PERCEPTION OF SWEDISH VOWEL QUANTITY: TRACING STAGES OF DEVELOPMENT

Dawn Behne*, Peter Czigler** and Kirk Sullivan**

*Norwegian University of Science and Technology, Trondheim Norway
**Umeå University, Umeå Sweden

ABSTRACT: Swedish adults generally use vowel duration to identify Swedish long and short vowel quantities. However, when the duration of a vowel is relatively long (due, e.g., to inherent duration, or vowel lengthening), adults listeners may also make use of vowel spectra to distinguish vowel quantities. In this respect, use of the vowel spectrum in special cases (e.g., when identifying the quantity of inherently long vowels) might be seen as the result of perceptual fine tuning to improve the processing efficiency of identifying vowel quantities. This project investigates how young, developing listeners acquire the perceptual use of vowel duration and the vowel spectra for identifying long and short vowel quantities. Of particular interest is whether younger listeners consistently use vowel duration when categorizing vowel quantities and whether the pre-adult listeners are as likely to use spectral cues to identify the vowel quantity of the inherently long vowels as adult listeners.

INTRODUCTION

Swedish is traditionally described as having a distinction between long and short vowel quantities (Elert 1964). This distinction is typically realized acoustically through differences in vowel duration, with a long quantity having a duration that extends over more time than a short quantity. The greater amount of time associated with a long vowel quantity also allows for an articulation using greater extremes of the vocal space than short vowel quantities, and consequently may also affect the vowel spectrum, in particular the first and second formant frequencies (F1 and F2).

In a classic perception study with Swedish adults, Hadding-Koch and Abramson (1964) investigated whether vowel duration or spectral attributes of a vowel have a more dominant perceptual role in distinguishing vowel quantities in Swedish. They established that vowel duration was a primary perceptual cue to Swedish vowel quantity, but were not able to exclude the possible role of the vowel spectrum.

More recent studies (Behne, Czigler & Sullivan 1996, 1998) have returned to this issue using resynthesized Swedish words with 10 stepwise adjustments of vowel duration and 10 stepwise adjustments of F1 and F2. The results indicate that listeners use vowel duration to identify the quantity of vowels in a continuum from [iː] to [ɪ] and from [oː] to [ɔ], whereas from [ɑː] to [a] adult listeners use both vowel duration and, to a lesser extent, spectral attributes of the vowel. These findings suggest that when a vowel has a relatively long duration (e.g., due to factors such as inherent vowel duration or postvocalic voicing), adult listeners will make use of both vowel duration and vowel spectrum to identify a vowel's quantity.

If vowel duration is a primary cue for identifying vowel quantity, use of the vowel spectrum in special cases (e.g., when identifying the quantity of inherently long vowels) might be seen as the result of perceptual fine tuning to improve the processing efficiency of identifying vowel quantities. If so, we would expect to see a change in how developing children use vowel spectra to identify vowel quantities, with vowel duration being used for vowel quantity identification at a relatively early age and the use of vowel spectra as a cue for vowel quantity identification coming later.

The project aims to offer an initial investigation of the developmental pattern of how vowel duration and the vowel spectra come to be used to distinguish long and short vowel quantities. Of particular interest is whether younger listeners consistently use vowel duration when categorizing vowel quantities and whether the pre-adult listeners are as likely to use spectral cues to identify the vowel quantity of the inherently long vowels (i.e. [ɑː] and [a]) as adult listeners.

METHOD

Materials

The materials used are identical to with those used with adult listeners in Behne et al (1996, 1998).

Recordings. A set of six /kVt/ pseudo words were used as targets. Each word was phonotactically possible in Swedish and contained one of the vowels [ɪ, ɔ, a, i:, o:, ɑ:].

Audio recordings were made of a young adult native male speaker of the Stockholm dialect of Swedish producing 10 random repetitions of the six target words in the sentence *Jag sa ___ igen.* ("I said ___ again.") at his natural speaking rate.

Measurements. From the 10 recorded productions of each target word, ESPS/*waves+*™ was used to measure the vowel duration, and the first three formant frequencies of the vowel (F1, F2 and F3) measured at the center of the vowel's most evident steady state. The closure duration of the postvocalic /t/ was also measured.

For each of the five measures, means were calculated from the 10 repetitions of the six target words. The production of a target word that best corresponded to the mean values for the target word were used as the basis for resynthesis. These most representative items will be referred to as "representative productions".

Word sets	Steps	Adjusted parameters		
		Vowel spectrum		Vowel duration
		F1 (Hz)	F2 (Hz)	
ki:t	step 1	262	2254	168
↓	step size	1	-4	-13
kɪt	step 10	274	2215	48
ko:t	step 1	295	528	182
↓	step size	9	29	-13
kɔt	step 10	378	788	64
kɑ:t	step 1	354	882	160
↓	step size	44	53	-10
kat	step 10	747	1362	68

Table 1: Parameter settings of the vowel and postvocalic consonant for the three sets of resynthesized /kVt/ words. In each set the vowel duration and F1-F2 were adjusted in 10 steps, as the closure duration of postvocalic /t/ was held constant.

Synthesis. The Kay Elemetrics LPC Parameter Manipulation/Synthesis program was used to resynthesize three sets of 100 words each. The synthesis of a set was started from the representative word containing a long vowel quantity (i.e. [ki:t], [ko:t] or [kɑ:t]). Within a set, vowel duration was first adjusted in 10 equal-sized steps toward the measured vowel duration of the selected word containing a short vowel quantity (i.e. [kɪt], [kɔt] or [kat] respectively). This resulted in ten synthesized words having vowel formant patterns identical to the long vowel quantity, but differing in vowel duration. Then, for each of the ten words with different vowel durations, F1 and F2 frequencies were simultaneously adjusted in ten steps, resulting in 100 resynthesized items differing both in vowel duration and vowel formant patterns. This was done to produce a set of synthesized words from [ki:t] to [kɪt], from [ko:t] to [kɔt], and from [kɑ:t] to [kat]. The vowel durations, formant frequencies and step sizes are summarized in Table 1.

Since little measured difference was observed in the F3 frequency of phonologically long and short vowels in the productions, the frequency of F3 and higher formants was unchanged in the synthesized items, and consequently was the same as the phonologically long vowel of the pair.

The duration of a postvocalic consonant is known to decrease as vowel length increases in Swedish (e.g., Elert 1964). This pattern was also observed from the durations of postvocalic /t/ measured in the

current study. To control possible perceptual effects of the postvocalic consonant duration, for all 100 items of a set, the /t/ closure duration was set at the mid-value between the /t/ closure duration postvocalic to the long and short vowel quantities.

Identification task

Pre-adult native Swedish listeners participated in the study. These included ten 14-year olds (four males and six females) and seventeen 15-year olds (five males and twelve females). In addition, eight 9-year olds (three males and five females) were included as a preliminary test of how younger listeners respond. The participants were all living in Umeå, Sweden at the time of the experiment, but had not necessarily always lived there.

The listeners were seated wearing headphones at a computer terminal with a monitor and mouse. For each trial, subjects heard a synthesized word and, at the same time, two real words (vit - vitt, våt - vått, or fat - fatt) were presented on the monitor. The two words on the monitor had the same vowel quality and postvocalic consonant as the word heard over the headphones, but differed from each other in vowel quantity.

The listeners were instructed to use the mouse to click on the word that rhymed with 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 subjects rarely encountered this upper limit.

The 14- and 15-year olds heard 5 randomized repetitions of each synthesized word for a total of 1500 items (3 word sets x 100 items x 5 repetitions). In order to reduce the number of trials for the younger children, the 9-year olds heard 50 items randomly selected from the three sets of 100 items, and had the option to do another 50 randomly selected trials if they chose. Since this results in a relatively small number of trials, the result for the 9-year olds is treated as preliminary. Before starting the experiment, the listeners had three practice trials, and after each set of 50 trials, they had the opportunity to take a short break.

Listneners´ responses and reaction times for each trial were logged to a data file.

RESULTS

Each group´s responses and reaction times were analyzed separately for each of the three sets of synthesized words. For each word set the mean percent of trials when listeners chose the rhyme word containing a long vowel quantity was calculated at each duration and spectral step. It should be noted that the materials used in the experiment are expected to slightly skew perception toward a long response since the resynthesis was done starting from the word with a long vowel quantity in each word set.

In Figure 1 the percent long responses and reaction times for the 10 duration steps and 10 spectral steps are presented for each of the three sets of synthesized words ([kiːt]-[kɪt], [koːt]-[kɔt] and [kɑːt] - [kat]). Responses and reaction times for the three groups of pre-adult listeners (9-, 14- and 15-year olds) are presented with those for the adults from Behne et al (1996). A plotted point at each duration step is an average of the 10 spectral steps for that duration step. Likewise, a plotted point at each spectral step is an average of the 10 duration steps for that spectral step.

With a passing look across the plotted data in Figure 1, one sees that the results for the 9-year olds take a notably different form than the 14- and 15-years olds.

Nine-year olds

In Figure 1 the results for the 9-year olds plotted across the duration steps and the spectral steps are generally more irregular than for older listeners. Recall that, as part of a preliminary investigation, the number of trials for the eight 9-year olds was much smaller than what was attained for the 14- and 15-year olds and adults. This and the random presentation of only 50 out of 300 items (3 word sets x 100 items in each set) to the 9-year olds lead to an ununiform number of trials for each of the 100 items in each word set, and across the word sets. Consequently, the unsmooth curves in Figure 1 are not completely unexpected. Whether the results can be fully accounted for on these grounds, is open to further investigation.

Based on the results gathered for 9-year olds so far, the percent long responses (upper half of Figure 1) suggest that the listeners are generally having a hard time with categorizing quantity using either vowel duration or the vowel spectra. This is the case for all 3 word sets. This can be seen from the non-categorical curves around 50% for percent long responses for vowel duration and for spectral steps in Figure 1. If this pattern becomes more evident with additional data, this would suggest that the use of perceptual cues to identify vowel quantity is acquired relatively late; that is, beyond 9 years old. While this not very likely, more data will be needed to say more.

The reaction times in the lower half of Figure 1 show that the range of reaction times by the 9-year olds is very similar to those by the adults, for the most part between 900ms and 1400ms. For data based on a small sample of trials, they are also relatively consistent. This reinforces impression the 9-years gave while participating in the task; that is, that they knew what the task was about and were not responding haphazardly.

Fourteen- and fifteen-year olds

As can be seen from the plotted s-curves across the 10 steps of duration steps (first row) for the three sets of materials in Figure 1, both groups of teenagers show clear use of vowel duration to identify vowel quantity. Like the adults, both 14- and 15-year old listeners show this general pattern for all three sets of materials. Comparing the two groups of pre-adults, the pattern of responses by the 14-year olds show a notably greater use of vowel duration for identifying vowel quantity. A closer look shows other differences in the use of vowel duration among the three groups. For the [ki:t] - [kɪt] word set, neither the 14- nor 15-year olds were able to clearly identify the short vowel quantity as consistently as the adults. Although this may be due to the skewed nature of the material preparation, it nevertheless is more evident for the pre-adults than the adults. A similar tendency is seen for the other two sets of materials, in particular for the 15-year olds, and is a point that deserves attention in further analyses of the data and future stages of the project.

Listeners' responses across the 10 spectral steps of the three word sets show a different pattern than was observed across the duration steps. Responses based on simultaneous adjustments of F1 and F2 show no sign of categorization in the [ki:t] - [kɪt] or [kɔ:t] - [kɔt] word sets, as can be seen from the flat curves across the 10 spectral steps. The pre-adults' responses are comparable to the adult responses for these materials. The pre-adults also show tendencies similar to the adults for the [kɑ:t] - [kat] word set, however they do not reach the same use of vowel spectra for categorizing vowel quantity as the adults. Notably, while the 14-year olds only show a slight tendency in this direction, the 15-year olds responses come closer to the adult pattern.

Supporting the patterns observed from the listeners' responses are the corresponding reaction times. The teenage listeners consistently had shorter reaction times than adults, but with essentially no difference between the 14- and 15-year olds. Their reaction times also did not vary as much across the 10 duration steps and spectral steps as the adult reaction times did. In other respects, however, the two groups of young listeners had reaction time patterns that were not unlike those of the adults. The reaction time curves increase slightly across the 10 duration steps for all three word sets in the same general pattern as is observed for the adult reaction times. Across the 10 spectral steps, their pattern of reaction times are as unchanging as the adults', with the possible exception of the [kɑ:t] - [kat] word set.

CONCLUSIONS

Two points have been of particular interest in this investigation, whether pre-adult listeners make use of vowel duration as a primary perceptual cue when identifying vowel quantity, and whether pre-adults listeners show signs of developing a more fine-tuned use of the available perceptual cues to vowel quantity.

As has been found with adults previously (Hadding-Koch & Abramson, 1964, Behne et al 1996, 1998), the younger listeners were expected to use vowel duration as a primary perceptual cue when identifying vowel quantity. For both 14- and 15-year olds, responses and reaction times show that vowel duration is clearly a more dominant perceptual cue for categorizing vowel quantity than spectral

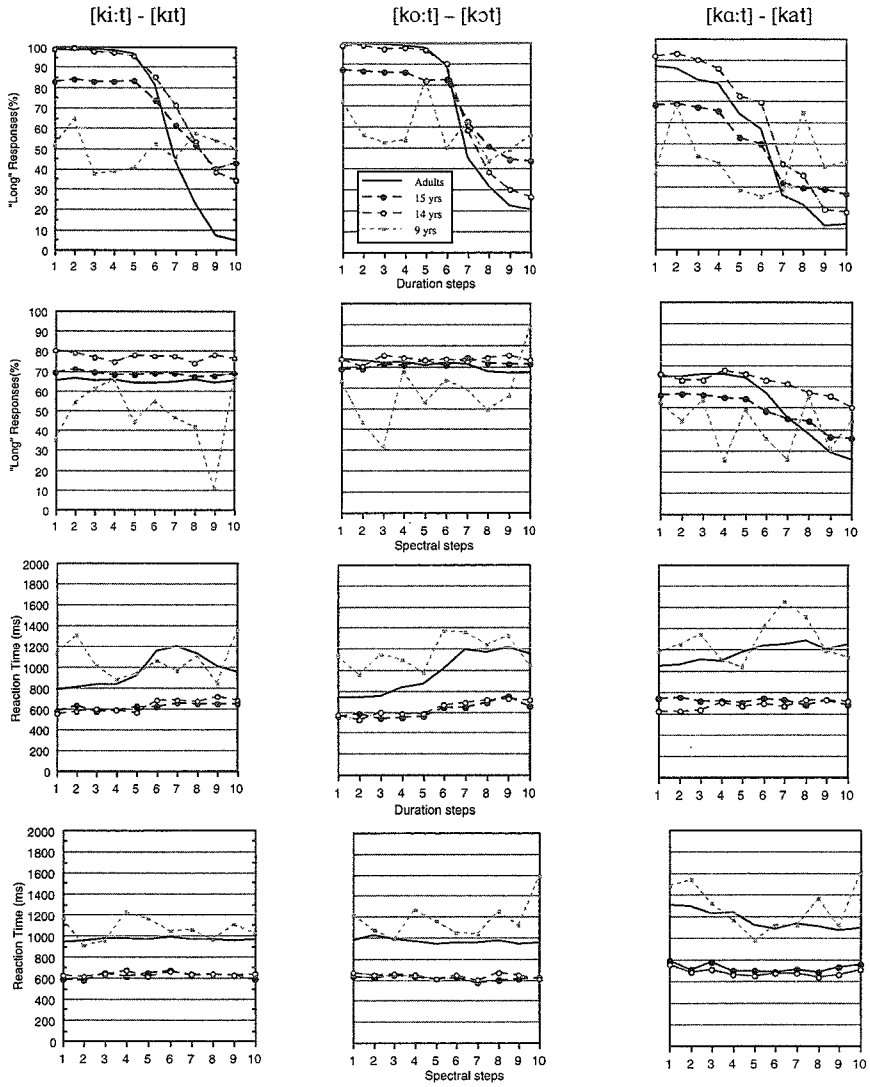

Figure 1: For the three sets of materials ([ki:t] - [kɪt], [ko:t] – [kɔt], and [kɑ:t] - [kat]), mean percent long responses and mean reaction times are plotted for the 10 synthesized duration steps and spectral steps.

information. The dominant role of vowel duration as a perceptual cue for identifying vowel quantity has lead us to speculate that it is a more fundamental, and consequently acquired earlier by native listeners. (See Bohn 1995 for a related point for non-native listeners.)

Among the three vowel pairs studied (Behne et al 1996, 1998), the use of spectral information by adults is unique to categorizing [ɑ:] and [a]. The current study shows a developmental progression, with the 14-year olds having only a slight tendency toward using the vowel spectra for quantity categorization and the 15-year olds being somewhat closer to the adult pattern.

These findings reconfirm the primary role of vowel duration as a perceptual cue for categorizing vowel quantity. They further suggest that the use of spectral information is acquired relatively late, still occurring between 14 and 15 years old, and may be a means of fine-tuning and improving the processing efficiency of vowel identification.

The 9-year olds listeners were included in order to attain a preliminary impression of whether the pattern emerging from the 14-years olds through to the adults extends down to younger listeners. As we continue to run more 9-year old listeners in the study and attain results for a more even distribution across the full range of materials for each word set, we expect that the pattern of results for this group will become more evident and will more clearly demonstrate how they use vowel duration and spectra to identify vowel quantity.

REFERENCES

Behne, D., Czigler, P. and Sullivan, K. 1996. Acoustic characteristics of perceived quantity and quality in Swedish vowels. *Speech Science and Technology '96*, 6, Adelaide, 49-54.

Behne, D., Czigler, P. and Sullivan, K. 1998. Perceived Swedish vowel quantity: Effects of postvocalic voicing. *PHONUM*, 6, 83-89.

Bohn, O.-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. *Speech perception and linguistic experience: issues in cross-language research*, ed. W. Strange, 273-304. Baltimore: York Press.

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

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