PERCEPTION OF VOWEL DURATION AND SPECTRAL CHARACTERISTICS IN SWEDISH

Dawn M. Behne
dawn.behne@hf.ntnu.no
Norwegian University of Science and Technology
7055 Dragvoll, Norway
Tel: +47 73 59 83 09 Fax: +47 73 67 70

Peter E. Czigler and Kirk P. H. Sullivan
czigler@ling.umu.se kirk@ling.umu.se
Umeå University
S-901 87 Umeå, Sweden
Tel: +46 90 16 63 67 Fax: +46 90 16 63 77

ABSTRACT

This project re-examines the perceptual weight of vowel duration and the first two vowel formant frequencies as determinants of phonologically short and long vowels in Swedish. Based on listeners' responses to synthesized sets of materials for [I]-[i:] [ɔ]-[o:] and [a]-[ɑ:], results indicate that vowel duration is of primary importance for distinguishing [I] from [i:] and [ɔ] from [o:], whereas both formant frequencies and vowel duration were found to influence the perception of [a] from [ɑ:].

1. INTRODUCTION

In many languages, vowels can be characterized by their contrastive use of vowel quality and quantity. Acoustically, vowel quality is primarily correlated with the first and second formant frequencies of the vowel spectrum, based generally on the length of the pharyngeal-oral tract, the position of a constriction and the degree of constriction (e.g., [1]). In addition, vowel qualities may differ in their inherent duration (e.g., [2]). For example, a vowel quality involving a more extreme articulation may require more time if it is fully realized and consequently be longer in duration. A distinction in vowel quantity is generally realized acoustically by the duration of a vowel, with a phonologically long vowel having a duration which extends over more time than a phonologically short vowel. In addition, the greater amount of time associated with a phonologically long vowel may also be associated with an articulation using greater extremes of the vocal space than phonologically short vowels, and consequently may also affect the vowel spectrum.

The vowel system of Swedish has traditionally been described as having a phonological distinction between short (e.g., [a] in [tak] *tack* "thanks") and long vowels (e.g., [ɑ:] in [tɑ:k] *tak* "roof") (e.g., [3]). Elert [3] demonstrated that the duration of short vowels is in general approximately 65% of the duration of long vowels. Other research has suggested that the quantitative differences between phonologically long and short vowels are also realized by qualitative differences, some of them greater than others. In particular, phonologically long vowels are generally known to be articulated with more closure than short vowels with the open articulation of the vowel pair [ɑ:]-[a] being a possible exception.

In 1964, Hadding-Koch and Abramsson investigated the role of duration and spectral features of a vowel in conveying the distinction between phonologically short and long vowels[4]. For three vowel pairs, tape recordings were carefully spliced with differences of 10-15 ms, resulting in approximately 5-8 steps from the phonologically long vowel to the short vowel. Although the role of spectral characteristics could not be excluded from being an important perceptual cue, their results show that length is the primary parameter distinguishing Swedish vowels. These issues are also raised in a study addressing whether teaching Swedish spelling to children should be based on vowel quality or vowel quantity ([5]). From Elert's measures [6], the duration of synthesized Swedish vowels in a /hVs/ frame were successively adjusted between "long" and "short". Based on an identification task, results suggest that the distinction "long-short" was generally more important than the distinction in quality for the Swedish vowels. However, quality was more important for the vowels /a/ and /ʉ/, and both length and quality was important for the vowel /o/. Notably, this is the only known study which has examined the full set of 9 Swedish phonologically long-short vowel pairs. Unfortunately the report does not include a full methodological description and in other respects appears to be problematic.

These findings have laid a foundation for understanding the perceptual role of vowel length and spectral characteristics. Still, much room for further investigation remains. With the technical developments of the past 30 years offering greater possibilities for accurate control of experimental environments and manipulation of both vowel durations and vowel formant, new investigations are motivated. The goal of this project is to examine the perceptual weight of vowel duration and the first two vowel formant frequencies in distinguishing three pairs of phonologically short and long vowels.

2. Method

1.1 Materials

1.1.1 Recordings

Six phonotactically possible Swedish non-words were developed as targets. The targets contained one of six vowels ([ɪ, ɔ, a, i:, o:, ɑ:]) and in all cases the initial consonant was /k/ and the postvocalic consonant was /t/.

Audio recordings were made of a young adult male native speaker of standard Swedish (Stockholm dialect).

The speaker produced 10 randomized repetitions of the six target words in the sentence "Jag sa__ igen." ("I said __ again.") at his natural speaking rate.

From these recordings five measurements were made within each target word: vowel duration, closure duration of the postvocalic consonant, and the first (F1), second (F2) and third formant (F3) frequencies. For each of the six vowel conditions, the mean value of these measures for the ten repetitions was calculated and the utterance which best corresponded to the mean values was chosen for resynthesis.

1.1.2 Synthesis

The most representative productions for [i:]-[I], [o:]-[ɔ] and [ɑ:]-[a], were the basis for three pairs of resynthesized words. For each vowel pair, the selected words were used as extreme points of a 10x10 synthesis matrix. Each matrix had 10 equal-sized steps of vowel duration intermediating the two original vowels and at each step of vowel duration, there were 10 equally sized degrees of synchronized F1 and F2 adjustment, resulting in 100 resynthesized items for each vowel pair. Little difference was observed for the third formant frequency in the productions. Consequently, for this study F3 and higher formant frequencies were not adjusted. In addition, for each vowel pair, the mean duration of the postvocalic consonant closure duration from the selected items was calculated, and the duration of the postvocalic consonant closure was adjusted to this mean for all 100 items of the vowel set. This was done to increase the sensitivity of stimuli near the perceptual phoneme boundary and to limit the number of stimuli that would be presented to subjects to that which could be done in a single sitting.

The three sets of stimuli were resynthesized using the Kay Elemetrics LPC Parameter Manipulation/ Synthesis program. Beginning from the values for [i:], [o:] and [ɑ:], and adjusting the signal in step sized increments toward the values of for [I], [ɔ] and [a], respectively, three series of 100 resynthesized items were developed.

1.2 Procedure

Twenty native speakers of Swedish between 20 and 38 years old participated in the study.

For each trial, subjects heard a synthesized target word over headphones and at the same time two real words (vit - vitt, våt - vått, or fat - fatt) were presented on a monitor. The words on the monitor differed in phonological length and had the same phonemes as the original two vowels the synthesized item was based on. Subjects were asked to chose which of the two words had the same vowel as the one they heard, and to respond on a keyboard as quickly as possible. Subjects heard 5 randomized repetitions of each synthesized word, a total of 1500 items. Subjects responses and their reaction times for each trial were logged to a data file.

3. RESULTS

The mean percent of responses that were "vit", "våt", or "fat" was calculated for each condition. These are referred to as "long responses" in the following discussion. For each of the three vowel sets, phoneme boundaries were calculated by estimating the 50%-point and slope of the curve Two-way analyses of variance were calculated with duration step and spectral step as independent variables for the percentage of long responses and reaction time. Results are presented in Figure 1. The discuss of results presented here will focus on general patterns.

3.1 Vowel Duration

The effects of duration step on the percent of long responses for each of the three vowel sets are shown in the top row of Figure 1. Reliable differences in percent long responses due to vowel duration were found for all three vowel sets. As expected, for [i:]-[I] [$F=1242.34$; $p<.0001$], [o:]-[ɔ] [$F=553.39$; $p<.0001$] and [ɑ:]-[a] [$F=290.59$; $p<.0001$] a higher percentage of long responses was observed for synthesized items which were longest in duration, and a much lower percentage was found for shorter durations. However differences can be observed among the three vowel sets. Perceived long and short responses across the 10 duration steps were more distinctively divided in the [i:]-[I] and [o:]-[ɔ] sets than in the [ɑ:]-[a] as is evident from the shape of the s-curves in Figure 1. For both the [i:]-[I] and [o:]-[ɔ] sets, the items having the first 6 duration steps, a duration range of 168-101 ms for [i:]-[I] and 182-116 ms for [o:]-[ɔ], were perceived as phonologically long 81-100% of the time. However for [ɑ:]-[a] only 57-87% of the items with duration steps 1-6, a range of 160-109 ms, were judged long. The phoneme crossover point of 50% long—50% short responses is at 90 ms for the [i:]-[I] set, 105 ms for the [o:]-[ɔ] set, and 107 ms for the [ɑ:]-[a] set. The phoneme crossover point occurs at 65% of the overall duration range for both set [i:]-[I] (duration range=120 ms) and set [o:]-[ɔ] (duration range=118ms), whereas for set [ɑ:]-[a] this point occurs earlier, at 58% of the overall duration range (92ms). As would be expected, the percent long responses is lowest for the shortest items, duration steps 8-10 for all three vowel sets. However, only in set [i:]-[I] does the percent long responses reach as low as 4.5%. These high values are likely due to having developed the stimuli from phonologically long items and maintaining all of the acoustic attributes of those items other than the select few parameters manipulated in the resynthesis. Consequently, although the vowel duration of items was based on that of both phonologically long and short vowels, other subtle acoustic cues which typically occur with a natural phonologically short vowel were not available to listeners.

As the reaction times in the second row of Figure 1 illustrate, for set [i:]-[I] the mean reaction time increases markedly [$F=32.42$; $p<.0001$], from 791ms to 1196ms, at the phoneme crossover point between steps 6 and 7, and then decreases to about 1000ms for the shortest items of the set at steps 8-10. A similar pattern is observed for set [o:]-[ɔ] [$F=44.78$; $p<.0001$], increasing from about 744ms to 1218ms at the phoneme crossover point. In addition, reaction times remained high for the shortest items of the set. For the [ɑ:]-[a] set, reaction times were consistently high, above 1000ms, which with the high standard deviations and flatter s-curve observed for percent long responses, reflects the greater difficulty

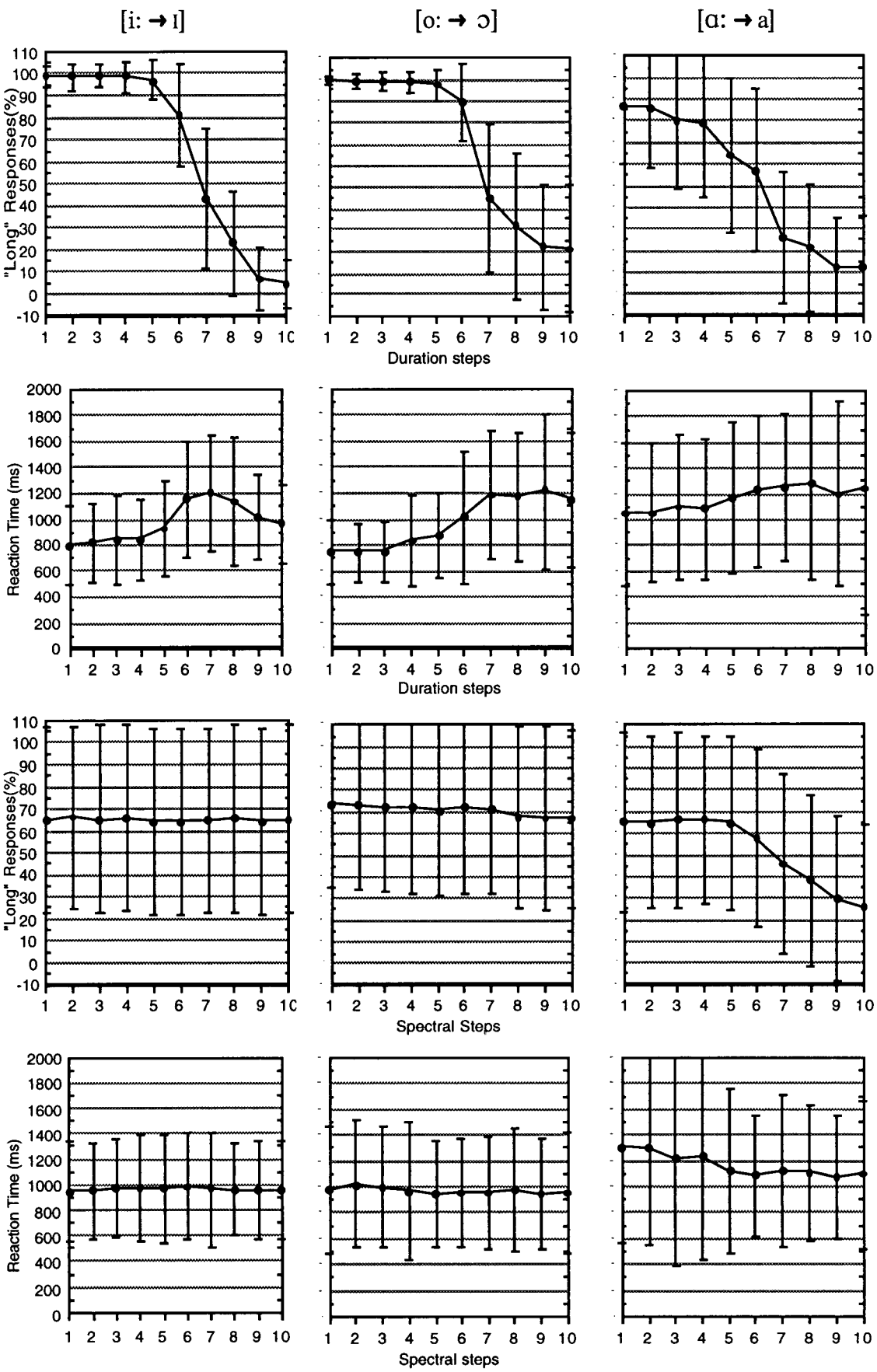

Figure 1. For vowel sets [i:]-[ɪ], [o:]-[ɔ], and [ɑ:]-[a], mean percent long responses and mean reaction times are plotted for the 10 synthesized duration steps and spectral steps. Standard deviations for each step are shown by vertical bars.

subjects had responding to items in this set. However, like for the other two vowel sets, reaction times for set [ɑ:]-[a] still increased near the phoneme crossover point, for this set from 1044ms to 1277ms [F=3.73; p<.0001].

3.2 First and Second Formants

The effects of spectral step on the percent of long responses for sets three vowel sets [i:]-[I], [o:]-[ɔ], and [ɑ:]-[a] are presented in the third row of Figure 1. Three different patterns of results are noticeable across the three vowel sets.

For vowel set [i:]-[I] no reliable differences in the percent long responses attributable to the concurrent adjustment of F1 and F2 frequencies were observed [F=0.36; n.s.]. Across the 10 spectral steps the mean percent long responses is consistently about 65%. The tendency toward slightly more than 50% long responses is likely due to developing the stimuli from [i:].

Although the pattern of long responses across spectral steps for set [o:]-[ɔ] appears similar to set [i:]-[I], a reliable difference was observed [F=2.76; p<.0033]. The mean percent long responses was slightly greater for spectral steps 1-6 than for spectral steps 7-10. However, notably, the adjustments of F1 and F2 frequencies alone were not enough either to strongly elicit a high percentage of long responses or to shift the mean of subjects' responses from more than 50% long responses to less than 50% long responses as would be expected if F1 and F2 frequencies were serving a role in categorically distinguishing [o:] from [ɔ].

Like sets [i:]-[I] and [o:]-[ɔ], the ceiling of the percent long responses for the [ɑ:]-[a] set was reached at about 65%. However unlike set [i:]-[I], a reliable difference in percent long responses due to the frequencies of F1 and F2 was observed [F=76,30; p<.0001], and unlike the [o:]-[ɔ] these spectral changes did appear to serve, to some degree, as a cue for distinguishing [ɑ:]-[a]. A higher percentage of long responses was given by subjects for items which were synthesized with F1 and F2 values closest to the original phonologically long vowels, and a lower percentage of long responses was observed of items spectrally more like phonologically short vowels. The items having the first 5 spectral steps were, with a range of 354-529Hz for F1 and 882-1091Hz for F2, were perceived as phonologically long only 65% of the time which, for having been developed from phonologically long to phonologically short, appears to be comparable to chance. As would be expected, the percent long responses is lowest for spectral step 10, however even in this case percent long responses only reaches as low as 25.8%.

Reaction times associated with spectral steps are presented in the bottom row of Figure 1. Mean reaction times for sets [i:]-[I] [F=0.21; n.s.] and [o:]-[ɔ] [F=0.72; n.s.] are reliably stable and slightly high at about 1000ms across the 10 spectral steps. However, mean reaction times for set [ɑ:]-[a] are generally even higher, decreasing gradually from 1310ms at spectral step 1 to 1094ms at spectral step 10. This finding, consistent with the close-to-chance percent long responses observed across spectral steps 1-5 for [ɑ:]-[a], suggests that, based on spectral information alone, there was in an increased difficulty with the task for items spectrally most like [ɑ:], comparable to that observed at the phoneme crossover points for duration steps. In addition, corresponding to the divergence from near-chance percent long responses at spectral steps 7-10, the difficulty of the task and corresponding reaction times, to some limited degree, appears to decrease.

4. CONCLUSION

The duration and resonance characteristics of vowels both play a role in distinguishing phonological length in Swedish. Results based on subjects responses and the corresponding cognitive load of the perception task reflected the concurrent patterns of standard deviation and reaction times, demonstrate two general patterns. Vowel duration appears to serve as the most dominant cue to listeners in distinguishing [i:] from [I] and [o:] from [ɔ], and although the results show no affect of F1 and F2 frequencies on perceived phonological length for these vowel pairs, other attributes of the vowels which were not addressed in this study did appear to progressively affect the variance and reaction time of responses to items acoustically most distant from the phonologically long vowels the synthesis was based on. For [ɑ:] and [a] the perceptually influence of vowel duration and spectral attributes appears to be more complex. The results clearly show that vowel duration serves as a dominant perceptual cue when distinguishing [ɑ:] and [a]. In addition, resonance also affects the perception of [ɑ:] versus [a]. In particular, the results suggest that although vowel duration is used in the perception of both [ɑ:] and [a], the first two formant frequencies appear to assist in the perception of [a], but not [ɑ:]. Addition acoustic cues must also be available for the clear perception of [ɑ:] in natural productions, although evidence is not available from the current study. Nevertheless, one can speculate that further investigation of the role of postvocalic consonant duration and investigation of other phonological vowel pairs of Swedish and other languages may shed light on this and other related issues.

ACKNOWLEDGMENTS

The authors thank Ola Andersson, research technician at the Dept of Phonetics, Umeå University, for developing the program used for the perception test.

REFERENCES

[1] K. Stevens & A. House "Development of a quantitative description of vowel articulation", *JASA* 27, pp. 484-493, 1955.

[2] G. Peterson & I. Lehiste "Duration of Syllable Nuclei in English" *JASA* 32, pp. 693-703, 1960.

[3] C-C. Elert "Allmän och svensk fonetik", Almqvist & Wiksell: Stockholm, 1966.

[4] K. Hadding-Koch & A. Abramson "Duration versus spectrum in Swedish vowels: some perceptual experiments" *Stud. Ling.*2, pp. 94-107, 1964.

[5] K. Johansson "Bör dubbelteckningsmetodiken bygga på LÄNGD- eller KLANGFÄRGS-skillnader?" *Lund University, Lärarhögskolan I Malmö, Rapport* 2, 1981.

[6] C-C. Elert "Phonologic studies of quantity in Swedish", Almqvist & Wiksell: Stockholm, 1964.