Perceptual Experiments with Duration versus Spectrum in Swedish Vowels

Kurt Johansson

1. INTRODUCTION

It is well known that the vowels in Swedish word pairs like vit:vitt, kal:kall, rot:rott generally differ both with regard to duration and to spectrum, i.e. formant frequencies (perceptually: vowel length and vowel quality, respectively). There are also differences, particulary durational, between the final consonants, short consonants after long vowels and long after short. Consistent durational differences, between vowels as well as consonants, only appear in stressed positions, while some vowel quality differences may be upheld also in unstressed positions where all vowels are short. (1)

Different phonological interpretations are discussed for instance by Elert (1964, 39 ff. and 1970, 54 ff.) and Hadding-Koch - Abramson (1966). In this paper I will not enter this discussion.

Long and short vowels are manifested rather differently in various Swedish dialects. Long vowels may be pronounced as monophthongs, but they are generally more or less diphthongized (Elert 1981). Short vowels may also be diphthongized but ordinarily they are more monophthongal. (2)

The perceptual investigation carried out by Hadding-Koch -

Abramson (op.cit.) dealt, as does this paper, with the question of whether vowel duration or vowel spectrum is the primary cue for distinguishing minimal pairs of the above type. They prepared their stimuli from 3 Swedish word pairs by using a tape cutting and splicing technique. (3) The male speaker and the listeners came from Southern Sweden. For practical reasons only non-diphthongized vowels were included. The main results were that vowel duration was shown to be the primary cue for [ɛ:]:[ɛ] and [ø:] : [œ] and formant frequencies, i.e. quality, for [ɯ:] : [ɑ](4).

PURPOSE OF THE PRESENT INVESTIGATION

One of the reasons for carrying out the experiments presented below was that I wanted to cover all 9 vowel pairs of Central Swedish:
([i:]:[I],[e:]:[ɛ],[ɛ:]:[ɛ],[y:]:[Y],[ø:]:[œ],[ɯ:]:[ɑ],
[u:]:[U],[o:]:[ɔ], and [ɑ:]:[a]). (5)

My hypothesis was that vowel duration would be the main perceptual cue for some of the vowel pairs and vowel spectrum for others. What is actually tested, then, is which manner of transcribing, for example /vi:t/:/vit/ or /vit/:/vIt/, is nearest to the perceptual reality. (6)

The background for this hypothesis was of course, beside the findings of Hadding-Koch - Abramson, the observations of many investigators that the magnitude of vowel spectrum differences, and sometimes also of vowel duration differences, may vary for different vowel pairs. My own informal tests with electronically gated speech also contributed to the assumption.

Another reason for my experiments was in fact to give a background for spelling methodology in Swedish schools. The use of single or double consonants in Swedish spelling is largely dependent on the previous vowel, and there has over the years

been much discussion on the issue of whether vowel length or vowel quality should be considered the appropriate methodological starting-point. One reason why I chose to work with Central Swedish was that dialects from Central Sweden have been normative for Swedish spelling.

In this paper, however, I do not intend to go into this discussion, as I have done this elsewhere (Johansson 1981) and my standpoint may be inferred from the results presented below.

3. PREPARATION OF THE TEST MATERIAL

In order to be able to control as many variables as possible I chose to work with synthetic speech. The stimuli were created with an OVE III speech synthesizer, controlled by an ALPHA LSI computer.

I preferred, as did Hadding-Koch - Abramson, to use non-diphthongized vowels. Diphthongized vowels will be dealt with elsewhere.

The test vowels appeared in 9 monosyllabic word pairs, always preceded by [h] and followed by [s] .(7)

The [h] was constructed as a voiceless counterpart to the adjoining vowel.

The [s] was characterized by two formant bands centered around 4 800 and 7 600 cps. It should be mentioned that the reason for choosing [s] was that it does not have any voiced counterpart in Swedish. Not much experimenting is needed to show that listeners tend to interpret, where this is possible, a consonant as voiced or unvoiced, not only in accordance with voicing or aspiration but also depending on the duration of a preceding vowel. In the experiments reported below this fact would without doubt have created an unnecessary complication.

In spite of the fact that the duration of a postvocal consonant is complementary in stressed words in most Swedish dialects, it has generally been assumed that durational differences between consonants are not distinctive, and this was also supported by the Hadding-Koch - Abramson experiments. I had planned to investigate this further, but an unfortunate location of some test stimuli made the results less reliable, and I have preferred to return to this question on another occasion. I am, however, apt to believe that differences in the durations of the consonants are of no consequence in the present investigation. Anyhow I choose a value for the [s] duration averaged between long and short according to Elert's measurements (1964, p 143), in this case 210 milliseconds.

The formant frequencies used were the ones reported by Fant for Central Swedish speakers (appendix 1). This was the only investigation available giving formant frequencies for both long and short vowels. In spite of the fact that the measurements had been made on isolated vowels the listeners did not react or comment on the resulting stimuli as being unnatural.

Steady state vowels, i.e. vowels without formant transitions, were chosen, which is of little perceptual consequence before [s] with its high intensity and strong intrinsic cues.

As fundamental frequency variations are known to influence the perception of length the frequency was held constant (at 100 cps).

In the test the vowel qualities were kept unchanged and the only thing that was varied was the duration of the vowels. The variations were made in 20 millisecond steps from 80 milliseconds to 200. (8)

In this way 126 different stimuli were created. On the test tape a buffer of 10 stimuli was recorded in order to let the

listeners become acquainted with the test procedure and as these stimuli had rather normal durations they may also have served to give the listeners a tempo reference. The rest of the stimuli were randomized and each stimulus recorded exactly in the same way three times in a row, with one-second pauses. Between different stimuli there was a pause of 2.5 seconds. After every tenth stimulus a longer pause of about 5 seconds was made. After 50 and 100 stimuli there was a longer pause. Each stimulus only appeared once in the test.

The whole test lasted for about 20-25 minutes.

4. LISTENERS

In all, 50 listeners took part in the test, 42 women and 8 men with an average age of 40. 35 of the participants were school teachers specializing in speech, reading, and spelling therapy, 10 were students or teachers of speech pathology, and the others were specialists in other fields.

5. LISTENING TEST

The test stimuli were presented to the listeners from a Revox A 77 tape recorder via Burwen PMB 6 head-phones.

The listeners were given a test form having two choises for each stimulus, and they were asked to underline the word they thought had been presented (forced choise).

6. RESULTS

As was mentioned earlier the formant frequencies of the vowels were kept constant and only the durations changed, the purpose being to investigate whether vowel duration or vowel spectrum

constitutes the major perceptual cue for distinguishing within the 9 word pairs.

If for instance a word his [hi:s] with a gradually shortened vowel eventually will be judged as hiss [hIs], in spite of the fact that vowel quality has not been changed, duration must be considered the essential cue. If, on the other hand, the identy of the word does not change, quality must be the important thing.

Figure 1, treating all listeners as one group and all vowels, reveals that on the whole duration must be considered the major cue. When the vowels are short, both vowels with qualities appropriate for long and for short vowels are judged as short. In the middle of the figure a change of identity takes place, and both types of vowels are judged as long when they have been lengthened enough.

If, however, we look at each vowel pair separately, it becomes clear that all pairs do not behave in the same manner. Most pairs change from short to long responses within the range of 120-180 milliseconds. (9) This is the case for [i:]:[I],[e:]:[ɛ],[ɛ:]:[ɛ],[y:]:[Y],[ø:]:[œ], and [u:] [U], i.e. 6 out of 9 pairs. The listeners tend to judge a vowel as short up to a duration of about 120 milliseconds and as long from about 160 (with 7 out of 12 vowels exactly at these values). These values may be compared with the average values reported by Elert for Stockholm Swedish (1964, p 109), 108 milliseconds for short vowels in his one-word list and 163 for long vowels. (The values in his sentence list are somewhat lower, but not very much.) The "short-point" above is 75% of the "long-point", a percentage that may be compared with what is generally reported for short vowels in relation to long (65-70%). (10)

The distance between the points where listeners judge the above vowels as short and long, respectively, can be described by

positive values, i.e. a vowel judged as long is always longer than a vowel judged as short, and within these vowel pairs it does not matter that vowel formant frequencies are different. The remaining 3 pairs, [ɯ:]:[ɑ], [ɑ:] :[a], and [o:]:[ɔ], give negative values, i.e. there is considerable overlap so that a long vowel may be shorter than the "short" vowel, and the short vowel longer than the "long" vowel, without making the listeners judge them as short and long, respectively. Vowel quality is obviously the important thing here.

In order to get a better picture of the contribution of duration versus spectrum differences within the word pairs, the data will be presented in another way.

Figure 2 shows what could be expected under ideal conditions, if duration alone were the distinctive cue. Along the x-axis we have gradually increased the duration of the vowel, along the y-axis we have the percentage of "long"-responses both for "long" and "short" vowels. If duration were the only cue, the results should be as indicated in this figure. Both "long" and "short" vowels should be judged as short or long in accordance with the duration of the vowel, and there should be an uncertainty region in between.

The expected results, if vowel quality were the only distinctive cue, should be in conformity with figure 3. A "short" vowel should be judged as such and give 0% "long"-responses, whatever the duration of the vowel, and in the same way a "long" vowel should give 100%.

In figure 4 all listeners are treated as one group, without any consideration of dialect, and all vowels are included . It is quite clear, as it was from figure 1, that duration is the primary cue. The curves for "long" and "short" vowels are very much the same as the one presented in figure 2, but the curve for the "short" vowel is consistently below the other curve, which indicates that quality is not unimportant.

It can be seen from figures 5-13 that the vowel pairs may be divided into 3 different groups:

The first group consists of the 6 pairs commented on earlier, [i:]:[ɪ], [e:]:[ɛ], [ɛ:]:[ɛ], [y:]:[ʏ], [ø:]:[œ], and [u:]:[ʊ] (figures 5-10). It might have been expected that [ø:] [œ] would have shown a greater dependence on the quality cue, but the results here are quite in agreement with the findings of Hadding-Koch - Abramson. Within the group the pair [e:]:[ɛ] displays the greatest distance between the curves, but there can be no doubt that duration is the essential cue here, too.

Only one vowel pair, [o:]:[ɔ], belongs to group 2 (figure 11). The curves are here considerably further apart than for the above group, and although they look very much like the curves for group 1, I am, considering what I said earlier regarding this pair, apt to consider quality as the distinctive cue. But duration is certainly not unimportant.

The third group, consisting of [ɯ:]:[ɑ] and [ɑ:]:[a], is of a different kind (figures 12-13). The curves are more of the type presented in figure 3, although they have the appearance of being turned on end as a result of the fact that duration even here is not completely unimportant. One reason for this behaviour of the curves might be that some of the listeners come from regions that do not have the usual quality differences between "long" and "short" vowels. In fact some of the dialect information given on the test forms indicates this, but the information is not detailed enough to allow a definite conclusion. However, the behaviour of the curves in figures 12-13, together with what was said about overlapping earlier, are to my mind indication enough that quality is the distinctive cue. Another thing that supports this conclusion is that the "long" vowels with their shortest durations, and the "short" vowels with their longest, do not change the responses further than to values indicating guessing.

7. DO LISTENERS FROM DIFFERENT DIALECTS BEHAVE DIFFERENTLY?

A question that has sometimes been discussed by Swedish phoneticians is if for instance people from Scania in the south of Sweden, with their tendency to diphthongize, particularly the "long" vowels, use these quality differences distinctively and not duration differences.

By coincidence there were among the listeners 15 Scanians and 15 people speaking various dialects of Central Swedish. The groups are of course too small to enable any definite answer, and as the listeners belong to different dialectal subgroups and are rather sophisticated they can not be regarded as representatives of genuine dialects.

If we, bearing this in mind, take a look at figure 14 we can see that the people from Central Sweden give more "short"-responses when the vowels are short and more "long"-responses when they are long. This is valid for "long" as well as "short" vowels. To put it another way, the Scanians are not quite as apt as the people from Central Sweden to judge according to duration.

The difference between the two groups is not very great, but together with what we know from other investigations, above all that Scanians tend to diphthongize more than most Swedes and that duration differences both between "long" and "short" vowels and consonants are less than for Central Swedish, we have indications that the role of quality variations should be investigated further. I am currently preparing some experiments that I hope will serve to elucidate this particular question.

8. SUMMARY

A listening test with synthetic stimuli, based on Central Swedish , was carried out using 50 listeners, the purpose being

to investigate whether duration or formant frequency is the primary cue for distinguishing between [i:]:[I], [e:]:[ɛ], [ɛ:]:[ɛ], [y:]:[Y], [ø:]:[œ], [ʉ:]:[ʊ], [u:]:[U], [o:]:[ɔ] [ɑ:]:[a] respectively.

The following conclusions may be drawn:

a) If we look at the whole material, duration must be considered the major cue.
b) The vowel pairs may be divided into three groups, depending on which is the essential cue:
    1 The duration group containing [i:]:[I], [e:]:[ɛ], [ɛ:]:[ɛ], [y:]:[Y], [ø:]:[œ], and [u:]:[U].
    2 The quality group containing [ʉ:]:[ʊ] and [ɑ:]:[a].
    3 The intermediate group, consisting only of [o:]:[ɔ]. For this pair both duration and quality are of importance. I am, however, inclined to consider this group as a subgroup of the quality group.
c) A comparison of a group of Central Swedish speakers and Scanians indicates, together with other things, that dialectal differences should be investigated further, particularly dialects with strong diphthongization.

NOTES

1) I will in this paper only deal with vowels in stressed positions.

2) As for instance in the Malmö dialect (Bruce 1970).

3) The word pairs used were: väg 'road':vägg 'wall', stöta 'push': stötta 'prop up', and ful 'ugly': full 'full'.

4) The symbols taken from the Swedish dialectal alphabet. The corresponding IPA symbols ordinarily used are [ʉ] and [ɵ], respectively.

5) As mentioned earlier Hadding-Koch - Abramson only covered three out of nine pairs, and they used a speaker and listeners from Southern Sweden. Furthermore they seem to mean (p 106) that vowel duration is the essential cue for all vowel pairs exept one, viz. [ɯ:] : [ɑ] , a view that was not compatible with my own experiences.

6) In spite of the fact that I have used a colon as a marker of length, I do not, so far, wish to take sides concerning the question of whether length should be considered a separate phoneme or a distinctive feature of the vowel.

7) Some of the resulting monosyllables are nonsensical, but they are all possible Swedish words.

8) Elert (1964, p 109) reports 108 milliseconds for short vowels and 163 for long as average duration values in his one-word list. Before [s] which was used in my experiments, he reports exactly the same durations in his sentence list (p 115 and p 118).

9) I have made no statistical treatment of the data in this paper but instead used a rough estimation and considered the listeners convinced that a certain word has been presented, when the percentage falls within the upper or lower third.

10) See for instance Gårding et al. (1974, p 108).

11) Occasional irregularities in the curves generally depend on preceding stimuli.

FREQUENCIES FOR THE FIRST THREE FORMANTS (after Fant 1957)

	F_1	F_2	F_3
[i:]	240	2050	3000
[I]	300	2050	2700
[e:]	340	2100	2500
[ɛ]	385	2000	2450
[ɛ:]	440	1800	2400
[ɛ]	385	2000	2450
[y:]	255	1950	2400
[Y]	300	2000	2400
[ø:]	350	1700	2200
[œ]	400	1550	2300
[ɯ:]	260	1600	2150
[ɤ]	450	1050	2300
[u:]	280	700	2250
[U]	340	700	2600
[o:]	410	750	2450
[ɔ]	500	850	2550
[ɑ:]	650	1000	2500
[a]	750	1250	2500

REFERENCES

Bruce, G. 1970. Diphthongization in the Malmö dialect. Working papers 3. Phonetics Laboratory, Lund University.

Elert, C.-C. 1970. Ljud och ord i svenskan. Stockholm: Almqvist & Wiksell.

Elert, C.-C. 1981. Ljud och ord i svenskan 2. Umeå: Almqvist & Wiksell.

Elert, C.-C. 1964. Phonologic studies of quantity in Swedish. Uppsala: Almqvist & Wiksell.

Fant, G. 1957. Den akustiska fonetikens grunder. Kungliga Tekniska Högskolan, Stockholm, Avd för Telegrafi-Telefoni, Rapport No 7.

Gårding, E., R. Bannert, A.-C. Bredvad-Jensen, G. Bruce and K. Nauclér. 1974. Talar skåningarna svenska? Svenskans beskrivning 8:107-117. Lund.

Hadding-Koch, K. and A.S. Abramson. 1966. Duration versus spectrum in Swedish vowels: some perceptual experiments. Studia Linguistica XVIII:94-107. Lund.

Johansson, K. 1981. Bör dubbelteckningsmetodiken bygga på längd- eller klangfärgsskillnader? Utvecklingsarbete och fältförsök, Rapport 2/81. Lärarhögskolan, Malmö.

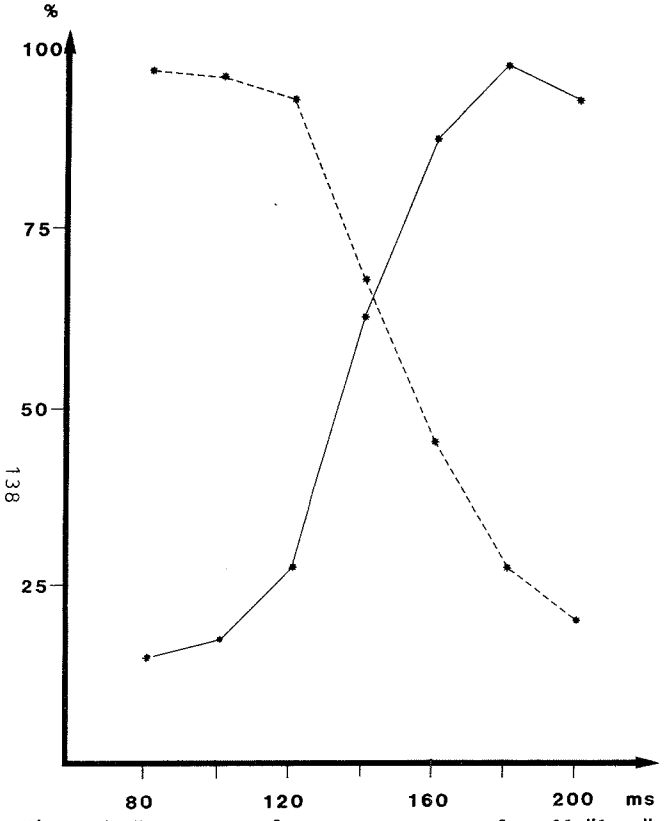

Figure 1 Percentage of correct responses for all "long" vowels (continuous line) and all "short" (dashed).

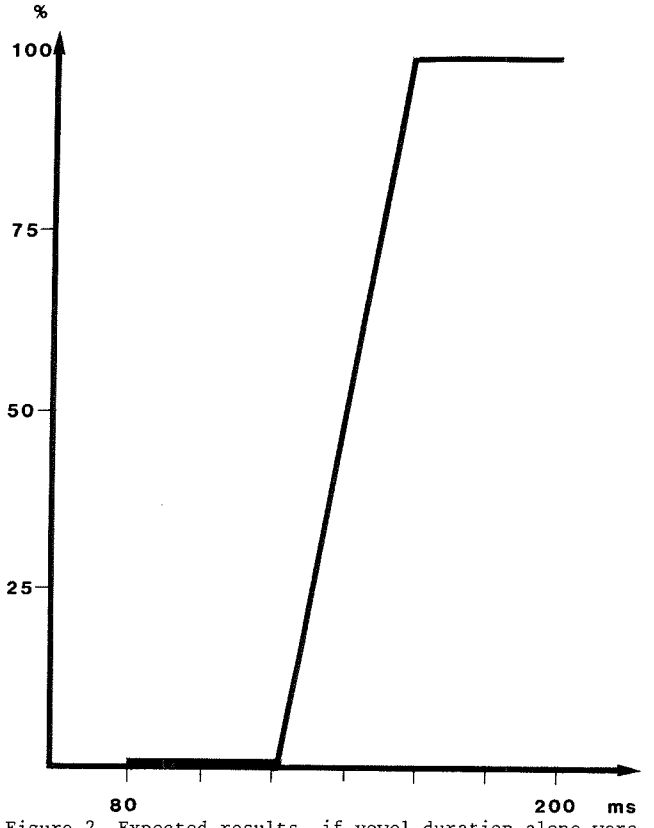

Figure 2 Expected results, if vowel duration alone were the distinctive cue. Duration along the x-axis, percentage of "long"-responses along the y-axis.

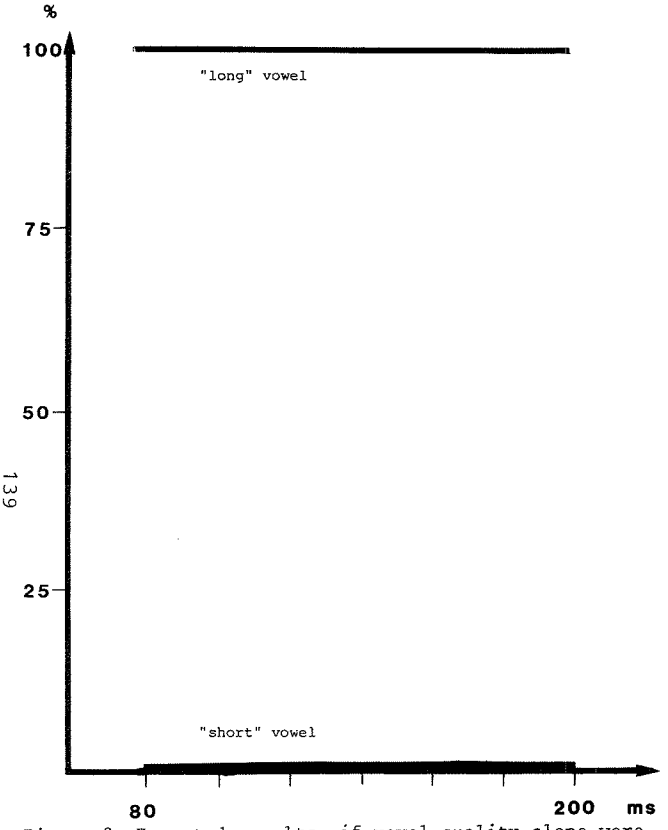

Figure 3 Expected results, if vowel quality alone were the distinctive cue.

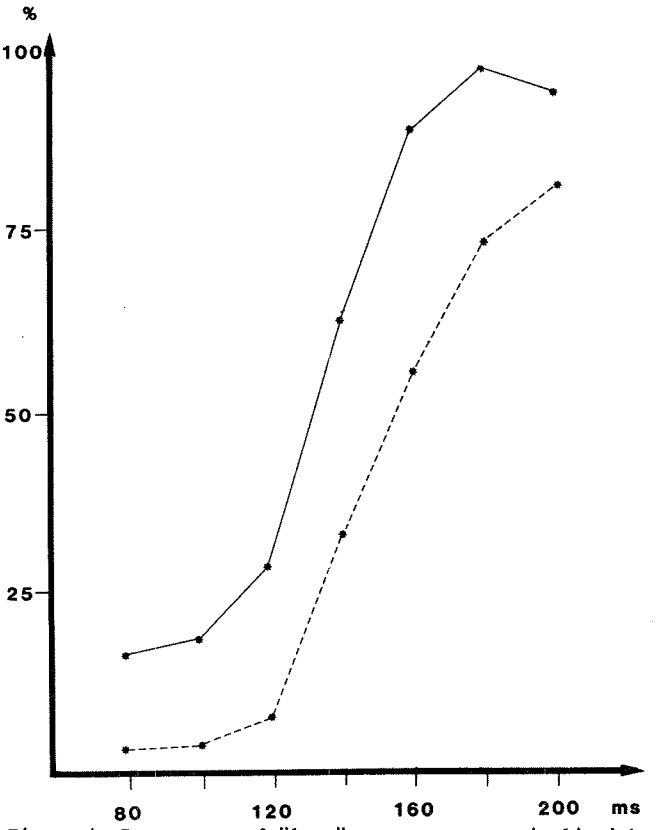

Figure 4 Percentage of "long"-responses to stimuli with gradually increased vowel duration. Continuous line = "long" vowels, dashed line = "short" vowels.

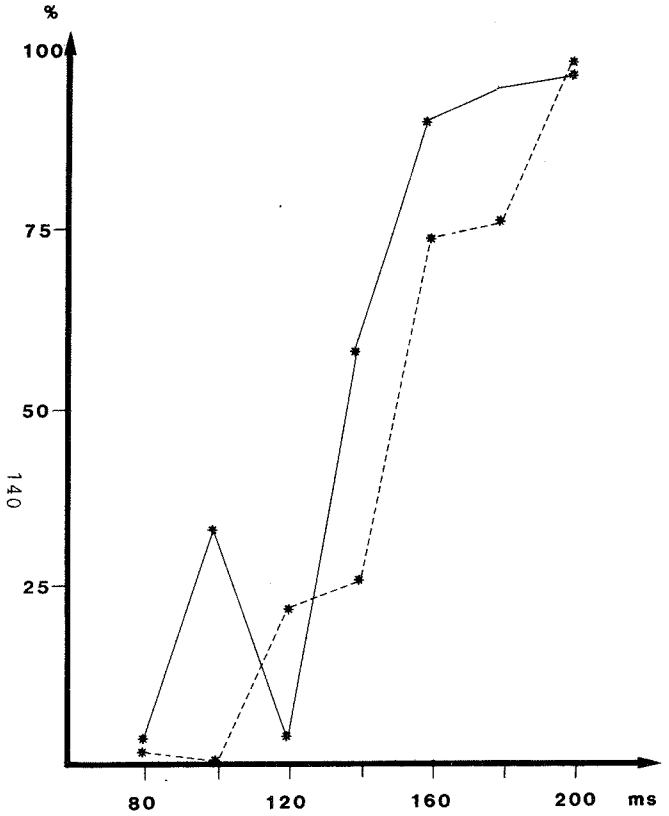

Figure 5 Percentage of "long"-responses for [i:] (continuous line) and [I] (dashed).

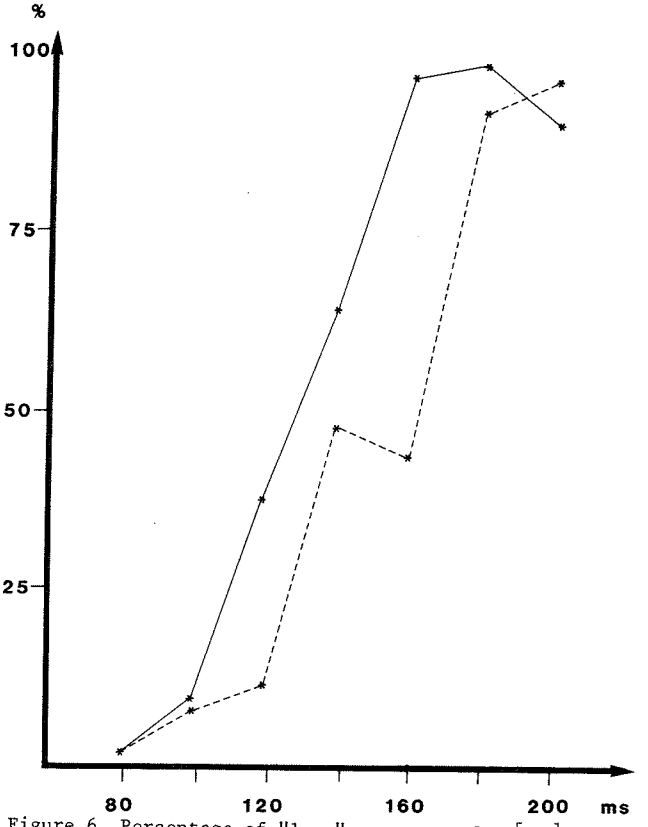

Figure 6 Percentage of "long"-responses for [e:] (continuous line) and [ɛ] (dashed).

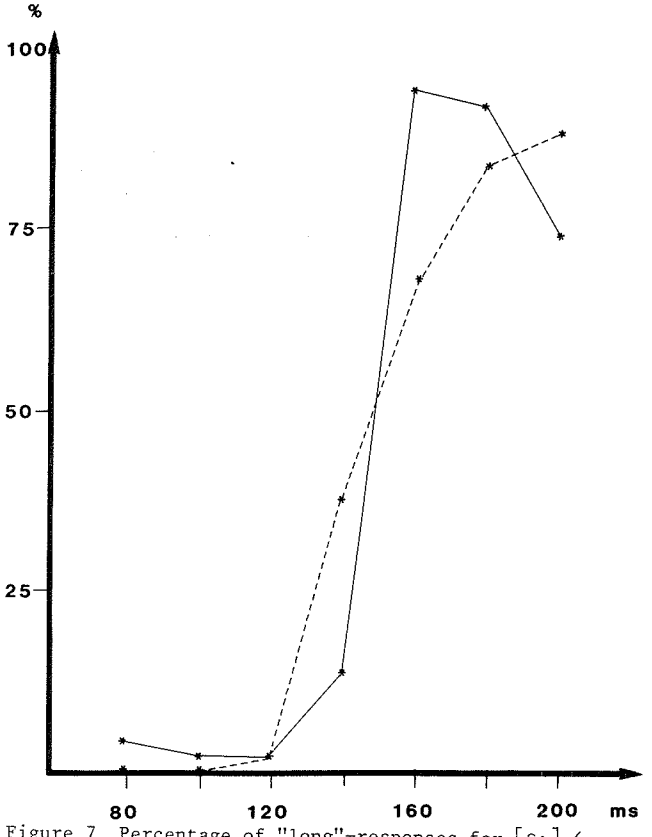

Figure 7 Percentage of "long"-responses for [ɛ:] (continuous line) and [ɛ] (dashed).

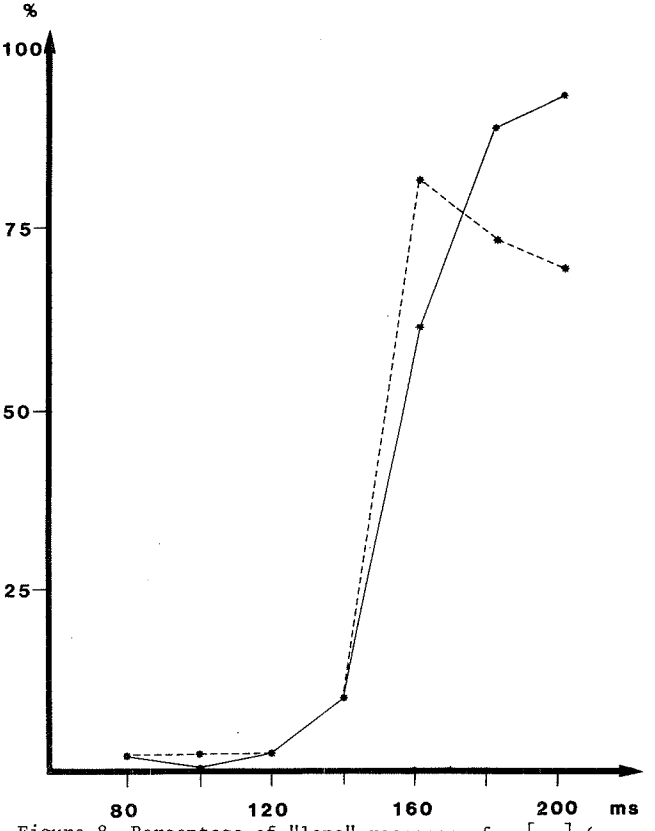

Figure 8 Percentage of "long"-responses for [y:] (continuous line) and [Y] (dashed).

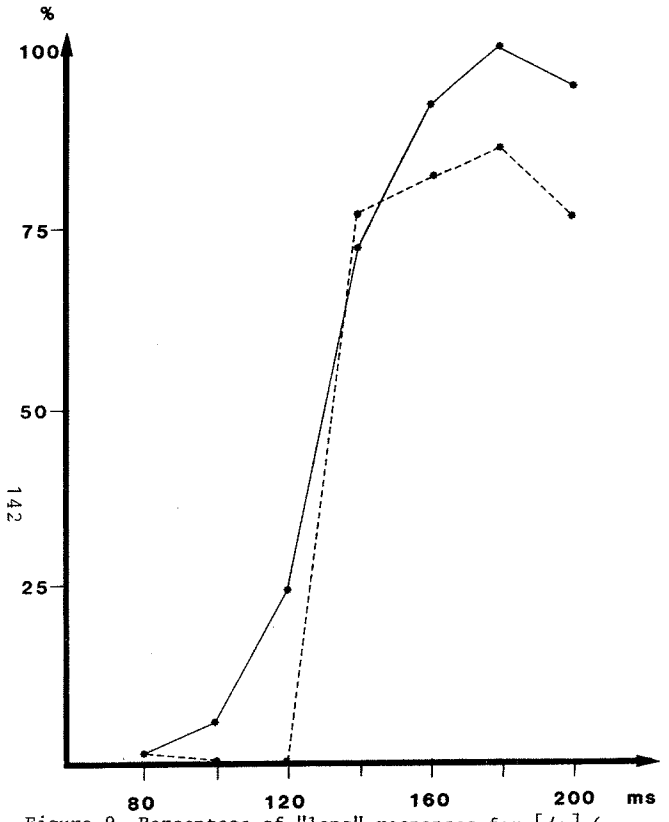

Figure 9 Percentage of "long"-responses for [ø:] (continuous line) and [œ] (dashed).

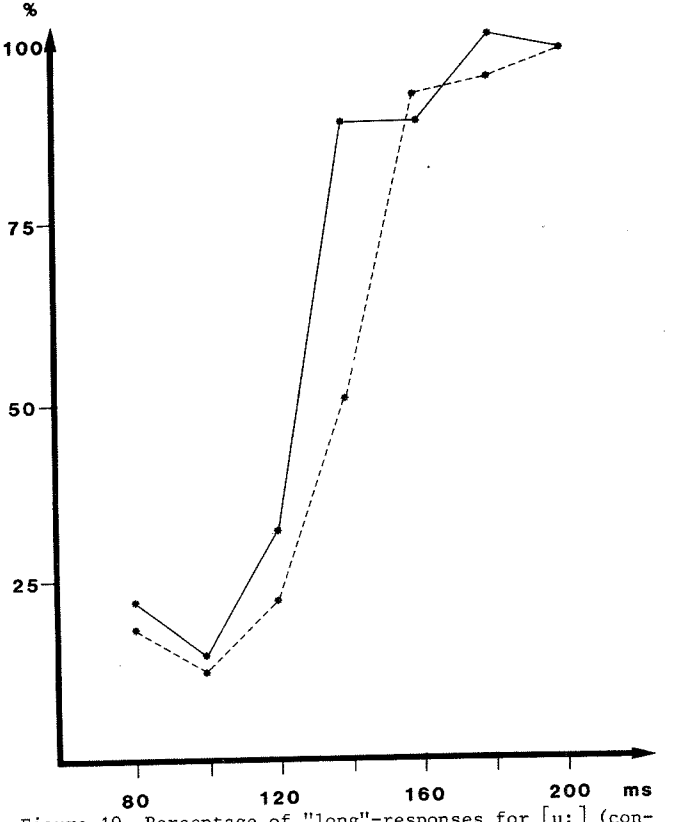

Figure 10 Percentage of "long"-responses for [u:] (continuous line) and [U] (dashed).

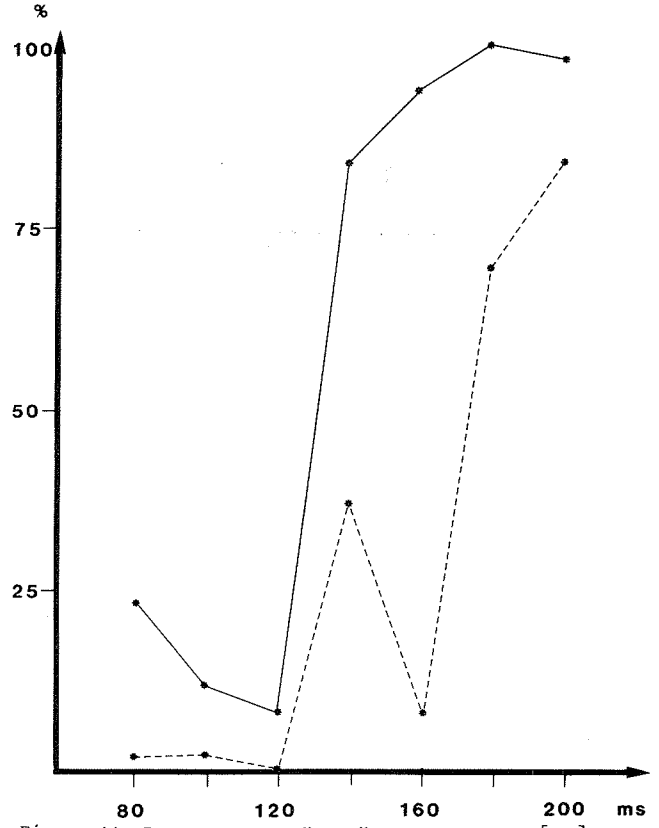

Figure 11 Percentage of "long"-responses for [o:] (continuous line) and [ɔ] (dashed).

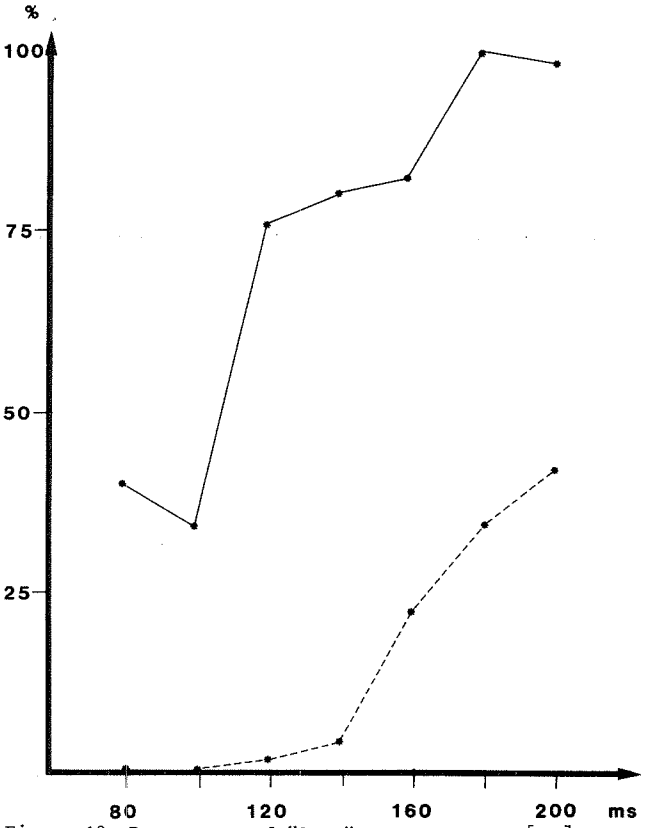

Figure 12 Percentage of "long"-responses for [ʉ:] (continuous line) and [ʉ̆] (dashed).

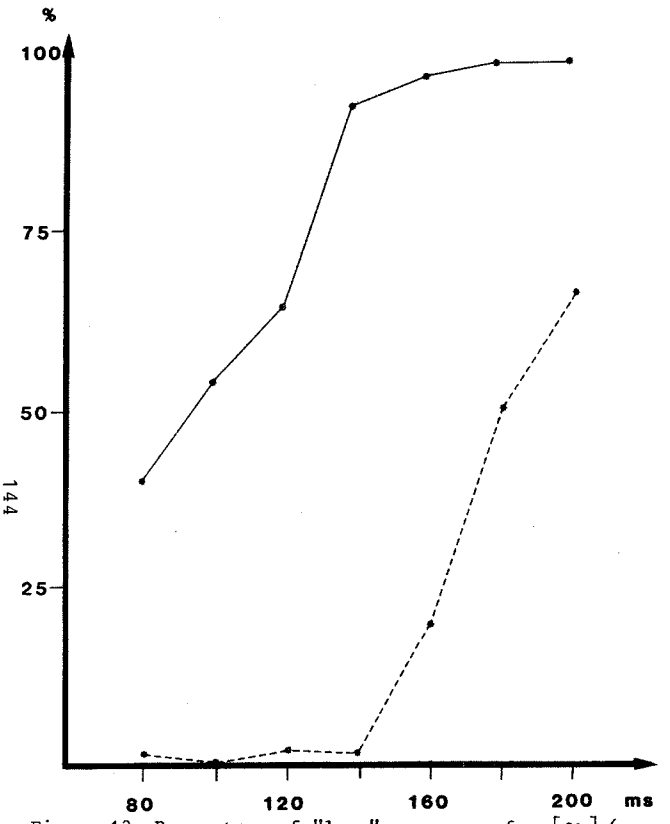

Figure 13 Percentage of "long"-responses for [ɑ:] (continuous line) and [a] (dashed).

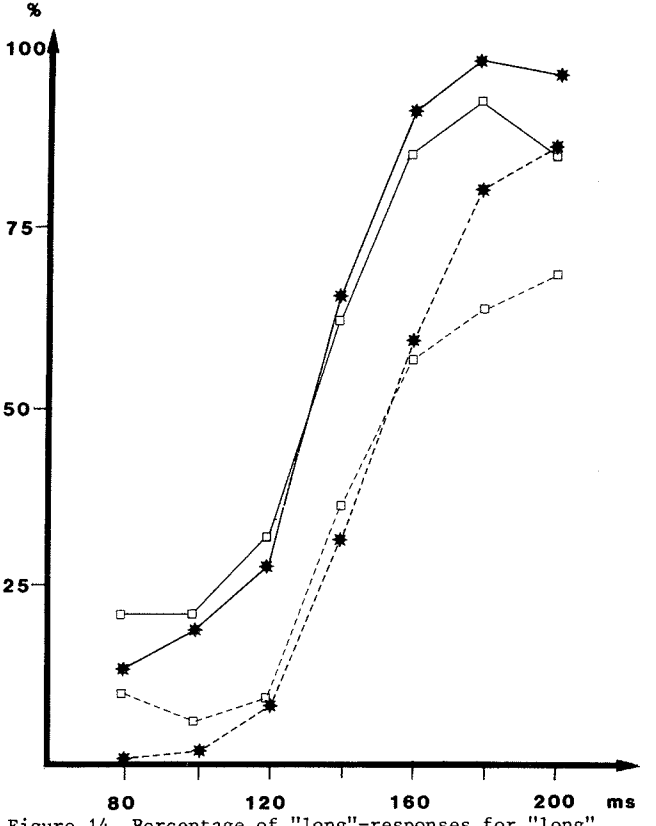

Figure 14 Percentage of "long"-responses for "long" vowels (continuous line) and "short" (dashed) for two groups of listeners, one from Central Sweden (stars) and one from Scania (squares).