Vowel Category Perception Affected by Microdurational Variations

Einar Meister¹, Stefan Werner²

¹Laboratory of Phonetics and Speech Technology,
Institute of Cybernetics, Tallinn University of Technology, Estonia
²Department of General Linguistics and
Language Technology, University of Joensuu, Finland

ein@ioc.ee, stefan.werner@joensuu.fi

Abstract

Vowel quality perception in quantity languages is considered to be unrelated to vowel duration since duration is used to realize quantity oppositions. To test the role of microdurational variations in vowel category perception in Estonian listening experiments with synthetic stimuli were carried out, involving five vowel pairs along the close-open axis.

The results show that in the case of high-mid vowel pairs vowel openness correlates positively with stimulus duration; in mid-low vowel pairs no such correlation was found. The discrepancy in the results is explained by the hypothesis that in case of shorter perceptual distances (high-mid area of vowel space) intrinsic duration plays the role of a secondary feature to enhance perceptual contrast between vowels, whereas in case of mid-low oppositions perceptual distance is large enough to guarantee the necessary perceptual contrast by spectral features alone and vowel intrinsic duration as an additional cue is not needed.

Index Terms: vowel perception, intrinsic duration, category boundary

1. Introduction

In quantity languages like Estonian and Finnish vowel quality perception is expected to be unrelated to vowel duration since this parameter is exploited in the realization of phonemic duration oppositions. In Finnish most sounds can occur in a short and a long quantity, in Estonian the quantity system is more complex: on the phonemic level a short vs. long opposition exists, on the foot level a three-way quantity contrast is possible [1], traditionally referred to as short (Q1), long (Q2) and overlong (Q3) quantity degrees. The acoustic measurements of Estonian vowel formants in CV/V/CV context show that quality differences of the stressed vowels in Q1, Q2 and Q3 are small and do not exceed 1 Bark in F1-F2 axis.

Experimental studies of microprosodic/intrinsic features of several non-quantity languages (see e.g. [3], [4], [5]) have shown that open vowels tend to have lower F0, higher intensity and longer duration than close vowels. Similar phenomena have been found also in the case of quantity languages, at least for Estonian and Finnish [6]. The role of intrinsic features in speech perception has not been studied thoroughly since according to the traditional approach suprasegmental (higher order) features “override” them in natural speech.

One of the few studies addressing the perceptual role of intrinsic duration of vowels [7] has shown that vowel duration is consciously controlled by Russian speakers to increase phonetic contrast in cases where other parameters, for example spectral parameters, lose their distinctive power. The acoustic measurements [8] show that the frequencies of the first two formants of Russian [i] and [e] are located in partly overlapping areas; the results of perceptual experiments also demonstrate auditory closeness of Russian [i] and [e]. The results of the experiments with synthetic vowels in CVC context demonstrate that in the case of ambiguous vowel quality with formant frequencies in the overlapping area of vowels [i] and [e], the frequency of stimulus perception as [i] is inversely proportional to vowel duration [7].

Intrinsic duration plays a role only in close-open (F1) dimension, as demonstrates the study on distinguishing vowel categories along front-back (F2) dimension in Finnish [9]. Thus, we should really look at the relations between vowel duration and category perception in the F1 domain only, as there is no evidence of the role of duration in F2 domain.

In quantity languages duration has a different function than in non-quantity languages – it acts as the main feature of quantity oppositions and not as e.g. a major prominence-inducing factor. This function of duration leads one to anticipate no relationship between vowel duration and quality perception or at most a much weaker relationship than in non-quantity languages. Nevertheless, the fact that intrinsic duration of vowels can be observed in both quantity and non-quantity languages, has lead us to the hypothesis that the contribution of intrinsic duration to vowel category perception could be universal.

In our pilot study [10] we investigated both Estonian and Finnish exemplars of the /i/-/e/ vowel pair; the results showed a clear effect of segment duration on vowel perception: the longer the duration of the stimulus, the more often it was identified as /e/. In a further experiment involving five vowel pairs along the close-open axis /i/-/e/, /y/-/æ/, /u/-/o/ and /o/-/æ/, four Estonian listeners showed considerable between-subject variation; nevertheless, the overall tendency – perceived vowel openness correlates positively with stimulus duration – was obvious [11].

The present study is intended to confirm the outcome of our earlier experiments with a larger number of subjects and to provide more reliable evidence for the role of microdurational variations in vowel category perception by Estonian subjects.

2. Experimental setup

The experimental setup involved a two-stage procedure: first, the perceptual category boundary of the two vowels in each pair had to be found; second, for each subject three stimuli at subject-specific category boundaries representing the most ambiguous formant structures were chosen for duration manipulation and thereafter presented to listeners.

For the experiments five vowel pairs representing quality oppositions in the close-open dimension were used: three
high-mid vowel pairs (/i/-/ɛ/, /i/-/ø/, /u/-/ø/) and two mid-low vowel pairs (/ɛ/-/æ/, /o/-/a/).

Ten native Estonian adults (5 male and 5 female) participated in both experiments; none of the subjects were phonetically trained or reported any hearing problems.

2.1. Experiment 1

Experiment 1 is designed to discover individual category boundaries of subjects in each close-open vowel pair, as intrinsic vowel duration is expected to play a role in vowel categorization only with stimuli of ambiguous quality. According to our preliminary results the locations of vowel category boundaries vary among subjects to a certain extent and applying averaged boundaries for all subject will lead to poorer discrimination results in the second experiment. Thus, adjusting the stimuli to subject-specific boundaries is necessary.

2.1.1. Stimulus corpus

The stimulus corpus was created by interpolating stepwise between prototypical values of the first three formants for each vowel pair and calculating evenly spaced steps through the F1/F2/F3 continuum. The respective seventeen to nineteen gradually changing stimuli were synthesized with KlattWorks [12], an implementation of a Klatt-type formant synthesizer [13]. The formant values of prototype vowels (see Table 2) are approximations of stressed-syllable vowel formants taken from an evaluation of the Estonian BABEL database [2]. All stimuli used in the first experiment were synthesized with constant F0 = 100 Hz and a duration of 160 ms.

2.1.2. Procedure

All testing was conducted in a sound-isolated room and stimuli were presented to subjects via high-quality headphones. The test was administered with Praat’s [14] multiple forced-choice test facility; each vowel was repeated three times in random order with no replay option. In the test listeners had to decide on vowel quality in a binary identification task with non-primed single stimuli. In total 264 stimuli (17-19 stimuli x 3 repetitions x 5 vowel pairs) were presented to subjects. The duration of the test turned out to be 20-25 minutes.

2.1.3. Results

Table 1 represents the stimulus numbers corresponding to individual category boundaries in five vowel pairs for the ten Estonian subjects S1..S10 discovered in the first experiment.

Table 1. Stimulus numbers corresponding to individual vowel category boundaries of Estonian subjects (S1...S10).

<table>
<thead>
<tr>
<th>Vowel pair</th>
<th>Individual boundary location (stimulus order number)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>/l/-/ɛl</td>
<td>9</td>
</tr>
<tr>
<td>/l/-/æl</td>
<td>9</td>
</tr>
<tr>
<td>/l/-/u/</td>
<td>9</td>
</tr>
<tr>
<td>/l/-/o/</td>
<td>9</td>
</tr>
<tr>
<td>/l/-/a/</td>
<td>9</td>
</tr>
</tbody>
</table>

The category boundary area for all subjects extends to 4 stimulus steps in the case of /l/-/ɛ/ (from 8th to 11th stimulus) and /l/-/æ/ (from 6th to 9th stimulus) series, and to 5 steps (from 8th to 12th stimulus) for the /l/-/a/ continuum. At the same time the individual subject-specific boundary areas are narrower (mainly 1-2 stimulus steps) in the case of all vowel continua.

The individual category boundaries served as the basis for the preparation of subject-specific stimulus sets for the second experiment.

2.2. Experiment 2

Experiment 2 was designed to answer the main question of the study – do microdurational variations affect the perception of vowel quality in a quantity language?

2.2.1. Stimulus corpus

For each subject three stimuli at subject-specific category boundaries representing the most ambiguous formant structures were chosen for duration manipulation (see Table 2 and Figure 1 for the values of formant frequencies). The durations of stimuli varied from 60 to 140 ms in 20 ms steps, fundamental frequency was kept constant at 100 Hz. ABX and BAX series with an inter-stimulus interval of one second were constructed, where A and B were the prototype vowels of a vowel pair and X a stimulus of ambiguous quality. The duration of prototypes A and B was also varied from 60 to 140 ms in 20 ms steps, according to the duration of X.

Table 2. Formant frequencies (in Hz) of vowel prototypes and ambiguous stimuli.

<table>
<thead>
<tr>
<th>Prototypes</th>
<th>Ambiguous stimuli</th>
<th>Prototypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>/l/-/ɛl</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>F1</td>
<td>250</td>
<td>330</td>
</tr>
<tr>
<td>F2</td>
<td>2220</td>
<td>2076</td>
</tr>
<tr>
<td>F3</td>
<td>3000</td>
<td>2776</td>
</tr>
<tr>
<td>/l/-/æl</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>F1</td>
<td>260</td>
<td>316</td>
</tr>
<tr>
<td>F2</td>
<td>1750</td>
<td>1875</td>
</tr>
<tr>
<td>F3</td>
<td>2160</td>
<td>2194</td>
</tr>
<tr>
<td>/l/-/u/</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>F1</td>
<td>400</td>
<td>522</td>
</tr>
<tr>
<td>F2</td>
<td>1950</td>
<td>1782</td>
</tr>
<tr>
<td>F3</td>
<td>2580</td>
<td>2495</td>
</tr>
<tr>
<td>/l/-/o/</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>F1</td>
<td>450</td>
<td>538</td>
</tr>
<tr>
<td>F2</td>
<td>800</td>
<td>888</td>
</tr>
<tr>
<td>F3</td>
<td>2400</td>
<td>2426</td>
</tr>
</tbody>
</table>

The category boundary area for all subjects extends to 4 stimulus steps in experiment 2.
2.2.2. Procedure

Testing conditions and subjects were the same as in the first experiment. The test was administered with Praat’s multiple forced-choice test facility using balanced permutation for stimulus randomization. Each stimulus was repeated three times in both ABX and BAX contexts, adding up to a total of 450 stimuli (5 vowel pairs x 3 formant structures x 5 durations x 6 repetitions). Subjects had to answer (by clicking in one of two response boxes on the screen) the question “Does the sound you heard last resemble more the first or the second vowel?” A replay option was available during the listening test which allowed up to five repetitions of a stimulus.

The ABX/BAX setup was chosen to compensate for the potential influence of the preceding stimulus which occurred in a pilot test with single stimulus ordering type (see e.g. [15] for a discussion of stimulus order effects in vowel perception). The experiment consisted of 5 blocks (each vowel pair in a separate block) with optional short breaks between blocks. On average, the test took about 45 minutes.

2.2.3. Results

Perception scores of high vs. low vowel in a pair for each formant setting and five duration steps were computed for each subject; the correlation between stimulus duration and perception scores was evaluated using Pearson’s product-moment coefficient.

The results show that there is a clear correlation between response and vowel duration in the expected direction at all high-mid category boundaries – the longer the stimulus the lower the score of high-vowel perceptions. The subjects showed the best results in the case of the /ɪ/-/ɛ/ vowel pair: all 10 subjects had significant correlation at least for one for mant setting; the results were slightly worse in the case of other high-mid vowel pairs. For mid-low category boundaries the results are different – stimulus duration does not affect vowel category perception significantly.

The values of Pearson’s product-moment coefficient are the following: for /ɨ/-/ɛ/ and /ɻ/-/ɛ/ it was -0.98-0.99 (p < 0.01), for /ɛ/-/æ/: -0.46 (p = 0.43), and for /ɻ/-/ɛ/: 0.07 (p = 0.91). Linear regression plots for perception scores of high-mid pairs are shown in Figure 2 (R² = 0.982 for /ɻ/-/ɛ/). R² = 0.978 for /ɻ/-/ɛ/ and R² = 0.984 for /ɻ/-/ɛ/ and for mid-low vowel pairs on Figure 3 (R² = 0.21 for /ɛ/-/æ/ and R² = 0.006 for /ɻ/-/ɛ/).

3. Discussion

The results have shown that duration of a vowel affects vowel category perception when spectral information is ambiguous, even in a quantity language. According to our tentative hypothesis this influence is based on differences between the intrinsic durations of high, mid and low vowels which are physiologically-based and thus outside the scope of speaker control. This hypothesis seems natural in the case of Estonian where duration is intentionally controlled by a speaker on higher (phonemic and syllabic/foot) levels in order to produce contrastive quantity oppositions. However, different results for high-mid versus mid-low vowel pairs cannot be explained by the hypothesis (assuming that the impact of intrinsic duration remains constant along the open-close dimension).

Another possible explanation is that intrinsic duration of Estonian vowels, attributed to physiological properties of the human vocal tract (see e.g. the tongue-pull hypothesis [21]), is actually consciously controlled by the speaker. This claim is supported by findings of studies on non-quantity languages like Russian [7] and Spain [16] where intrinsic features of speech are found to be under control of a speaker to increase phonetic contrast between different phonological categories.

Yet, we would like to keep our initial hypothesis based on the physiological grounding of intrinsic features and to provide a different explanation for our results. According to theories modeling sound system evolution in different languages [17], [18], vowels tend to maximize perceptual distances in the perceptual plane F1 vs. F2 (F2 is the so-called “perceptual second formant” integrating F2, F3 and F4). Depending on the number of vowels in a language, secondary articulatory contrasts (e.g. nasality, length, pharyngealisation) may be combined with (primary) quality contrasts in order to achieve the goal of maximal perceptual contrast [18].

The Estonian vowel system includes 9 vowel phonemes, divided according to tongue height as: /ɪ/, /ɻ/, /ɻ/ high, /ɛ/, /ɔ/, /ɑ/, /ɩ/ mid, and /æ/, /ʌ/ low vowels (the mid-vowel /ʌ/ is actually a non-low vowel as it extends to the high vowel area, too) [20]. Using the formant values F1 and F2 of the Estonian vowel prototypes from [19] for calculating Euclidean distances as a measure of perceptual distance between vowel types, we can see that perceptual distance between mid-low vowels is larger than between high-mid vowels (Table 4). In the case of front vowels the perceptual distance between high-mid and mid-low pairs is 1.6-2.1 Bark (/ɻ/-/ɑ/ and /ɪ/-/ɛ/) vs.
4.3 Bark (/u/-/o/), in back vowels correspondingly 1.7 Bark (/a/-/a/) vs. 2.8 Bark (/o/-/a/).

We can hypothesize that in the case of shorter perceptual distances (high-mid area of vowel space) intrinsic duration plays the role of a secondary feature to enhance perceptual contrast between vowels. In case of mid-low oppositions the perceptual distance is large enough, thus the necessary perceptual contrast is guaranteed by spectral features alone and an additional duration cue is not needed.

<table>
<thead>
<tr>
<th>Vowel pair</th>
<th>Perceptual distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/-/e/</td>
<td>2.1</td>
</tr>
<tr>
<td>/y/-/æ/</td>
<td>1.6</td>
</tr>
<tr>
<td>/a/-/o/</td>
<td>1.4</td>
</tr>
<tr>
<td>/e/-/æ/</td>
<td>4.3</td>
</tr>
<tr>
<td>/o/-/a/</td>
<td>2.9</td>
</tr>
</tbody>
</table>

4. Conclusions

Our experiments confirm that microdurational variations do affect the perception of vowel quality also in a language with phonemic quantity oppositions. Nevertheless, different results for high-mid and mid-low vowel pairs cannot be explained solely by the different intrinsic durations of high, mid and open vowels. Thus, in addition, the perceptual distance of vowels in the close-open dimension should be taken into account.

5. Future work

The perception scores (especially at mid-low category boundaries) show considerable variation among subjects. Also, the context of stimulus presentation (ABX vs BAX) had a certain impact on vowel categorization in all vowel pairs. These variations deserve further attention which may shed some light on the influence of other possible factors on vowel quality perception.

On-going perception experiments with Finnish subjects will give us additional evidence on the role of intrinsic duration on vowel category perception in quantity languages.

6. Acknowledgements

This research was partially supported by the target-financed theme No.032270506 of the Estonian Ministry of Education and Research and by the National Programme for Estonian Language Technology. Authors' cooperation has been supported both by the Academy of Finland and the Estonian Academy of Sciences.

7. References