ACOUSTIC CHARACTERISTICS OF PERCEIVED QUANTITY AND QUALITY IN SWEDISH VOWELS

Dawn M. Behne*, Peter E. Czigler** and Kirk P. Sullivan**

*English Department
Norwegian University of Science and Technology, Norway

**Phonetics Department
Umeå University, Sweden

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

INTRODUCTION

In many languages, vowels can be characterized by their contrastive use of vowel quality and quantity. Vowel quality refers to the relative phonological resonance or timbre of a sound, whereas vowel quantity refers to the phonologically distinctive length relative to one or more vowels of similar quality.

The Swedish vowel system has traditionally been described as having a phonological distinction between short (e.g., [a] in [tak] *tack* "thanks") and long vowels (e.g., [ɑː] in [tɑːk] *tak* "roof") (e.g. Elert, 1966). Accompanying this vowel length distinction is an inverse affect on postvocalic consonant length (Elert, 1964). Elert demonstrated that the duration of short vowels is, in general, approximately 65% of the duration of long vowels. Further research has suggested that the quantitative differences between phonologically long and short vowels are also realized by resonance. In particular, phonologically long vowels are generally known to be articulated with more closure than short vowels (Gårding, 1974), with the open articulation of [ɑː] and [a] being a possible exception.

In 1964, Hadding-Koch and Abramson investigated whether the duration or spectral aspects of a vowel had the more dominant role in distinguishing phonologically short and long vowels. For three vowel pairs, tape recordings were carefully spliced with differences of 10-15ms, 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 (Johansson, 1981). From measures presented in Elert (1964), the duration of synthesized Swedish vowels were successively adjusted from "long" to "short" and "short" to "long". 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 formants, 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. To this end, this study will examine how vowel duration and the first and second formant frequencies of vowels each affect the perception of vowel quantity and quality.

METHOD

Materials

Six non-words were developed as targets. The targets were phonotactically possible in Swedish and

contained one of six vowels, [ɪ, ɔ, a, iː, oː, ɑː]. In all cases the initial consonant was /k/ and the postvocalic consonant was /t/.

Audio recordings were made of a young adult male student of phonetics who is a native speaker of standard Swedish (Stockholm dialect). The speaker produced 10 randomized repetitions of the six target words imbedded in the carrier sentence "Jag sa ___ igen." ("I said ___ again."). The speaker was asked to speak at his natural speaking rate.

From these recordings five measurements were made within each target word using ESPS/*waves+*™. Vowel duration was measured from the onset to the end of the periodic energy. The closure duration of the postvocalic consonant was measured from the start of the closure to the beginning of the release. The first (F1), second (F2) and third formant (F3) frequencies were measured at the center of the vowel's most evident steady state. 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. The means for the six conditions and the measured values from the six items which were the basis for resynthesis are presented in Table 1.

The most representative productions for [iː]-[ɪ], [oː]-[ɔ], and [ɑː]-[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 equal-sized steps 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 the frequency of F3 and higher formant frequencies were not adjusted. Step sizes for F1, F2 and vowel duration are presented in Table 2, as are the corresponding values from the selected reference items (also in Table 1) which served as endpoints for the series of resynthesized items for each vowel set.

Vowel Condition	Source of Values	Vowel				Stop Closure Duration (ms)
		F1 (Hz)	F2 (Hz)	F3 (Hz)	Duration (ms)	
[iː]	Means	256 (9)	2270 (50)	3339 (74)	166 (28)	111 (10)
	Selected item	262	2254	3413	168	102
[ɪ]	Means	280 (10)	2220 (59)	3129 (113)	54 (7)	167 (15)
	Selected item	274	2215	3126	48	154
[oː]	Means	309 (13)	574 (50)	2159 (288)	177 (15)	107 (8)
	Selected item	295	528	2020	182	110
[ɔ]	Means	359 (32)	805 (33)	2555 (322)	60 (7)	160 (8)
	Selected item	378	788	1942	64	154
[ɑː]	Means	377 (32)	895 (20)	2338 (173)	167 (23)	107 (9)
	Selected item	354	882	2120	160	100
[a]	Means	761 (20)	1388 (61)	2339 (47)	58 (7)	160 (17)
	Selected item	747	1362	2394	68	158

Table 1. For [iː] [ɪ] [oː] [ɔ] [ɑː] and [a], means from the 10 repetitions and the measured values from the productions selected for resynthesis are presented for F1, F2 and F3 frequencies, vowel duration, closure duration of the postvocalic stop. Standard deviations are shown in parentheses.

Vowel Pairs	Spectral Steps				Duration Steps (ms)	
	F1 (Hz)		F2 (Hz)			
	Reference	Step size	Reference	Step size	Reference	Step size
[iː]	262	-1	2254	-4	168	13
[ɪ]	274		2215		48	
[oː]	295	-9	528	-29	182	13
[ɔ]	378		788		64	
[ɑː]	354	-44	882	-53	160	10
[a]	747		1362		68	

Table 2. For each vowel pair, step sizes and corresponding values from the reference vowels are shown for F1, F2 and vowel duration.

Since the duration of a postvocalic consonant is also known to decrease as vowel length increases in Swedish (Elert 1964), the duration of the postvocalic consonant was also adjusted. For each vowel pair, the mean duration of the postvocalic consonant closure duration from the selected items was calculated. The duration of the postvocalic consonant closure was adjusted to this mean for all 100 items of the vowel set. On this basis, for the series of resynthesized items for [iː]-[ɪ], [oː]-[ɔ] and [aː]-[a], the duration of the postvocalic consonant closure was adjusted to 128 ms, 132 ms, and 129 ms respectively. This was done to increase the sensitivity of stimuli near the phoneme boundary in the perception task and at the same time to limit the number of stimuli.

The three sets of stimuli were resynthesized using the Kay Elemetrics LPC Parameter Manipulation/ Synthesis program. The numerical editor was used to adjust the values of the first two vowel formants and scale the time of the LPC frames for the vowel by the step sizes shown in Table 2. The closure period of the postvocalic consonant was adjusted to the corresponding constant length. Beginning from the values for [iː], [oː] and [aː] and adjusting the signal in step sized increments toward the values of [ɪ], [ɔ] and [a] respectively, three series of 100 resynthesized items were developed.

Procedure

Twenty native speakers of Swedish (11 females and 9 males) participated in the study. All of the subjects were between 20 and 38 years old (median = 24 years), had no known history of speech or hearing impairment and came from different parts of Sweden.

Subjects were seated wearing headphones at a computer terminal with a monitor and keyboard. For each trial, subjects heard a synthesized target word and at the same time two real words (vit - vitt, våt - vått, or fat - fatt) were dimly presented on the monitor in 24 point font. Simultaneous with the end of the auditory signal, the word pair became more visible. The words 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. They were asked to respond as quickly as possible and were allowed up to 10 seconds to respond before the beginning of the next trial, although most subjects never encountered this upper limit. Subjects heard 5 randomized repetitions of each synthesized word, a total of 1500 items (3 sets x 100 items x 5 repetitions). Subjects responses and their reaction times for each trial were logged to a data file. Before starting the experiment, subjects had three practice trials, and after each set of 50 trials, subjects had the opportunity to take a short break. Subjects were run in groups of 1-4 at a time

RESULTS

The mean percent 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. Statistical differences with a probability less than five percent were accepted as reliable. Nonsignificant contrasts are indicated by "n.s.". Results are presented in Figure 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 [iː]-[ɪ] [$F=1242.34$; $p<.0001$], [oː]-[ɔ] [$F=553.39$; $p<.0001$] and [aː]-[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 [iː]-[ɪ] and [oː]-[ɔ] sets than in the [aː]-[a] as is evident from the shape of the s-curves in Figure 1. For both the [iː]-[ɪ] and [oː]-[ɔ] sets, the items at the first 6 duration steps, a duration range of 168-101 ms for [iː]-[ɪ] and 182-116 ms for [oː]-[ɔ], were perceived as phonologically long 81-100% of the time. However for the [aː]-[a] set only 57-87% of the items at duration steps 1-6, a range of 160-109 ms, were judged long. The percent long responses decreases sharpest between duration steps 6 and 7 for all three vowel sets, with a slope of -37.5 %/duration step (-2.8 %/ms) for [iː]-[ɪ], 44.4%/duration step (-3.4 %/ms) for [oː]-[ɔ] and -31.3 %/duration step (-3.1 %/ms) for [aː]-[a]. The phoneme crossover point of 50% long—50% short responses is at 90 ms for the [iː]-[ɪ] set, 105 ms for the [oː]-[ɔ] set, and 107 ms for the [aː]-[a] set.

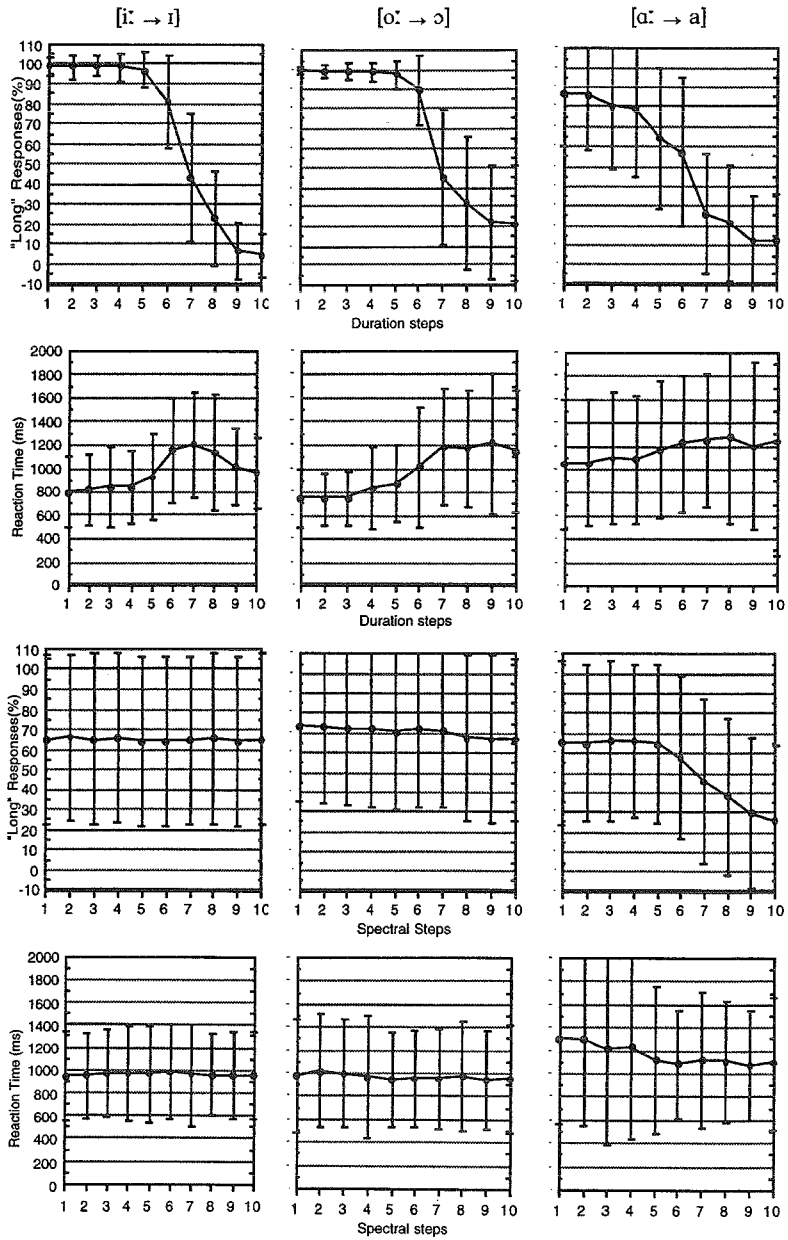

Figure 1. For vowel sets [iː]-[ɪ], [oː]-[ɔ], and [ɑː]-[a], mean percent long responses and mean reaction times are plotted for the 10 synthesized duration steps and spectral steps. Standard deviations for each step are shown by vertical bars.

The phoneme crossover point occurs at 65% of the overall duration range for both set [i:]-[ɪ] (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:]-[ɪ] does the percent long responses reach as low as 4.5%. For [o:]-[ɔ] the percent long responses reaches only as low as 21% and for [ɑ:]-[a] 12%. These high values is likely due to having developed the stimuli from phonologically long items and maintaining all the acoustic attributes, with the exception of 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 might occur with a natural phonologically short vowel were not available to listeners.

The results for the [i:]-[ɪ] and [o:]-[ɔ] sets also differ from the [ɑ:]-[a] set in observed patterns of standard deviation. For the [i:]-[ɪ] and [o:]-[ɔ] sets standard deviation increases markedly as vowel duration gets shorter, most likely attributable to having developed the stimuli from phonologically long vowels. Standard deviation also increases near the phoneme crossover point, at duration steps 6- 8, reflecting the added difficulty of the task near the phoneme boundary. For [ɑ:]-[a] the standard deviation across duration steps remains relatively stable and is generally higher than for both [i:]-[ɪ] and [o:]-[ɔ], increasing only slightly at the phoneme crossover point.

Corresponding to the pattern of standard deviation, are reaction times. As is shown in the second row of Figure 1, for set [i:]-[ɪ] 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, consistent with the percent long responses and standard deviation observed for set [o:]-[ɔ], reaction times remained high for the shortest items of the set, at steps 8-10. 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 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].

First and second formants

The effects of spectral step on the percent of long responses for sets three vowel sets [i:]-[ɪ], [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:]-[ɪ] 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 the stimuli having been developed from [i:].

Although the pattern of long responses across spectral steps for set [o:]-[ɔ] appears similar to set [i:]-[ɪ], 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:]-[ɪ] and [o:]-[ɔ], the ceiling of the percent long responses for the [ɑ:]-[a] set was reached at about 65%. However unlike set [i:]-[ɪ], 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. The percent long responses decreases gradually between spectral steps 6 and 10, with a slope of --11.6 %/spectral step (-0.3 %/Hz for F1 and

-0.2 %/Hz for F2) between spectral steps 6 and 7. The phoneme crossover point is at 600Hz for F1 and 1176Hz for F2 , occurring at 63% of the overall frequency range for both F1 (frequency range=393Hz) and F2 (frequency range=470Hz). 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%.

Although the pattern of percent long responses differs among the three vowel sets, the observed standard deviation is consistently high, at about 40%, across the spectral steps for sets [iː]-[ɪ], [oː]-[ɔ] and [ɑː]-[a]. This is higher than the standard deviation observed at the crossover point of the duration steps for any of the vowel sets, but notably the standard deviation of percent long responses does not have the tendency to increase across spectral steps 1 through 10 as it did across duration steps.

Reaction times associated with spectral steps are presented in the bottom row of Figure 1. Consistant with the standard deviations for spectral steps, the mean reaction times for sets [iː]-[ɪ] [F=0.21; n.s.] and [oː]-[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 [ɑː]-[a] are generally even higher, decreasing gradually from 1310ms at spectral step 1 to 1094ms at spectral step 10. This finding, consistant with the close-to-chance percent long responses observed across spectral steps 1-5 for [ɑː]-[a], suggests that, based on spectral information alone, there was in an increased difficulty with the task for items spectrally most like [ɑː], 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.

CONCLUSIONS

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 [iː] from [ɪ] and [oː] 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 [ɑː] 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 [ɑː] and [a]. In addition, resonance also affects the perception of [ɑː] versus [a]. In particular, the results suggest that although vowel duration is used in the perception of both [ɑː] and [a], the first two formant frequencies appear to assist in the perception of [a], but not [ɑː]. Addition acoustic cues must also be available for the clear perception of [ɑː] 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, and have appreciated supporting research grants from NOS-H and NorFA.

REFERENCES

Elert C-C (1964). *Phonologic studies of quantity in Swedish.* (Almqvist & Wiksell: Stockholm).

Elert, C-C. (1966) *Allmän och svensk fonetik.* (Almqvist & Wiksell: Stockholm).

Hadding-Koch, K. & Abramson, A.S. (1964) "Duration versus spectrum in Swedish vowels: some perceptual experiments", Studia Linguistica, 1964:2, 94-107 .

Johansson, K. (1981) "Bör dubbelteckningsmetodiken bygga på LÄNGD- eller KLANGFÄRGS-skillnader?" Lund University, Lärarhögskolan I Malmö, Inst. för stadie- och specialmetodik, Rapport 2.