Development of vowel quantity perception in late childhood

Dawn M. Behne\textsuperscript{a}, Peter E. Czigler\textsuperscript{b}, and Kirk P.H. Sullivan\textsuperscript{b}

\textsuperscript{a}Norwegian University of Science and Technology, Trondheim Norway
\textsuperscript{b}Umeå University, Umeå Sweden
dawn.behne@hf.ntnu.no

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 language development by children 9 to 13 years old. Of particular interest is whether these developing 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 the developmental use of vowel duration and spectra as cues for vowel quantity are described.

1. Introduction

1.1. Background

Swedish is traditionally described as having a distinction between long and short vowel quantities\cite{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, F2).

In a classic perception study with Swedish adults, Hadding-Koch and Abramson\cite{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\cite{3, 4} have returned to this issue using resynthesized Swedish words with 10 stepwise adjustments of vowel duration and 10 stepwise adjustments of F1 and F2. Results from a word rhyming task show that adult Swedish listeners use vowel duration to identify the quantity of vowels in a continuum from [i:] to [i] 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), 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 some 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 coming later.

In a recent study\cite{5} we ran the same perception experiment as in\cite{3, 4} with 9, 14, and 15 year old children to trace the developmental use of vowel duration and spectra in vowel quantity identification, and to determine the extent to which the adult task could be used with children. The results showed that 14 and 15 year olds had responses which were similar, but not fully like, the adult results. The 9 year old’s results were much less systematic, suggesting that the 9 year olds either were only beginning to use these acoustic cues in vowel quantity identification, or that the rhyming task was too difficult for them. Since the task required the children to select a printed word which rhymed with the word they heard, and consequently involved reading, we suspect that the task may have edged on being difficult for them.

The aim of this project is to offer an initial investigation of the progressive developmental pattern of how vowel duration and the vowel spectra come to be used to distinguish long and short vowel quantities by children 9 to 13 years old using an age-appropriate task. Of particular interest are whether the young listeners consistently use vowel duration when categorizing vowel quantities and whether the young 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 those used with adult listeners in\cite{3, 4} and with children in\cite{5}.

2.1.1. Recordings

A set of six /kVt/ words were used as targets. Each word was phonotactically possible in Swedish and contained one of the vowels [i, æ, a, i, ɔ, 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.
1.1.2. Measurements

From the 10 recorded productions of each target word, ESPS\textsuperscript{waves+TM} was used to measure the vowel duration, and the first three formant frequencies of the vowel (F1-F3) measured at the center of the most evident steady state. The postvocalic /t/ closure duration 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.

1.1.3. 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: /i:/-\(\alpha:/, [o:]-\alpha:/, and \[a:]-\alpha:/.

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. More specifically, starting from the selected production of /ki:t/, /ko:t/ and /ka:t/, the vowel duration of each word was adjusted in 10 equal-sized steps toward the measured vowel duration of the selected productions of /kt/, /kt/ and /kat/ 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 /kt/, /kt/ and /kat/ respectively. This resulted in three sets of 100 resynthesized items: one set each for /i:/-\(\alpha:/, [i:]-[i:], for [o:]-[\alpha:] and [a:]-[\alpha:].

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 /t/ measured in production in the current study (section 2.1.2). For each vowel pair the duration of the /t/ closures from the selected productions were used to calculate an intermediate duration for /t/. The post-vocalic closure duration of all 100 items in each set was adjusted to this value. The postvocalic consonant durations are shown in Table 1.

1.2. Identification task

1.2.1. Participants.

Eighty native Swedish listeners between 9:0 and 13:0 years participated in the study, and based on their closest birthday, were organized in five groups: 9, 10, 11, 12, and 13 year olds. The participants were all living in Umeå, Sweden at the time of the experiment.

During the experimental task listeners were seated wearing headphones at a computer terminal with a monitor and mouse. For each trial listeners heard a synthesized word and, at the same time, two pictures were presented on the monitor. The two pictures on the monitor represented objects which have names differing from each other in vowel quantity, but containing the same vowel quality and, in most cases also the same postvocalic consonant, as the word heard over the headphones.

Participants were instructed to use the mouse to click on the picture whose name rhymed with the word 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.

Participants heard a semi-randomized subset from the full set of 1500 items used in [3] (3 vowel series x 100 items x 5 repetitions). The semi-randomization allowed for each listener to respond to an equal number of items from each vowel series and five repetitions of an item. This was done as an attempt to keep inter-subject variability at a minimum. Before starting the experiment, participants had three practice trials, and after every 30 trials, they had the opportunity to take a short break. Most children responded to 60 items; others responded to as many as 180 items (\(\mu =80\) items). Listeners’ responses and their reaction times for each trial were logged to a data file.

3. Results

Responses and reaction times from the five age groups were analyzed separately for the three sets of synthesized words. Only the responses will be discussed here. Responses for the picture having a label with a long vowel quantity are referred to as “long” responses in the following discussion. It should be noted that the materials 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 for the five age groups. At each duration step in the figures, a data point is the result of averaging over the 10 spectral steps; similarly, for each spectral step a data point is the result of averaging over the 10 duration steps.

3.1. Variability among children’s responses

A first look at the results in Figure 1 shows a great deal of variability in mean responses evident across age groups for each vowel set.

Although each of the 20 adults in [3] responded to 1500 trials, each child participated in 60 to 180 items from the full set of 1500 items (total = 6360 trials across

| Table 1. Parameter settings of the vowel and postvocalic consonant for the three sets of resynthesized /kVt/ words. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Word pair       | Spectrum F1 (Hz) F2 (Hz) | Vowel duration | /t/ closure duration |
| /ki:t/          |                               |                |                         |
| ↓ Step 1        | 262                           | 2254           | 168                      | 128                      |
| Step size       | 1                             | -4             | -13                      | 0                        |
| /kt/            | Step 10                      | 274             | 2215                      | 48                       | 128                      |
| /ko:t/          | Step 1                       | 295             | 528                       | 182                      | 132                      |
| Step size       | 9                             | 29             | -13                      | 0                        |
| /kat/           | Step 10                      | 378             | 788                       | 64                       | 132                      |
| /ka:t/          | Step 1                       | 354             | 882                       | 160                      | 129                      |
| Step size       | 44                            | 53             | -10                      | 0                        |

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 /t/ measured in production in the current study (section 2.1.2). For each vowel pair the duration of the /t/ closures from the selected productions were used to calculate an intermediate duration for /t/. The post-vocalic closure duration of all 100 items in each set was adjusted to this value. The postvocalic consonant durations are shown in Table 1.
the five age groups). In total the number of trials with the 80 children spread across the five age groups in this study (see Table 2) correspond to the number of trials with only 425 adults in [3]. So the difference in regularity of the children’s data in Figure 1 is not surprising. In addition, by randomly selecting trials for the children from the full set, the number of responses for each item will not be equal until approximately 18 children have participated in each age group. (We continue to collect data). Notably, the need to use more children than adults to attain the same number of trials is very likely to increase the inter-subject variance in the results. These factors all contribute to the greater variability of the mean responses from the children, compared with the adults.

With this in mind, the results presented for the five groups of children in Figure 1 include curves of best fit. These curves represent the general tendency of responses from the five age groups based on results to date.

3.2. Development from 9 to 13 years

In the corresponding study with mature Swedish listeners [3], adults were found to use vowel duration, but not vowel spectra, to identify vowel quantity in the [iː] - [ɪ] and [ɑː] - [ɔː] sets. Adults also clearly used vowel duration to identify quantity in the [ɑː] - [a] series, but also made use of spectral attributes of the vowel.

3.2.1. Vowel duration

The left half of Figure 1 shows plotted raw data and best-fit curves for the three vowel series. Results for the 9 year olds are in the first row, with data-plots from progressively older children in the rows below.

Across the 10 duration steps for the three vowel sets in Figure 1, the relatively flat curves show that vowel duration is not being used to identify the vowel quantities by the nine year olds. Although minor negative slopes are observed for [iː] - [ɪ] and [ɑː] - [ɔː], they are very slight. For [iː] - [ɪ] this tendency has a slight s-curve by 10 years old, and between 10 and 13 years old the two categories of quantity become slightly more distinct. However, based on the data so far, even 13 year olds appear to have not yet fully developed the use of vowel duration for identifying the quantity of [iː] - [ɪ].

Although the 9 year old’s results for the [ɑː] - [a] series show no indication of even starting to use vowel duration for vowel quantity identification, the pattern that develops is similar to the [ɑː] - [ɔː] series. For the [ɑː] - [ɔː] series the minor negative slope in the plot of 9 year olds becomes sharper by 10 years old. For the [ɑː] - [a] series there are also indications of vowel duration being used. By 11 and 12 years old children are using vowel duration slightly more, and by 13 years old, it is clearly being used as an identification cue for vowel quantity in a way similar to the adults. The difference between these two series at 13 years old is similar to how they differ for adults.

3.2.2. First and second formants

Of the vowel pairs investigated the adults in [3] only used the simultaneous adjustments of F1 and F2 to identify vowel quantity in the [ɑː] - [a] series. F1 and F2 results for the children are shown on the right side of Figure 1.

As was observed for the adults, children show no indications of using F1 and F2 to identify vowel quantity in either the [iː] - [ɪ] or [ɑː] - [ɔː] series, evident from the comparatively flat curves across the 10 spectral steps for these two series. Although there are some slight variations, this general pattern is observed for children 9 through 13 years old.

Like the adults, the children also used F1 and F2 to identify vowel quantity in the [ɑː] - [a] series. Initial signs of the adult pattern occur already at 9 and 10 years old. By 11 and 12 years old the s-curve in the lower half of the plot is approximates the adult results, and is maintained through 13 years old.

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. Adult responses show vowel duration is 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.

The current study shows a developmental progression, with 9 year olds showing few indications of being able to identify vowel quantity using vowel duration and/or F1-F2. But from 10 to 13 years old, they develop patterns approximating those observed for adults. Results form 14-15 year old children [6] show that use of these perceptual cues continues to be refined for a couple more years, but is likely complete by around 15 years old.

These findings suggest that the developmental course for using vowel duration and F1-F2 for vowel quantity identification likely occurs in parallel from about 10 years old. However, it may be that, after vowel duration is already in full use for vowel quantity identification, the use of spectral cues continues being fine-tuned and improving processing efficiency of vowel identification.

5. Acknowledgements

We are grateful to Takayuki Arai for his valuable input, in particular related to curve analyses.

6. References


<table>
<thead>
<tr>
<th>Vowel duration</th>
<th>F1 and F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i:] - [i]</td>
<td>[i:] - [i]</td>
</tr>
<tr>
<td>[o:] - [o]</td>
<td>[o:] - [o]</td>
</tr>
<tr>
<td>[α:] - [α]</td>
<td>[α:] - [α]</td>
</tr>
</tbody>
</table>

![Graphs](image)

Figure 1. For the three sets of materials ([i:] - [i], [o:] - [o], [α:] - [α]), mean percent long responses are plotted (---) for the 10 synthesized duration steps (left three columns) and spectral steps (right three columns) for each of the five age groups (from the top to bottom row): 9 year olds, 10 year olds, 11 year olds, 12 year olds, and 13 year olds. Best-fit curves are plotted as solid lines.