ABSTRACT
A distinction in vowel quantity is typically realized acoustically by vowel duration. Research on the perception of Swedish vowel quantity by adult native speakers supports this. It further suggests that when the duration of a vowel is relatively long (due, e.g., to inherent duration), listeners may also make use of vowel spectra to distinguish vowel quantities.

The current project investigates the perceptual cues used to distinguish vowel quantities in late stages of language development. Of particular interest is whether pre-adult listeners use spectral cues to identify the quantity of vowels which have a relatively long inherent duration. Results are compared with the findings for Swedish adults and discussed in terms of the perceptual role of vowel duration and spectra as cues for vowel quantity.

1. INTRODUCTION

1.1 Background
Swedish is traditionally described as having a distinction between long and short vowel quantities [1]. This distinction is realized acoustically through differences in vowel duration, with a long quantity having a duration which 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 [2] 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 [3, 4] have returned to this issue using resynthesized Swedish words with 10 stepwise adjustments of F1 and F2. The results suggest that listeners use vowel duration to identify the quantity of vowels in a continuum from [i] to [l] and from [o] to [O], whereas from [A] to [a] adult listeners use both vowel duration and, to a lesser extent, spectral attributes of the vowel. These findings suggested 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 then make use of both vowel duration and vowel spectra to identify the vowel's quantity.

1.2 Current study

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 aim of this project is 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 /a:/ and /a/ as adult listeners.

2. METHOD

2.1 Materials
The materials used are identical to with those used with adult listeners in [3, 4].

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

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 which 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 “selected productions”.

Synthesis. The Kay Elemetrics LPC Parameter Manipulation/Synthesis program was used to resynthesize three sets of 100 words. Each set was based on the measurements from the selected productions of a pair of long-short vowel quantities: /kI:t/, /kO:t/, and /kA:t/.

For each set, the measurements of the selected productions were used as extreme points of a 10x10 synthesis matrix, having ten degrees of vowel duration and ten degrees of simultaneous first and second formant frequency adjustment. Starting from the selected production of /kI:t/, /kO:t/ and /kA:t/, the vowel duration was adjusted in 10 equal-sized steps toward the measured vowel duration of the selected productions of /kI:t/, /kO/ and /kA/ respectively. Then, for each of the ten vowel durations in each set, F1 and F2 frequencies were simultaneously adjusted in ten steps toward the formant values of the selected productions of /kI:t/, /kO/ and /kA/, respectively. This resulted in three sets of 100 resynthesized items. Since little measured difference was observed between the F3 frequency of long and short vowel quantities in the productions, the frequency of F3 and higher formants was unchanged in the synthesized items, and consequently was the same as the long vowel quantity of the pair. The vowel durations, formant frequencies and step sizes are summarized in Table 1.

<table>
<thead>
<tr>
<th>Word pairs</th>
<th>Spectrum</th>
<th>Vowel duration</th>
<th>/l/ closure duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1 (Hz)</td>
<td>F2 (Hz)</td>
<td></td>
</tr>
<tr>
<td>/kI:t/</td>
<td>262</td>
<td>2254</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-4</td>
<td>-13</td>
</tr>
<tr>
<td>/kI:t/</td>
<td>274</td>
<td>2215</td>
<td>48</td>
</tr>
<tr>
<td>/kO:t/</td>
<td>295</td>
<td>528</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>29</td>
<td>-13</td>
</tr>
<tr>
<td>/kO:t/</td>
<td>378</td>
<td>788</td>
<td>64</td>
</tr>
<tr>
<td>/kA:t/</td>
<td>354</td>
<td>882</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>53</td>
<td>-10</td>
</tr>
<tr>
<td>/kA:t/</td>
<td>747</td>
<td>1362</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>129</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

The duration of a postvocalic consonant is also known to decrease as vowel length increases in Swedish [1] and was observed from the durations of postvocalic /l/ measured in the current study. For each vowel pair, the mean duration of the /l/ closure duration between the representative phonologically long and short items was calculated. The closure duration of all 100 items in each set was then adjusted to this mean. 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 postvocalic consonant durations are shown in the last column of Table 1.

2.2 Identification task

Participants. Two groups of native Swedish listeners participated in the study: 10 14-year olds (4 males and 6 females) and 17 15-year olds (5 males and 12 females). The participants were all living in Umeå, Sweden at the time of the experiment, but had not necessarily always lived there.

Subjects 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, vät - vätt, 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.

Subjects were instructed to use the mouse to click on the word which 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.

Subjects heard 5 randomized repetitions of each synthesized word for a total of 150 items (3 vowel series x 10 items x 5 repetitions). 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 responses and their reaction times for each trial were logged to a data file.

3. RESULTS

Each group’s responses and reaction times were analyzed separately for each of the three sets of synthesized words. Responses for the rhyme word having a phonologically long vowel are referred to as “long” responses in the following discussion. 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 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. Responses and reaction times for the two groups of pre-adult listeners are presented with those of the adults [3]. At each duration step in the figure,
a data point is the result of averaging over the 10 spectral steps and vice-versa.

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, all three age groups show clear use of vowel duration to indentify 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/ materials, 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 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 the /ki:t/-/kɪt/ or /ko:t/-/kɑt/ materials, 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 /kʌt/-/kat/ materials, however not reaching 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 young listeners, perhaps not surprisingly, 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 which were not unlike those of the adults. The reaction time curves increase slightly across the 10 duration steps for all three sets of materials in the same general pattern as is observed for the adult reaction times. Across the 10 spectral steps, their pattern of reaction times are are as unchanging as the adults’, with the possible exception of the /kʌt/-/kat/ materials.

4. DISCUSSION AND 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 previously with adults [2,3,4], 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 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.

Among the three vowel pairs studied [3,4], the use of spectral information by adults is unique to categorizing [ʌ] and [ɑ]. The current study shows a developmental progression, with the 14 year olds having only a slight tendency toward this pattern and the 15 year olds being 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 and may be a means of fine tuning and improving the processing efficiency of vowel identification.

5. REFERENCES


Figure 1. For the three sets of materials (/kiːt/-/kɪht/ /koːt/-/koʊt/ and /kæt/-/kat/), mean percent long responses and mean reaction times are plotted for the 10 synthesized duration steps and spectral steps.