(ARIPUC 8, 1974)

THE INFLUENCE OF TONGUE HEIGHT ON THE PERCEPTION OF VOWEL DURATION IN DANISH

Niels Reinholt Petersen

1. Introduction

In numerous languages there is a tendency for high vowels to be shorter than low vowels, other things (quantitative category, environments, etc.) being equal. This tendency has been reported for Danish by Eli Fischer-Jørgensen (1955), for Swedish by C.-C. Elert (1964), for English by Peterson and Lehiste (1960) and by Holtse (1972), for Thai by Abramson (1962), and for Dutch by Nooteboom (1972).

Nooteboom also investigated the perception of vowel duration in a series of experiments in which the subjects adjusted the vowel duration of synthetic words according to a criterion of naturalness. In these experiments he found that the rules describing the effect on vowel duration of factors such as e.g. number of syllables in a word also described the effect on the durations preferred by his subjects in the perceptual tests. For instance, it was found in spoken words that the duration of a stressed first vowel in a word decreased as a function of the number of succeeding syllables in the word, and so did the preferred duration of that vowel in the perceptual tests. For the effect of tongue height on vowel duration, however, no such agreement between the production and perception of speech was found.

This is in disagreement with the results of a pilot experiment carried out by the present author. In that experiment

listeners were asked to judge whether a vowel embedded in a word was phonemically short or long, the duration of the vowel being varied in 10 ms steps from 80 to 200 ms. The vowels were Danish /i/ and /ɛ/. The identification functions showed a cross-over point from short to long /i/ at about 110 ms, and from short to long /ɛ/ at about 140 ms, i.e., the listeners wanted /ɛ/ to be longer than /i/ in order to judge it to be phonemically long. The listeners were also asked to compare /i/ and /ɛ/ (embedded in words) of varying duration. When /i/ and /ɛ/ were of equal physical duration, /i/ was perceived to be the longer, and when /ɛ/ was physically longer (about 20 ms) than /i/ (as is the case in natural speech), they were perceived to have the same duration. The results of this experiment were based on a very limited material, however, and cannot be considered entirely conclusive.

Thus the aim of the experiments reported below was to obtain data which could make it possible with more certainty to answer the question whether the perception of vowel duration is influenced by the tongue height[1] of the vowel perceived.

2. Stimuli and test procedure

2.1 Stimuli

The vowels chosen for the experiment were the Danish phonemes /i/, /i:/, /ɛ/, and /ɛ:/ (IPA [i] and [ɛ̝]). In order to avoid that the vowels were perceived as non-speech sounds they were embedded in the surroundings l-sə, giving meaningful Danish words for both long and short /i/ and /ɛ/, viz.: [lisə] lidse, 'lace', [li:sə] lise, 'relief' (or a personal name), ['lɛsə] læsse, '(to) load', and [lɛ:sə] læse, '(to) read'.

---

1) By the term tongue height is meant the properties of the acoustic signal which convey information on that feature.

The words were synthesized on the parallel speech synthesizer of the Institute (Rischel 1969; Rischel and Lystlund 1972). The build-up of the synthetic words was based on spectrograms of the words spoken by four Danish speakers. Information about the formant frequencies and levels of the vowels [i] and [ɛ] (see table 1) was kindly provided by Mr. P. Holtse. These data corresponded to formant data that had given 100 per cent identification as /i/ and /ɛ/ in an identification test carried out by him (Holtse 1973). (There is only a slight difference between formant frequencies of short and long /i/ and /ɛ/ in normal spoken Danish.)

TABLE 1.

Formant frequencies and levels of the vowels [i] and [ɛ]. Formant frequencies in cps and levels in dB relative to the level of F1.

	i		ɛ	
	formant freq.	level	formant freq.	level
F1	226	0	416	0
F2	2326	-26	1970	-14
F3	3391	-19	2491	-18
F4	3800	-20	3750	-28
F5	4430	-26	4450	-26

The F_o contour was rising, beginning at 92 cps in the [l] and ending at 110 cps in the [ə]. In the vowel, however, F_o was kept constant at 100 cps; otherwise the variation of vowel

duration would have caused a variation in the rate of F_0 change during the vowel, which might influence the perception of vowel duration.

The vowels [i] and [ɛ] were built up by four 'acoustic segments'. (Dynamic control of connected speech on the synthesizer is obtained by means of a function generator producing varying or constant voltages in 20 successive steps - 'acoustic segments'. Each of these segments holds information on the parameter values and the transition time and can be varied in duration from normally 5 to 100 ms, see Rischel 1969; Rischel and Lystlund 1972.) The duration of one of the acoustic segments could be varied continuously by means of a remote control device which was to be used by the subjects for the adjustment of the vowel duration to meet certain criteria (see below). For this adjustment procedure it would have been most expedient if the subjects could cover the total range of expected durations in one sweep of the remote control. However, for the criteria employed in the tests, the expected adjustments varied from less than 100 ms to more than 250 ms. It turned out that this wide range could only be achieved at the expense of precision of adjustment and of linearity within the range. The widest range of continuous variation by means of the control device which gave an acceptable precision and linearity with our function generator was of the order of 90 ms (from 10 ms to 100 ms). Thus, in order to cover the total range of expected durations it did not suffice to vary one acoustic segment: it was necessary to select suitable durations of the other three acoustic segments of the vowel, depending on the type of stimulus and on the subject. The total duration of the parts of the vowel which were not affected by the remote control was thus set somewhere between 50 ms and 205 ms. If adjustments were to be made in the short vowel area, the three segments might, for instance, be given a joint duration of 50 ms by the experimenter,

and the subject could then adjust vowels ranging from 60 to 150 ms. Similarly, if adjustments of longer durations were to be made, the three segments were lengthened, thereby shifting the range of the remote control device to cover adjustments of higher values of duration (e.g. from 175 to 265 ms with the duration of the three segments set at 165 ms). The highest range which could be obtained was 215 to 305 ms. The duration of the segment connected to the control device could be read by the experimenter from an electronic counter, and the total vowel duration was then the sum of that segment and the three segments set by the experimenter.[1]

2.2 Test procedure

The subjects' task was to adjust the vowel duration according to the following criteria by means of the remote control device:

1. Normal duration for phonemically short vowel.
2. Normal duration for phonemically long vowel.
3. Boundary between phonemically short and long vowel.

During the experimental sessions the subject was seated at the control device listening to the output from the synthesizer via ear phones (Sennheiser HD 414). The synthesizer was

---

1) The actual value of vowel duration may deviate from that specified by the function generator. It seems that there is an inherent source of error due to the fashion in which the amplitude is controlled: the point in time at which the sound is programmed to start, i.e. at which the voice source amplitude gate is activated to produce an increasing amplitude, is independent of the repetition rate of voice source pulses. Since the amplitude gate is placed before the formant filters, this means that the first pulse to produce an appreciable excitation of the formant filters (and hence a vowellike output) may occur with varying time-lags relative to the programmed onset, i.e. signals differing up to one period in duration (e.g. 10 ms at F_o = 100 cps.) can be generated by exactly the same programming of the function generator. There is a similar possibility of error at the end of the synthesized vowel. - However, the results of perceptual experiments carried out at the present time suggest that the influence of these inaccuracies on the perception of duration is smaller than should be expected, but nothing conclusive can be said about the problem until investigations in progress have been completed.

set to repeat the programmed word once every two seconds, during which interval the subject could operate the control device to obtain a duration corresponding to the criterion given. When he found that he had adjusted the optimal duration the subject gave notice to the experimenter, who took down the reading of the electronic counter.

Because of the limitations of the apparatus mentioned above (section 2.1), every experimental session was initiated with a series of pilot adjustments to the criteria of that session. The purpose of these adjustments was to determine the best range of duration for the subject in question. This was done by presenting the subject with different ranges, e.g. 60 to 150 and 80 to 170 ms for the adjustment of normal short vowel durations, and choosing the one where his adjustments were well within the limits. This range was then kept for the succeeding adjustments to the criterion in question.

Adjustments to criteria 1 and 2 above were carried out in one experimental session, short and long vowels being adjusted alternately. Each subject adjusted 20 short and 20 long vowels of each quality.

In order to establish the perceptual phoneme boundary between short and long vowel, the subjects alternately started the adjustment from the lower limit of the range chosen, increasing the duration until he perceived a long vowel (i.e. perceived the word to be [li:sə] or [lɛ:sə], respectively), and from the higher limit of the range, decreasing the duration until he perceived a short vowel (i.e. perceived the word to be [lisə] or [lɛsə], respectively). Twenty 'short to long' and twenty 'long to short' adjustments were made by each subject for each vowel quality.

All [i]-adjustments were carried out first, and then, after a couple of weeks, the [ɛ]-adjustments were made.

Six subjects took part in the experiment. Five were

university students of phonetics, and one (NT) was a member of the staff of the institute. All subjects made adjustments to criteria 1 and 2. Only four (BH, JJ, EBC, and NT) did criterion 3.

3. Results

Arithmetic means and standard deviations of the vowel durations adjusted are given in table 2, and the means are graphically displayed in fig. 1.

TABLE 2

Average adjustments ($\bar{X}$) and standard deviations (s) for each of the six subjects

		adjustments of normal durations				adjustments of boundaries			
		short vowel		long vowel		long to short		short to long	
		i	ɛ	i	ɛ	i	ɛ	i	ɛ
EBC	$\bar{X}$	75	116	163	196	97	141	99	143
	s	5.4	8.6	4.2	7.3	3.5	5.1	5.2	3.1
JJ	$\bar{X}$	84	108	148	208	95	126	106	156
	s	6.7	8.0	7.0	8.2	5.2	5.0	5.2	5.2
PA	$\bar{X}$	86	118	150	213				
	s	7.9	9.3	9.7	9.3				
BH	$\bar{X}$	86	109	151	216	98	138	96	138
	s	2.9	7.1	4.4	7.2	4.2	4.8	3.6	5.8
MJ	$\bar{X}$	89	110	182	226				
	s	6.0	11.0	10.1	11.8				
NT	$\bar{X}$	107	130	206	230	123	177	130	193
	s	8.4	5.5	8.7	10.9	5.1	7.9	5.6	4.7
grand mean diff.			27 ms		48 ms				

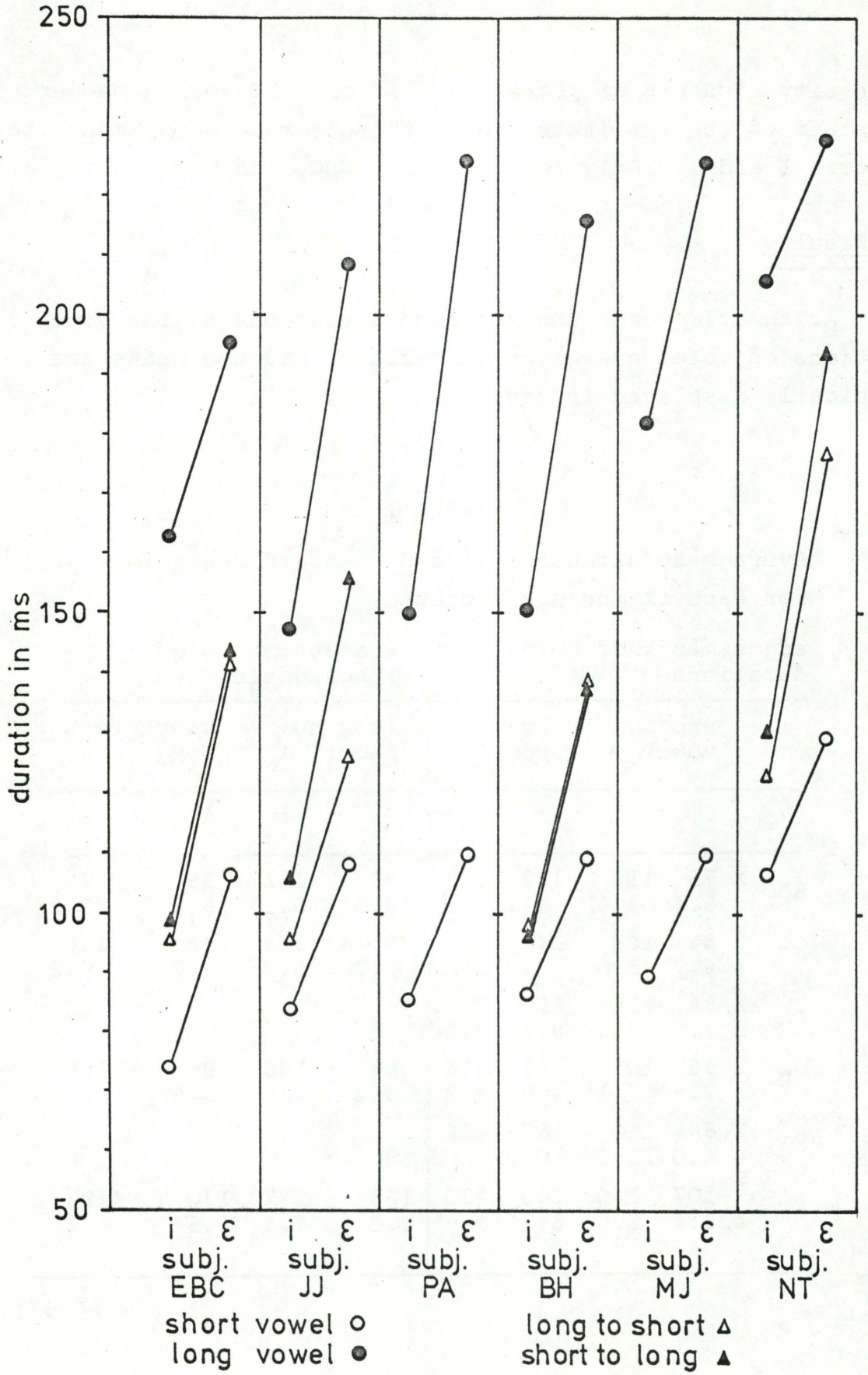

Fig. 1

Average adjustments of normal long and short vowels and phoneme boundaries for each of the six subjects.

From the data it appears to be quite evident that the subjects wanted [ɛ] to be longer than [i], when the vowels were adjusted in accordance with the same criterion. In all cases the difference between means was statistically significant at the 1 per cent level of significance.

Thus it seems reasonable to conclude that the perceptual duration of a vowel is influenced by the acoustic properties conveying information on the tongue height of that vowel, and, further, that the agreement between production and perception of vowel duration as found by Nooteboom for some influencing factors (see section 1 of this paper) may be extended to include tongue height as well.

References

Abramson, A.S. 1962: The vowels and tones of Standard Thai: Acoustical measurements and experiments (Bloomington Ind.: Indiana University)

Elert, C.-C. 1964: Phonologic studies of quantity in Swedish (Stockholm, Göteborg, Uppsala)

Fischer-Jørgensen, E. 1955: "Om vokallængde i dansk rigsmål", NTTS 15

Holtse, P. 1972: "Spectrographic analysis of English vowels", ARIPUC 6

Holtse, P. 1973: "Identification and discrimination of closely spaced synthetic vowels", ARIPUC 7

Nooteboom, S.G. 1972: Production and perception of vowel duration (dissertation)

Peterson, G.E. and I. Lehiste 1960: "Duration of syllable nuclei in English", JASA 32

Rischel, J. 1969: "Constructional work on a function generator for speech synthesis", ARIPUC 3

Rischel, J. and S.-E. Lystlund 1972: "A formant-coded speech synthesizer", ARIPUC 6