ACOUSTIC STUDIES OF VOWEL REDUCTION IN SWEDISH

LENNART NORD

Department of Speech Communication and Music Acoustics
Royal Institute of Technology (KTH), Box 70014
S-100 44, Stockholm, Sweden

ABSTRACT

An acoustic analysis has been performed on a number of Swedish vowels, spoken in varying context. A complementary matching experiment using synthetic speech was made to see whether the analyzed differences in vowel quality were perceptually significant. Subjects had to adjust the vowel quality in words produced by an interactive rule-synthesis computer program. The purpose with these investigations was to describe and quantify the relations between vowel quality and influences of stress and position.

INTRODUCTION

In this study we focus on the concept of vowel reduction, here taken to mean the reduction in phonetic contrast between vowels. For a number of languages it has been found that the vowel space is reduced as the level of stress placed upon the vowels is reduced. Acoustic studies by Tiffany /1/, Shearme & Holmes /2/, Delattre /3/ Stålhammar, Karlsson & Fant /4/, Koopmans-van Beinum /5/ and others have shown that vowels in unstressed positions are displaced towards a more central (neutral) position in the vowel plane. A number of factors contribute to obscure vowel color in speech, see for example the study by Delattre (ref. /3/) who lists factors such as stress, rhythm, duration and contextual assimilation.

PRESENT STUDY

The aim with the present study was to study in detail some of the factors that contribute to vowel reduction in Swedish. We need a better understanding of these problems to improve the quality of synthetic speech, a typical impression being that synthesizers often over-articulate unstressed syllables.

TEST HYPOTHESIS

The phonetic context that will influence the formant pattern of a given vowel in a two-syllable word is: i) surrounding consonants, ii) the neutral position of the vocal tract, and iii) the second syllable, especially its vowel nucleus.

We focus on one aspect of the reduction phenomenon; is there a difference in formant pattern between two vowel samples of the same duration, one stressed and the other unstressed but of equal duration due to final lengthening? If there is a difference, could it be accounted for in terms of varying degrees of contextual influence?

Four types of two-syllable words were chosen with the following structure: The lexical stress on the initial or the final syllable, with dental consonants surrounding the analyzed vowel: CVC´S2, ´S1CVC, ´CVCS2 and S1´CVC, with V = the short Swedish vowels /a,i,e,ʉ/ and C-C = /s-l, l-s, s-s/, S1 and S2 =first and second syllable. This means that each analyzed vowel was placed either in initial or final position or in a stressed or unstressed syllable in an invariant consonantal frame. The words were read in isolation with no carrier phrase.

MATCHING EXPERIMENT

We were also interested in testing the perceptual significance of the results from the acoustic analysis by means of an interactive matching paradigm. That is, how reliably would subjects be able to adjust F1F2-values for given synthetic words in order to match to some internal criteria? A number of phonetic details can be tested with this type of interactive matching paradigm, using the specially developed rule synthesis program (Carlson & Granström, /6/). As long as the quality of the speech is acceptable to the subjects, segmental as well as suprasegmental cues can be evaluated. One could, for instance, let the subjects manipulate duration, pitch, intensity, etc. Few matching experiments of this type have been reported /7//8/ on segment duration. Öster /9/ also used the Carlson & Granström rule-synthesis program to systematically map typical features of the speech of deaf children.

EXPERIMENT: ANALYSIS OF NATURAL SPEECH

A high-quality recording was made of four Swedish male speakers from the Stockholm area reading twice a list of 38 lexically meaningful words with no carrier phrase in an anechoic chamber. The words contained the short vowels /a,i,e,ʉ/ and were constructed as described above. Formant frequencies and vowel durations were measured manually. The sample point for the formant measure was chosen by means of broad-band spectrograms in the middle of the vowel segment. The actual measurements were made from narrow-band spectral sections. In a few cases of uncertainty, the measurements were adjusted after comparison with selective inverse filtering which was used to display one single formant ringing at a time and enable measurements in the time domain. Based on systematic comparisons with measurements on synthetic vowels, we estimated the accuracy of the formant measurements to be +/- 20 Hz.

As it was impossible to always find content words of the right format, a few proper names were used. Also the demand on invariant CVC-syllable forced us to modify the consonantal frame for the different vowels, but still only use dentals. C-C were for the most words /s-1, 1-s, s-s/. Accordingly, a comparison across vowels has to be taken into account the difference in consonantal coarticulations that occurred.

RESULTS AND DISCUSSION. VOWEL ANALYSIS

To find out the sensitivity of formant perturbations to changes in word material, a set of words were tried with variation of consonantal frame: /s,l,d,t/ as well as a change of vowel nuclei in the other syllable. The spread turned out to be small and the tendencies the same. Therefore, it was decided to consider the influence of the different dental consonantal frame negligible and base mean values of the entire word list material.

By placing the vowel in stressed or unstressed, initial or final CVC-syllables, we obtained vowel durations ranging from 70 to 190 msec. Comparing vowels in final unstressed syllables with vowels in initial stressed syllables, we were also able

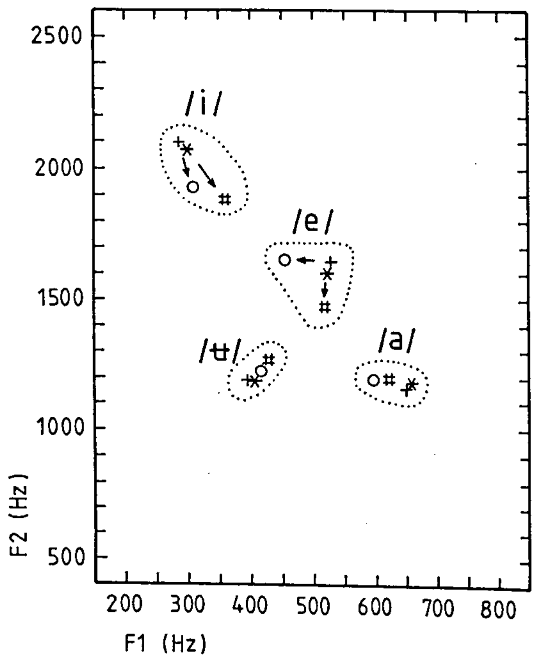

Fig. 1. Results from the analysis of real speech. Mean values of first and second formant of the short Swedish vowels /a,i,e,ʉ/. 4 male speakers, 8-40 samples/point.

	initial	final (position)
stressed	o	#
unstressed	+	*
(syllable)		

to study the influence of stress in those cases where the duration did not differ between vowels, i.e., word categories 2 and 3, which both got duration values around 125 msec.

In Fig. 1 the mean values of first and second formants are shown. As can be seen, the unstressed initial and final vowels are displaced away from the target values of the stressed vowels. For the short /e/ and /i/, it is evident that the unstressed initial samples (o) are displaced differently compared to the unstressed final samples (#) as the arrows indicate. This difference could be expressed as a difference in coarticulation: the unstressed vowels coarticulate with the consonantal frame, i.e., they move towards the dental locus of approximately 350/1650 Hz for F1/F2, while the unstressed final vowels are reduced towards a more neutral place in the vowel plane (500/1500 Hz). Formant values for the initial stressed vowels (+), that are of the same duration as the finally lengthened unstressed vowels (#) are thus not identical. Duration is thus not the sole determinant of the degree of reduction. These tendencies are not as evident for all the vowels, probably depending on the relative position of the target, the consonantal locus and the neutral vowel. For the /a/ vowel it is thus not possible to distinguish a perturbation caused by neutralization or by coarticulation as both effects will lower F1 and raise F2.

Another way of showing this effect for /e/ is to plot F2 as a function of vowel duration, see Fig. 2. As can be seen, the duration alone cannot predict the formant value. The stressed initial vowel (+) has approximately the same duration value as the unstressed final vowel (#), but speakers choose different formant values depending

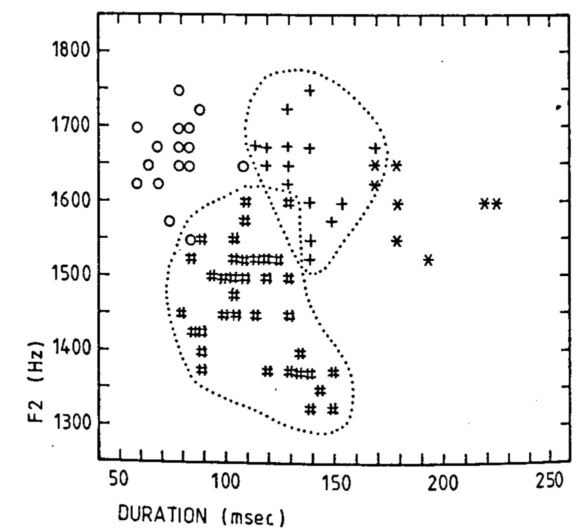

Fig. 2. Second formant value as a function of vowel duration for the short vowel /e/. Each point represents the mean value of the two readings of each speaker.

	initial	final (position)
stressed	o	#
unstressed	+	*
(syllable)		

on the syllable context (in terms of stress and position).

The intention was to maintain an invariant C-C frame for each vowel (for the four-word categories). Due to the demands for meaningful words, the consonantal frame differs somewhat for the different vowels, but still being dental. Thus, for /a/ and /e/: /s-1/, for /i/: /1-s/ and for /ʉ/: /s-s/. This causes the differences in vowel duration. /ʉ/ and /i/ become shorter than /a/ and /e/ as they are followed by a voiceless consonant.

One might also wonder whether all speakers behave in the same way. An analysis of individual performances for the four speakers shows that the tendencies vary. Two of the speakers, one of which was used in an earlier study show clear tendencies /10/. The other two speakers perform a bit differently. One shows less F1-perturbations and the other has small vowel areas in general. The vowel /ʉ/ differs appreciably between the speakers, probably due to sociolect differences.

MATCHING EXPERIMENT

As a complement to the acoustic analysis, an interactive rule-synthesis program (see ref. /6/) including an OVE III formant synthesizer, was used in a matching experiment. The task of the subject was to listen to synthetic words taken from the list of previously analyzed material, and by means of a joystick connected to the computer, adjust the quality of one vowel at a time in a word to make it sound as natural as possible. This interactive method has been used earlier for duration studies (ref. /8/). For this experimental set-up the x- and y-coordinates of the joystick were programmed to give the F1- and F2-values of the synthetic vowel that was tested. The quality of the vowel could thus, instantly, be changed by moving a cursor around in a grid pattern on the terminal screen. Different scaling and offset values were used for each test word in order to avoid learning effects. A minor modification of the duration rules made the unstressed finally lengthened vowel of equal duration as the initially stressed vowel.

With this paradigm it is possible to get valuable information about the perceptual importance of acoustic parameters. Here, where one aim is to improve the synthetic speech with regard to the vowel dynamics, it is especially interesting trying to optimize the setting of the synthesis parameters directly, using the rule synthesis.

The test was run in the following way: The subject had a list of test words with one vowel in each word marked. The task was to listen to a synthetic version of one word at a time and adjust the phonetic quality of the marked vowel to sound as natural as possible. The subjects were first instructed on the task of moving the joystick and listen to the result. The test demanded some effort in terms of concentration by the subjects so it was felt necessary to limit the word list. The same type of test as for the reading list was made to evaluate the influence due to different dental C-C frames, comparing /s-1/, /1-s/ etc. The variation in matching did not change systematically with the different frames. Therefore, the mean values are pooled over the entire word list.

Preliminary tests with phonetically untrained subjects showed that they could manage the task quite well. However, in order to keep the variability as low as possible, it was decided to use phonetically non-naive subjects.

The matched formant values were automatically stored by the program and could be analyzed immediately after each session.

Eight subjects participated, among them the four speakers in the previous test. Each subject performed two matchings on a list of 25 words; one to three words for each vowel and word category. The amount of words were limited to a selected part of the reading list as the test was rather exhausting.

MATCHING EXPERIMENT. RESULTS AND DISCUSSION

The results from the matching experiment are shown in Fig. 3 where the mean values of the first and second formants are plotted. The vowel areas are smaller than for the the spoken samples, cf., Fig. 1, but the same tendencies can be seen, although to a lesser extent. Thus, the unstressed final /e/ is matched differently than the initial stressed one, the former moving towards schwa, the latter towards the dental locus.

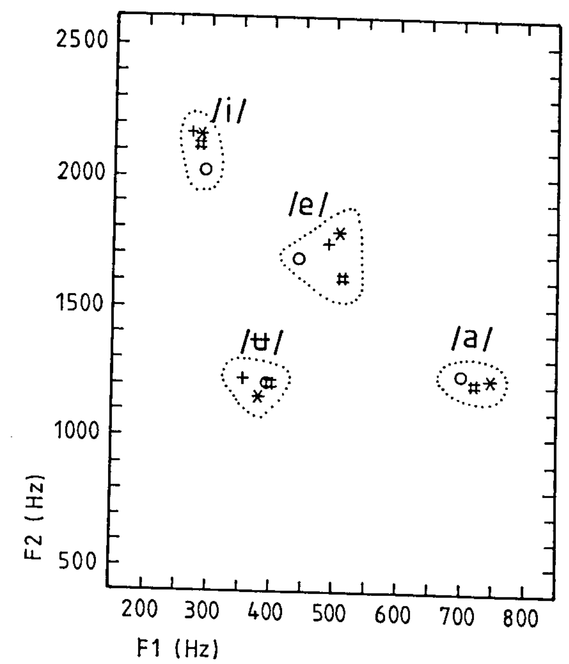

Fig. 3. Results from the matching test with synthetic words. Mean values of first and second formant of the short Swedish vowels /a,i,e,ʉ/. 8 subjects, 16-48 matchings/point.

	initial	final (position)
stressed	o	#
unstressed	+	*

A number of reasons can account for the discrepancy between the two tests. The synthetic quality will probably affect the matchings depending on the subject's acceptance of the voice quality. As only F1 and F2 were manipulated while higher formants were kept constant, especially the F2 of high, front vowels will differ from F2 of natural vowels. The matching session was experienced as a difficult but manageable task by the subjects. Also the spread between subjects was small. In conclusion, the method seems to be useful for this type of optimizations.

SUMMARY

The first and second formants were measured for four Swedish short vowels /a,i,e,ʉ/ in varying context, the purpose being to investigate factors of vowel reduction, such as stress, position and duration. The vowels were placed in stressed and unstressed, initial and final syllables in two-syllable words.

The result supports the findings in the previous pilot study by Nord (see ref. /10/). A tentative explanation to the distribution of formant data is that the perturbations are caused by contextual influence of surrounding consonants and in unstressed final position by a neutralization gesture, which in this word list material with no carrier phrase also belongs to the immediate context. If we do not reach for a phonological rule to explain the observations, specifically regarding the unstressed short /e/ in final syllable position, we could formulate the vowel reduction process in the following manner: irrespective of their duration, unstressed vowels coarticulate strongly with context: in non-final syllable position with surrounding phonemes and in final syllable position with a neutral position corresponding to a centralized schwa vowel. These tendencies were seen in varying degrees, probably depending on the relative locations of vowel targets, schwa and consonantal loci.

A supplementary study was performed using synthetic speech in order to evaluate the perceptual importance of formant perturbations in the realization of vowels in varying contexts. During the experiment, subjects were exposed to synthetic words of the same structure as in the previous experiment. The task was to adjust the quality of one vowel in each word by means of a joystick, connected to the rule-synthesis program, controlling the first and second formant of that particular vowel.

The results from this test were compared with the previous analysis. The same tendencies were seen, although to a lesser extent. This was probably due to the design of the experiment. As only two formants were manipulated, there were some difficulties in finding suitable vowel qualities during the matching procedure. Also the synthetic quality of the stimuli might have had some influence on the subjects' matching strategies. Although the task was rather difficult, subjects performed well with small deviations. One conclusion from this test is that the matching procedure using synthetic stimuli is an efficient way of evaluating perceptual cues and testing theories of speech dynamics.

REFERENCES

/1/ W.R. Tiffany, "Non-random sources of variation in vowel quality", J. Speech Hearing Res. 2, pp. 305-317, 1959.

/2/ J.N. Shearme, J.N. Holmes, "An experimental study of the classification of sounds in continuous speech according to their distribution in the Formant 1 - Formant 2 plane", pp. 234-240 in A. Sovijärvi & P. Aalto, eds., Proc. 4th Int.Congr.Phon.Sci., Mouton & Co., The Hague, 1962

/3/ P. Delattre, "The general phonetic characteristics of languages. An acoustic and articulatory study of vowel reduction in four languages", Final Report, Univ. of California, Santa Barbara, CA, USA. 1969

/4/ U. Stålhammar, I. Karlsson, G. Fant, "Contextual effects on vowel nuclei", STL-QPSR 4/1973, pp. 1-18 (KTH, Stockholm).

/5/ F.J. Koopmans-van Beinum, "Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions", Doct. thesis, Univ. of Amsterdam, 1980

/6/ R. Carlson, B. Granström, "A text-to-speech system based entirely on rules", pp. 686-689 in Conf. Record, 1976 IEEE Int.Conf. on Acoustics, Speech and Signal Processing, April, Philadelphia, PA, 1976

/7/ S.B. Nooteboom, "Production and perception of vowel duration. A study of durational properties of vowels in Dutch", Doct. thesis, Univ. of Utrecht, 1972.

/8/ R. Carlson, B. Granström, "Perception of segmental duration", STL-QPSR 1/1975, (KTH, Stockholm), pp. 1-16.

/9/ A-M. Öster, "The use of a synthesis-by-rule system in a study of deaf speech", STL-QPSR 1/1985 pp. 95-107 (KTH, Stockholm).

/10/ L. Nord, "Vowel reduction - centralization or contextual assimilation?", pp. 149-154 in G. Fant, ed., Speech Communication, Vol. 2, Almqvist & Wiksell Int., Stockholm. 1975