestations, is in agreement with the model's predictions for the Gothenburg intonation pattern.

Sentence accent is manifested in the same way as in Stockholm (2A), i.e., with a rapid rise after the word accent, followed by a plateau up to the postfocal word accent. When there is no postfocal accent, there is still a sharp rise after the word accent, occurring in the final vowel.

Lessebo dialect

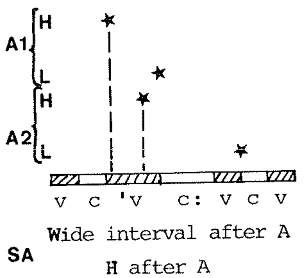


Fig. 5.

We can see from the data presented in Fig. 5 that the word accent manifestations in the Lessebo dialect, for both word accents, agree well with my interpretation of the Gothenburg pattern.

Sentence accent is consistently manifested as in Växjö, except for compounds. While the compounds in the Växjö dialect showed a pattern similar to that of Stockholm, with a rise in the secondary-stressed vowel, Lessebo shows a sharp rise in the final vowel, as in Gothenburg.

Väckelsång dialect

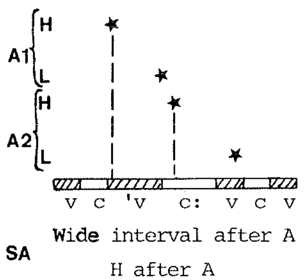


Fig. 6.

Accent 1 shows the same pattern as in Gothenburg, Växjö and Lessebo, while accent 2 is more similar to a Scanian Malmö pattern. Sentence accent is manifested as in Växjö.

Möckeln dialect (Fig. 7.)

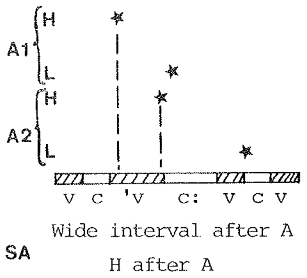


Fig. 8.

This speaker shows strong similarities to the Växjö dialect with regard to both word accent and sentence accent manifestation. An important difference, however, is that in compounds there is no rise in the secondary-stressed vowel in the Möckeln dialect, but rather a rise in the final vowel. The manifestation of focal accent 1 appears to be variable. The stressed vowel can be either rising or falling (cf., for example, the final focus position in Fig. 7. A similar variation was seen in a speaker from Växjö who is not presented here.)

Bökeberga dialect (Fig. 9.)

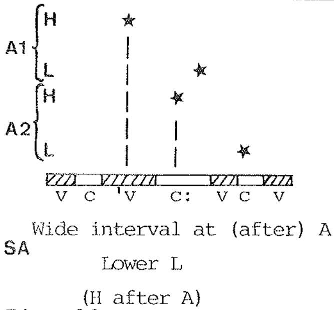


Fig. 10.

This informant, who is the sole northern Scania area representative in the material, differs greatly from all the others. Her word accent patterns are the same as those of the Malmö dialect. The manifestation of sentence accent is somewhat harder to describe: In final position, focus (sentence accent) is manifested by a rise in the final vowel, as in Småland especially for accent 2. Focus in earlier positions has a lower focal turning point which can, but need not be followed by a rise. Thus there is always a wide frequency interval within the focal word, and occasionally after it as well.

Discussion and summary

Although the data-base is somewhat limited, it does suggest a number of characteristic features of the prosody of the dialects of Småland. The manifestation of word accent, for nonfocal accent 1, as well as the manifestation of sentence accent, with a sharp rise after the word accent, seems to be characteristic for these speakers. A final accent 2 word in focus has a rise in the final vowel.

Compounds in Småland are characterized by varying patterns; They either have a rise in the final vowel (as in Möckeln and Lessebo) or a rise in the vowel carrying secondary stress (as in Växjö and Väckelsång). It is interesting to note that the only major difference found between Växjö and Möckeln concerned the manifestation of sentence accent for final compounds.

Compounds in Bökeberga, which is situated in the northern part of Scania (that is, just outside of Småland), show no change in F_0 after the stressed syllable. This pattern agrees well with the prototype for the 1A area (Malmö).

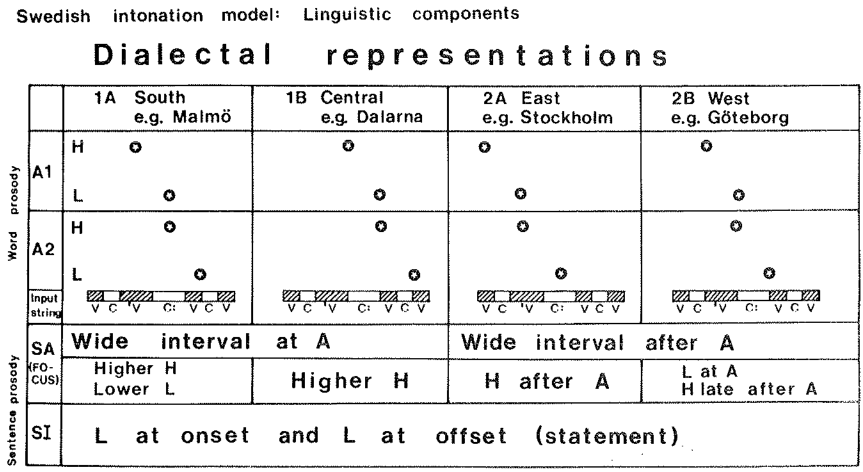


Fig.1. Prosodic variables in the model for Swedish intonation patterns. From Bruce & Gårding 1978.

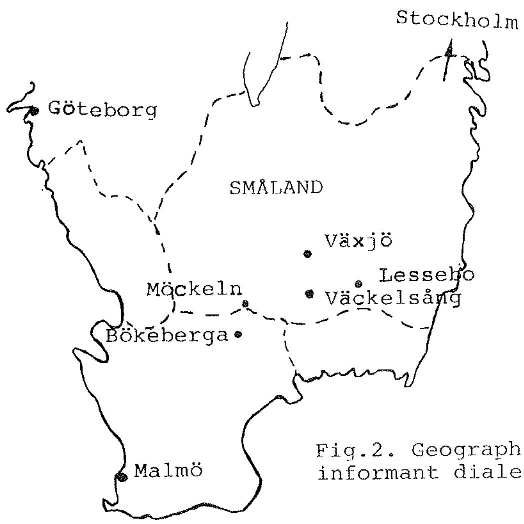


Fig.2. Geographical distribution of informant dialects.

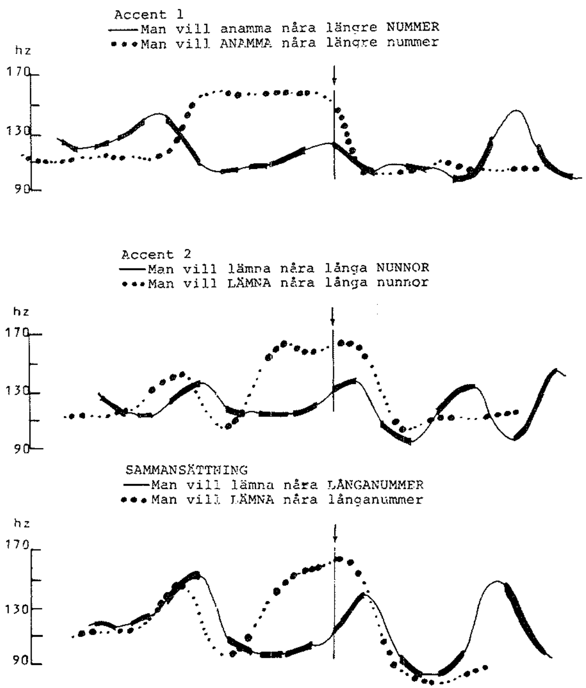


Fig.3. Representative patterns for Växjö*.

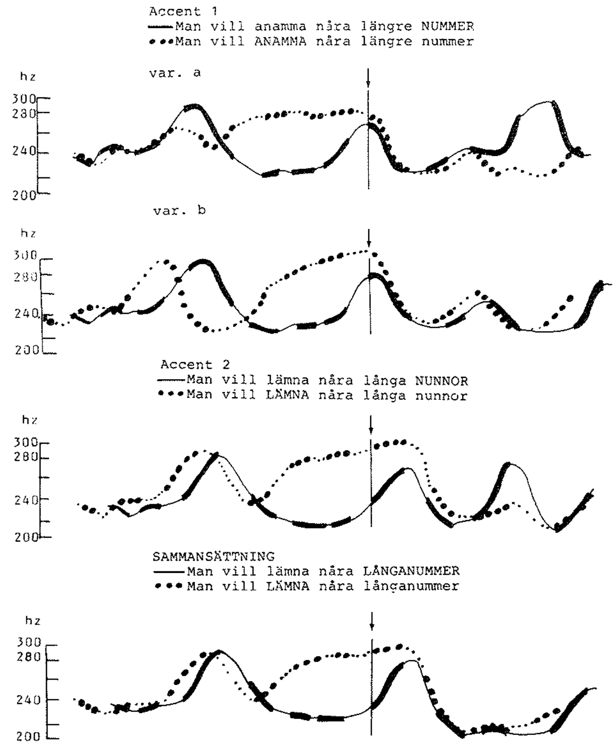


Fig.8. Representative patterns for Möckeln*.

* Thicker portions of the curves represent vowel segment, thinner portions represent consonant segments.

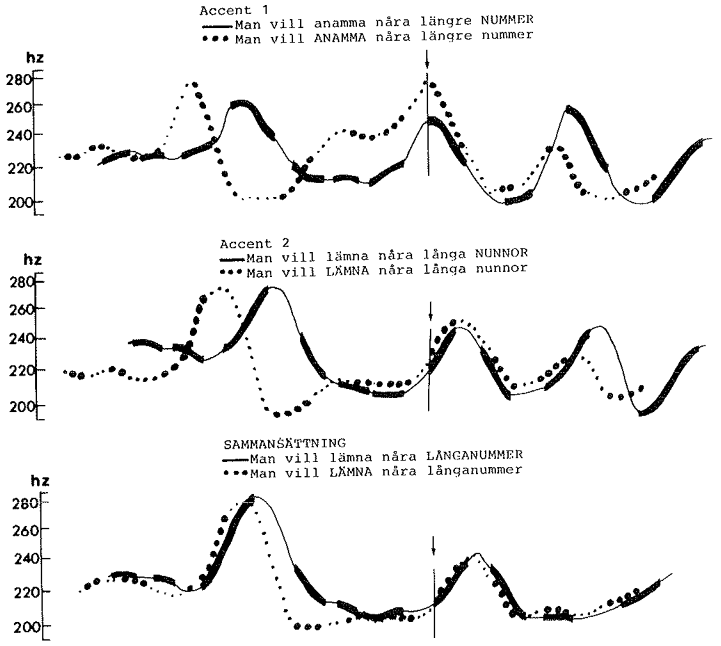


Fig. 9. Representative patterns for Bökeberga*.

Footnote

1. *Långa nunnor* was established as a new kind of pastry analogic to *munkar*, (lit: 'munks') 'bismarks'. *Långanummer* are numbers from the town of Långa.

References

Bruce G. 1977: Swedish word accents in sentence perspective. Gleerup Lund.

Bruce G. & Gårding E. 1979: (1978) A prosodic typology for Swedish dialects. Ur Nordic prosody, Travaux XIII.

Gårding E. 1973: Toward a prosodic typology for Swedish dialects. Proc. 2nd Int Conf. of Nordic and General Linguistics (utg. Dahlstedt).

Willstedt U. 1980: En undersökning av prosodin i några småländska dialekter. Stencil, Institute of linguistics, Lund University.

COMPUTER RESYNTHESIS OF SPEECH
ON PHONETIC PRINCIPLES

L. Nordstrand and S.E.G. Öhman
Dept of Linguistics, Uppsala

Introduction

In Zetterlund, Nordstrand and Engstrand (1978) and Engstrand, Nordstrand and Zetterlund (1979) it was demonstrated how the prosody (fundamental frequency contour and physical segment durations) of a tape recording of spoken utterances can be manipulated using LPC methods without much affecting the voice characteristics of the speaker or the phonetic identity of the so-called 'segmental' features of the utterance[1]. Using this method it has been possible to perform certain phonetic experiments that lend support to the theory that intonation in Swedish (and probably in many other languages also) is determined by certain (relative) pitch levels that must be reached simultaneously with certain definite nonprosodic sound effects (primarily the vowels) produced during the course of the speech act. On this assumption it appears to be possible to predict the durations of the physical segments from a knowledge of 1) the sequence of phonemes[2] and 2) the relative pitch levels of the phonemes that carry such levels. This should be a basic principle - in so far it holds - for any synthesis-by-rule system.

To carry the phonetic experiments further one should now like to be able to manipulate the parameters of the 'segmental' sound effects, also. I.e. , one should like, in a given recording of a certain utterance, to be able to change the sense-discriminating features[3] of certain occurrences, in that utterance, of various segmental phonemes. It would of course also be desirable to be able to manipulate other kinds of feature, e.g. expressive features, and, in fact, any kind of phonetically relevant feature.

It is not at all obvious, however, how segmental parameters can be controlled in an LPC analysis/synthesis system. The problem can be solved on certain conditions, however. First, one has to find a one-to-one transformation from the filter coefficients to a set of parameters whose relation to the synthesized acoustic phonetic effect is explicitly known. Second, in order to make the method easy to use, there should be included a process of interpolation along the time axis between given phonetic sound effects in a phoneme string. Third, to insure that the intermediate sound effects generated by this interpolation procedure do not strike the ear as independent 'intruding phonemes', yet another one-to-one transformation has to be found, one that maps the filter coefficients into a further parameter set.

What parameters to control

Theory as well as experience with formant synthesis motivate us to use spectral parameters for control and manipulation purposes. The transformation that maps the LPC coefficients on the spectral domain is defined by the fundamental theorem of algebra, relating coefficients and roots of a polynomial equation. LPC formants and bandwidths are defined through the roots of the complex polynomial whose coefficients are equal to the LPC coefficients. Efficient algorithms to solve for the roots of the polynomials are available in standard computer systems, and the converse computation of coefficients from roots can be performed by means of a simple well-known formula.

How to interpolate

Any interpolation method to be used must meet two requirements:

(1) it must guarantee that the synthesis filters are stable,

(2) it must not introduce 'intruding phonemes'.

The parameter set that first comes to mind as a possible candidate for these purposes is the set of filter coefficients themselves, i.e. , the LPC coefficients, as they are also called. Linear interpolation of these parameters does not guarantee stability, however [Markel & Gray (1976)] . On the other hand, linear interpolation of the roots (i.e. formant frequencies and bandwidths) of the LPC polynomial always guarantees stability, but this method entails a sorting problem (the so-called formant tracking problem) since the roots in question do not possess any obvious partial ordering. A number of other methods have recently been worked out (see e.g. Markel

& Gray -76), and the parameter sets defined by these transformations are tabulated in Table 1 where data relating to stability are also given. Fig 1 shows how the poles move in the z-plane under linear interpolation of the various parameter sets (in each case using six filter coefficients).

On the basis of the above discussion and informal experiments linear interpolation has been tried using area coefficients, log area coefficients and arc sine of the reflexion coefficients. It was found that synthesis based on interpolation of the arc sine of the reflexion coefficients gave a quality superior to the other methods. Probably this is due to the fact that the arc sine transformation produces a more even distribution of the reflexion coefficients near their peak values, i.e. , when these values are close to 1.

Using this method the following phonetic experiment was performed. The phonetically crucial instants of time in a given utterance was marked with help of a display of the pressure/time wave form. These moments were picked in accordance with phonetic principles sketched in Öhman et al. (1979). LPC coefficients were computed at the crucial instants and stored along with the relevant time coordinate. With this as input data the gaps between the crucial instants were filled with interpolated LPC coefficient sets, so that a new speech wave could be resynthesized.

The signal thus obtained sounded suprisingly similar to the original. One will recognize the voice of the speaker, his dialect, and, of course, what he says without much more distortion than a simple LPC analysis-and-resynthesis (without intervening parameter manipulations) will cause.

Recordings of these tests were played at the conference (see Fig 2).

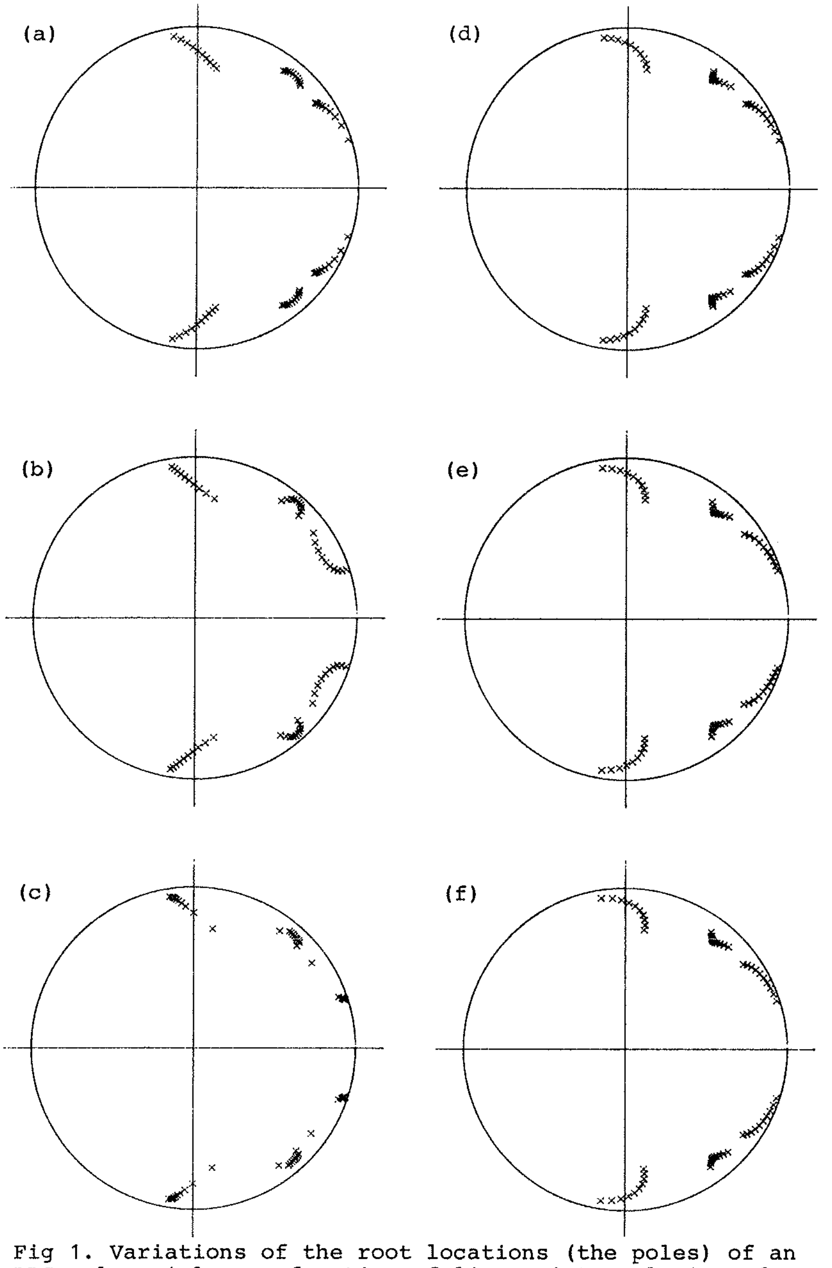


Fig 1. Variations of the root locations (the poles) of an LPC polynomial as a function of linear interpolation of a) filtercoefficients, b) cepstral coefficients, c) autocorrelation, d) reflection coefficients, e) area functions and f) arcsin of reflection coefficients.

Table 1. The various parameter sets based upon the LPC formulation.

Name	Stability
filter coefficients	No
cepstral coefficients	Yes
autocorrelation	Yes
reflection coefficients	Yes
area functions	Yes
arcsin of reflection coeficients	Yes

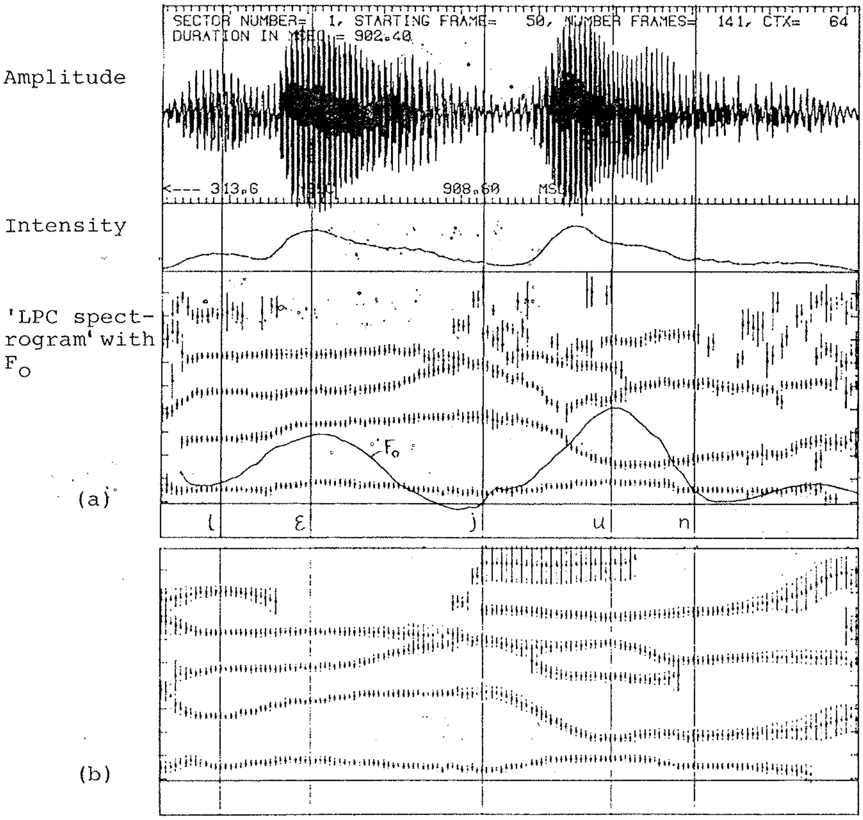


Fig 2. Amplitude, intensity, F_o and 'LPC spectrogram' as a function of time, of the word "lejon" (lion) in the test utterance. a) analysed 'LPC spectrogram', b) interpolated 'LPC spectrogram'. The 'crucial instants of time' are marked with vertical lines.

Notes

1. It is this process of analysis followed by manipulation of analysis parameters followed by synthesis from the parapeters thus obtained that is referred to by the word Resynthesis of the title of this paper.
2. The concept of phoneme intended here is sketched briefly in Öhman, Zetterlund, Nordstrand and Engstrand (1979).
3. Cf. Jakobson & Waugh (1979) for this term.

References

Engstrand, O., L. Nordstrand and S. Zetterlund.1979. Some observations on the role of prosodic parameters in the perception of phrase structure in Swedish. Paper given at The Swedish-Danish Symposium on Phonetics, Speech Transmission Laboratory, Dept of Speech Comunication, KTH, Stockholm Sep. 1979

Jakobson, R. and Linda R. Waugh.1979. The Sound Shape of Language. Harvester Press,Sussex

Markel, J.D. and A.H. Gray Jr.1976. Linear Prediction of Speech. Springer-Verlag, New York

Zetterlund, S., L. Nordstrand and O. Engstrand.1978. An Experiment on the Perceptual Evalution of Prosodic Parameters for Phrase Structure Decision in Swedish. Paper given at The Symposium on the Prosody of the Nordic Languages, Phonetic Laboratory, Dept of general Linguistics, Lund University, June 14-16,1978

Öhman, S.E.G., S. Zetterlund, L. Nordstrand and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. In Proceedings of the 9th International Congress of Phonetic Sciences held in Copenhagen Aug. 6-11,1979. Vol II