DIFFERENCES IN FINNISH FRONT VOWEL PRODUCTION AND WEIGHTED PERCEPTUAL PROTOTYPES IN THE F1-F2 SPACE

Osmo Eerola[a,b] *& Janne Savela*[c]

[a]Department of Biomedical Engineering, Tampere University of Technology, Tampere, Finland;
[b]Centre for Cognitive Neuroscience, University of Turku, Turku, Finland;
[c]Department of Information Technology, University of Turku, Turku, Finland
osmo.eerola@tut.fi; jansav@utu.fi

ABSTRACT

The perception and production of the mid-front Finnish vowels /i/, /e/, /y/, and /ø/ were investigated in fourteen Finnish-speaking subjects. In the perception experiment, synthesized long vowels were used as stimuli in order to identify category prototypes. For production, the subjects were asked to pronounce words including these vowels as short and long variants. This study introduces a new concept of *weighted perceptual prototype*, which is compared with the estimated absolute prototypes obtained in the perception experiment. The calculated mean Euclidean distance in the F1-F2 space between the produced vowels and their weighted category prototypes was 111 mel for short and 116 mel for long vowels. At an individual level, the F1 and F2 values of the weighted perceptual prototypes correlated significantly with the F1 and F2 values of the produced short and long vowels. Statistically significant differences were found between the mean values of the weighted category prototypes and estimated absolute prototypes for /i/, /e/, and /ø/ but not for /y/.

Keywords: vowel perception and production, weighted prototypes

1. INTRODUCTION

The existence of internal structures of phonetic categories and prototypical category representatives has been shown in many reports [10-16]. The phoneme prototype (P) is traditionally defined as the best representative of a phoneme category, and experimentally determined as the highest ranking category member in goodness evaluation tests [8].

Irrespective of the experimental approach, the measured prototype represents an estimate of the absolute or 'true' category prototype, marked here as Pa_{est}. The goodness of the estimate depends on the number of stimuli used in the grid to cover the investigated vowel space; decreasing the step size of the synthesis parameters will rapidly increase the number of stimuli unpractically large for use in listening experiments. To overcome this problem, novel optimizing methods have been presented [2, 6]. The weighted prototype (Pω) approach enables us to avoid some of these experimental problems. The Pω is robust in the sense that it represents the center of gravity of the category: the absolute prototype can most likely be found within the area of the vowel space where the majority of the stimuli with high goodness values lie.

Phoneme prototypes are the natural candidates for the 'auditory targets', which many models assume to be the elementary neural representations used in the template matching of speech perception, and for control references in speech production [3, 4].

The Finnish vowel system includes eight vowels: /a/, /e/, /i/, /o/, /u/, /y/, /æ/, and /ø/, which all can occur as short (single) or long (double) in any position of a word. This study concerns the perception and production of the Finnish mid-front vowels /i/, /y/, /e/, and /ø/ and consists of two experiments: a combined vowel identification and rating experiment, and a subsequent vowel production experiment. The purpose of this study was to test the hypothesis that the acoustic features (as implemented in F1 and F2) of an individual's perceptual vowel prototype correlate with the same acoustic features of the produced vowel, and to compare the Euclidean distances of perceived and produced vowels in the F1-F2 space. Additionally, since Finnish is an example of an extreme quantity language, the effect of vowel duration on the articulated vowel quality was investigated as well: it was assumed that the long vowels better achieve the articulatory targets (prototypes), in other words, they have smaller distances from the prototypes than the short vowels have.

2. EXPERIMENT 1: PERCEPTION

2.1. Subjects

Fourteen (14) normally hearing young adults aged 17-31 and speaking the modern educated Finnish of South-West Finland volunteered as subjects (7 male, 7 female) in both experiments. All subjects were screened for hearing impairments by means of an audiometer (Amplivox 116).

2.2. Stimuli and procedure

Forty-six (46) vowel variants were synthesized using the Klatt serial mode speech synthesizer [7] to represent the long Finnish /e:/, /i:/, /y:/, and /ø:/ vowels with a duration of 250 ms. On the basis of the earlier reported typical formant values of the relevant vowels occurring in Finnish words [1], a tentative category center was determined for each vowel category (Table 1, upper part).

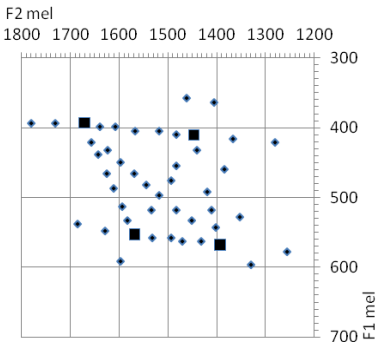
Forty-two (42) vowel variants were then synthesized around the category centers to form the grid of stimuli shown in Figure 1. The F1 and F2 of these stimuli varied in approximately similar steps of 30 mel in the psychoacoustic mel scale (Table 1, lower part). The other formants were fixed: F3 was 2400 Hz for /y:/, 2460 Hz for /ø:/, 2800 Hz for /e:/, and 2980 Hz for /i:/, and F4 was 3200 Hz for /y:/, 3300 Hz for /ø:/, 3800 Hz for /e:/, and 4000 Hz for /i:/.

The f0 contour rose from 112 Hz to 122 Hz during the first 50 ms and then decreased to 102 Hz until the end of the 250 ms stimulus. A linear window of 10 ms was used at the beginning and end of the stimulus in order to prevent audible clicks. The stimuli were presented in an acoustically dampened room (27 dB_A) via Sennheiser PC161 headphones that were calibrated for each session by Brüel & Kjaer Type 2235 SPL meter to deliver 83 +/- 0.5 dB_A.

Table 1: The F1 and F2 values (in Hz and mel) of the tentative category centers, and the range of F1 and F2 variation (in Hz) of the grid of synthesized stimuli presented in Figure 1.

Vowel	F1 Hz	F1 mel	F2 Hz	F2 mel
/e:/	435	553	2170	1568
/i:/	285	393	2460	1671
/y:/	300	410	1865	1447
/ø:/	450	568	1740	1393
Range	F1 min	F1 max	F2 min	F2 max
/e:/	370	475	1980	2500
/i:/	285	335	2170	2800
/y:/	255	340	1500	2040
/ø:/	375	480	1450	1920

Figure 1: The grid of synthesized long vowels in Experiment 1. The category centers (from left to right, and from top to down; /i:/, /y:/, /e:/, and /ø:/) determined on the basis of literature are marked with large squares. The horizontal F2 and vertical F1 axes are in mels.



The EMFC tool of the Praat software was used for stimulus delivery and data collection. The stimuli were presented in 10 blocks of 46 stimuli, each variant occurring 10 times in a random order. After each block, the subject was allowed to take a short break. The test started with a training block consisting of 30 vowels.

In the perception experiment, the subjects were first asked to identify the vowels as belonging to one of the four categories /e:/, /i:/, /y:/, or /ø:/, and then to rate the goodness of each vowel stimulus. A rating scale of 1-7 was employed. The highest score (7) represented a natural sounding, good exemplar of the relevant vowel category, whereas the lowest score (1) represented a poor exemplar. If the subject was not able to categorize the stimulus into the given categories, then the subject was instructed to select the null goodness score (0).

2.3. Analysis and results

For each subject, the identifications of the 46 stimulus variants were counted. This resulted in a categorization rate (%) for each stimulus. For those stimuli that were classified as belonging to one and the same category at a rate of ≥70%, a mean goodness score value was calculated based on the ratings on the scale 1-7. The highest scoring stimulus token in each category signifies an estimate of the absolute prototype $Pa_{est}(F1,F2)$.

The weighted prototype $P\omega(\mathbf{F1,F2})$ of each category was formed by using the equation

(1) $\mathbf{Fi}=(a_1r_1Fi_1+ a_jr_jFi_j +...+a_nr_nFi_n) /(a_1r_1+a_jr_j +...+ a_nr_n)$

where **Fi** = weighted formant frequency, i=1,2,
Fi_j = formant i of stimulus j, j=1,2, ...,n,
a_j = evaluation mean score (1-7), j=1,2, ...,n,
r_j = identification consistency (0.7-1.0), j=1,2, ...,n,
n=number of stimuli identified as category members

Pω(F1,F2) thus represents a point in the F1-F2 space (mel) that is obtained by weighting the F1 and F2 mel values of each stimulus identified as a category member (≥70%) by the goodness rating value. The mean values and standard deviations of the F1 and F2 frequencies (mel) of the estimated absolute category prototypes (Pa_{est}) and the mean values of the weighted prototypes (Pω) of the 14 listeners are presented in Table 2.

Table 2: The mean F1 and F2 values (mel) of perceived /e:/, /i:/, /y:/, and /øː/ vowels given as the estimated absolute prototypes (Pa_{est}) and weighted prototypes (Pω). Standard deviations are in the parentheses.

Vowel	Pa_{est} F1	Pa_{est} F2	Pω F1	Pω F2
/e:/	558 (24)	1639 (23)	541 (17)	1628 (15)
/i:/	392 (12)	1733 (60)	401 (5)	1701 (23)
/y:/	388 (33)	1483 (48)	402 (12)	1462 (11)
/øː/	569 (20)	1375 (53)	544 (14)	1412 (33)

The mean differences between the two methods of obtaining the prototype are 9-25 mel for F1 and 11-36 mel for F2. The Wilcoxon signed ranks test showed that there were significant differences between the estimated absolute and the center-of-gravity type (i.e. weighted) prototypes ($p < 0.05$) in the categories /e:/, /i:/, and /øː/, but not in /y:/.

3. EXPERIMENT 2: PRODUCTION

3.1. Procedure

In the production experiment, the articulation of the utterances [tili], [ti:li], [teli], [te:li] [tyli], [ty:li], and [tøli], [tøːli] (Finnish words and non-words) was recorded from the subjects of Experiment 1. They were asked to utter each word five times successively using their normal speech style. The recording was carried out in an acoustically dampened room by using a high quality AKG D660S microphone that was connected via an amplifier to a PC. The recordings were made at a sampling rate of 44.1 kHz, and saved as sound files for later analysis. Praat SW was used both for the recordings and analysis.

3.2. Analysis and results

The sound samples were automatically analyzed using a text grid in which the steady state part of each target vowel was windowed varying between utterances. Five vowel formants (F1-F5) were analyzed by using the Burg method in which short-term LPC coefficients are averaged for the length of an entire sound. The Praat formant analysis settings were 0.025 s for the Window length, and 5000 Hz (male) and 5500 Hz (female) for the Maximum formant.

The mean values and standard deviations of the F1 and F2 frequencies (mel) of the produced short and long /e/, /i/, /y/, and /ø/ vowels of the 14 listeners are presented in Table 3. ANOVA showed no effect of the vowel quantity on the F1 or F2 values across the four vowel categories. The Euclidean distances in the F1-F2 plane between the short and long vowels produced by the 14 subjects were 29 (SD 16) mel for /e/, 49 (SD 24) mel for /i/, 51 (SD 44) mel for /y/, and 42 (SD 31) mel for /ø/. These distances are of the order of the combined F1 and F2 difference limens (DL) reported in the literature [5, 9], indicating that the quality differences of short and long Finnish mid-front vowels spoken in citation form words are hardly noticeable.

Table 3: The mean F1 and F2 values (mel) of produced short and long /e/, /i/, /y/, and /ø/ vowels. Standard deviations are in the parentheses. dE Pa_{est} is the Euclidean distance in the F1-F2 plane between the produced vowels and estimated absolute prototypes, and dE Pω is the Euclidean distance between the produced vowels and weighted prototypes.

Vowel	F1	F2	dE Pa_{est}	dE Pω
/e/	601 (43)	1560 (113)	131 (51)	141 (53)
/i/	461 (43)	1658 (125)	176 (64)	143 (51)
/y/	445 (34)	1445 (60)	106 (49)	80 (16)
/ø/	580 (46)	1431 (62)	95 (47)	86 (46)
/e:/	602 (47)	1583 (115)	123 (44)	138 (48)
/i:/	441 (38)	1693 (133)	162 (54)	135 (52)
/y:/	436 (37)	1442 (90)	112 (55)	93 (41)
/øː/	588 (52)	1416 (75)	99 (45)	97 (40)

4. RESULTS AND DISCUSSION

The average perceptual Euclidean distance between the Finnish /e:/, /i:/, /y:/, and /øː/ categories was 218 mel (SD 15, N=14), when calculated as the mean distances between the weighted prototypes. Correspondingly, the average distances between the category centers of produced short /e/, /i/, /y/, and /ø/ vowels were 204 mel (SD 68, N=14), and of produced long /e:/, /i:/, /y:/, and /øː/ vowels 205 mel (SD 37, N=14).

The differences between individual weighted prototypes and articulated short and long vowels are presented in Figure 2. The lengths and directions of the vectors indicate that, on the average, the individual production (vector arrow) is more central and/or lower than the relevant perceptual target (vector start point). The Euclidean distances in the F1-F2 plane between the produced and perceived short and long vowels of

the 14 subjects are shown in Table 3. The mean dE Pa_{est} is 127 mel (SD 36) for short vowels and 125 mel (SD 29) for long vowels, and the mean dE Pω is 113 mel (SD 34) for short vowels and 116 mel (SD 24) for long vowels. At the group level (N=14), the produced vowels were always closest to the weighted prototypes of the category in question (Table 4).

Figure 2: Individual Euclidean distances (dE Pω) for each vowel category plotted in the F1-F2 space (mel). The upper panel represents the short and lower panel the long vowels.

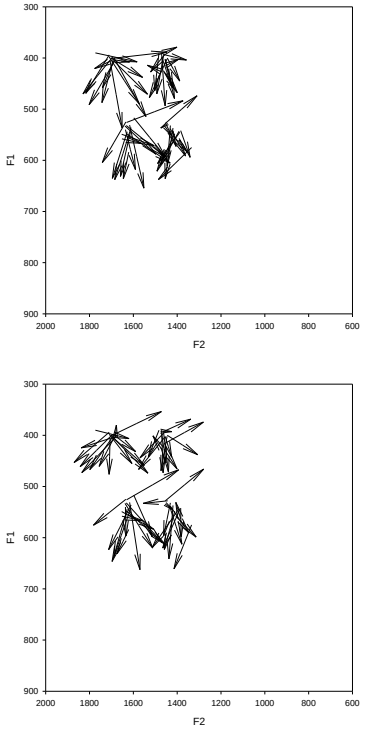


Table 4: The mean Euclidean distances (mel) of the produced short and long /e/, /i/, /y/, and /ø/ vowels from the weighted category prototypes. Standard deviations are in the parentheses.

Vowel	Pω /e/	Pω /i/	Pω /y/	Pω /ø/
/e/	**141 (53)**	269 (46)	242 (71)	178 (106)
/i/	146 (65)	**143 (51)**	213 (119)	273 (120)
/y/	205 (64)	262 (63)	**80 (16)**	124(54)
/ø/	216 (61)	329 (44)	192 (34)	**86 (46)**
/e:/	**138 (48)**	261 (47)	253 (77)	199 (109)
/i:/	213 (43)	**135 (52)**	240 (127)	311 (129)
/y:/	219 (91)	264 (88)	**93 (41)**	141 (57)
/øː/	223 (74)	347 (61)	208 (34)	**97 (40)**

The relationship between the weighted prototypes and produced vowels was tested by using Pearson correlation. The F1 and F2 values of the weighted individual perceptual prototypes correlated significantly with the F1 and F2 values of the articulated short and long vowels: between Pω and short vowels for F1 ($r=0.860$; $p<0.01$; $df=55$) and for F2 ($r=0.666$; $p<0.01$; $df=55$), and between Pω and long vowels for F1 ($r=0.882$; $p<0.01$; $df=55$) and F2 ($r=0.708$; $p<0.01$; $df=55$).

5. REFERENCES

[1] Aaltonen, O., Suonpää, J. 1983. Computerized two-dimensional model for finnish vowel identifications. *Audiology* 22, 410-415.

[2] Benders, T., Boersma, P. 2009. Comparing methods to find a best exemplar in a multidimensional space. *Proc.10th Ann. Conf. of Int. Speech Com. Ass.*, 396-399.

[3] de Boer, B. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28, 441-465.

[4] Guenther, F. 2006. Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders* 39, 350-365.

[5] Hawks, J. 1994. Difference limens for formant patterns of vowel sounds. *J. Acoust. Soc. Am*. 95, 1074-1084.

[6] Iverson, P., Evans, B. 2003. A goodness optimization method for investigating phonetic categorization. *Proc. 15th ICPhS* Barcelona, 2217-2220.

[7] Klatt, D. 1980. Software for Cascade/Parallel Formant Synthesizer. *J. Acoust. Soc. Am.* 53, 8-16.

[8] Kuhl, P. 1991. Human adults and human infants show a "perceptual magnet effect" for prototypes of speech categories, monkeys do not. *P&P* 50, 93-107.

[9] Mermelstein, P. 1978. Difference limens for formant frequencies of steady-state and consonant-bound vowels. *J. Acoust. Soc. Am*. 63, 572-580.

[10] Miller, J. 1997. Internal structure of phonetic categories. *Language and Cognitive Processes* 12, 865-869.

[11] Miller, J., Connine, C., Schermer, T., Kluender, K. 1983. A possible auditory basis for internal structure of phonetic categories. *J. Acoust. Soc. Am.* 73, 2124-2133.

[12] Nåbelek, A., Czyzewski, Z., Crowley, H. 1993. Vowel boundaries for steady-state and linear formant trajectories. *J. Acoust. Soc. Am.* 94, 675-687.

[13] Nearey, T. 1989. Static, dynamic, and relational properties in vowel perception. *J. Acoust. Soc. Am*. 85, 2088-2113.

[14] Repp, B., Crowder, R. 1990. Stimulus order effects in vowel discrimination. *J. Acoust. Soc. Am.* 88, 2080-2090.

[15] Rosch, E. 1975. Cognitive reference points. *Cognitive Psychology* 7, 532-547.

[16] Strange. W. 1989. Evolving theories of vowel perception. *J. Acoust. Soc. Am.* 85, 2081-2087.