

# On the expected variability of vowel quality in Finnish informal dialogue

Mietta Lennes

Department of Phonetics, University of Helsinki, Finland  
mietta.lennes@helsinki.fi

## Abstract

Eight vowel qualities are phonologically distinct in Finnish. All of them may occur as either long or short, or they may combine into diphthongs. However, the actual variability of vowel qualities in conversational Finnish speech is unknown. One of the factors affecting this variation is probably the predictability of word forms. The influence of simple word form frequency was addressed in this preliminary study. Five informal Finnish dialogues were recorded and transliterated, and the frequencies of different word forms were obtained from the material. For four speakers, the formant values of F1 and F2 were calculated at the midpoints of vowel segments. Different F1/F2 charts were plotted for vowels according to four word form frequency categories. The results indicate that vowel segments within tokens of common word forms are phonetically more variable than vowels within tokens of rare word forms.

## 1 INTRODUCTION

Casual conversational speech exhibits great phonetic variability, which is a difficult problem for, e.g., the development of speech technological applications. However, this variation is not random, and it may be assumed that people in fact utilize it to support their communication — both speaking and speech comprehension.

The Finnish language uses eight different vowel qualities /a e i o u y æ ø/ for marking phonological contrasts. All of these vowels may occur phonologically as either long or short, or they may combine into diphthongs. However, there are phonotactic restrictions on the occurrence of different vowels (e.g., vowel harmony), and not all vowel types are equally common in all parts of everyday speech. Due to lexical differences, the distribution of the vowel types mentioned above may be different in written and spoken Finnish. Specific phonetic knowledge concerning informal Finnish speech is still lacking.

It is known that vowel segments occurring in accented

positions of utterances or in the stressed syllables of words tend to be pronounced with more articulatory effort or precision, i.e., they are phonetically less “reduced” than unaccented vowels. In turn, more reduced vowels are prone to coarticulation and exhibit more acoustic variability. Phonetic vowel reduction has sometimes been associated with a “centralization” effect seen on the F1/F2 formant chart (the most popular method for visualizing vowel quality). However, this centralization does not necessarily concern the speaker’s “articulatory target”, but is probably caused by averaging over vowel measurements: the more variable formant values, the closer the mean usually shifts to the centre of the chart [1].

Common word forms are not likely to be accented since they are more predictable than rare words. The most frequent Finnish word forms (often particles, pronouns, auxiliary verbs, etc.) are mono- or bisyllabic, but some forms may have many more syllables, thanks to the rich inflection system and the tendency to form new compound words “on the fly”. Finnish does not exhibit lexical word stress, but for any clearly pronounced or accented word, the primary word stress is regularly perceived on the first syllable, which is also phonotactically allowed the greatest complexity. However, in non-prominent utterance positions, word stress may be practically non-existent.

In this light, fewer vowel contrasts are necessary within common and predictable words, and more acoustic variability may be expected in these vowel occurrences.

## 2 METHODS

Five informal dialogues were recorded from ten young Finnish adults (five females) in an anechoic room. The speakers in each dialogue condition knew each other well, and they were allowed to chat freely and unmonitored for 45–60 minutes. Each speaker’s voice was recorded with a head-mounted microphone on a separate channel of a DAT recorder. The dialogues were transferred to a computer and downsampled to a rate of 22.05 kHz. The two channels of each stereo signal were separated, resulting in one high-quality speech

Category	Word form frequency ranks	Vowel set		Word form examples
		Initial syllable	Non-initial syllable	
I	1–20	 2	 0	se et nii niinku joo sit ei mut nyt tota
II	21–200	 13	 3	just joku kaikki varmaa juttu semmonen menee pääsee esimerkiks periaattees
III	201–2000	 18	 8	muistan oikeestaan kuule matka urbaani kivaa ilta kaupunkii lapset treenikämppäsystemi
IV	2001–5769	 14	 9	siideriä rikollisjengejä puoli opiskelukaveri niittenki kysyy juteltii hyperaktiivinen biisii avara

**Table 1:** Sets of vowels that were used by four Finnish speakers within word forms of different frequencies. Circle size indicates the relative frequency of a long or short vowel phoneme within all the analyzed syllables in the category. In addition to the long and short vowels, the number of diphthongs is shown at the bottom right corner of each vowel set. The frequency dictionary was built from five dialogue recordings.

signal per speaker. The Praat program [2] was then used for both labeling work and acoustic analyses.

Each speaker’s utterances were transliterated following Finnish orthographic conventions. Written and spoken Finnish are known to differ to some extent, and therefore no attempts were made to correct grammatical structure. A frequency dictionary of 5769 word forms was created from a total of 29102 word tokens in the five dialogues. Since morphological analyses have

not yet been completed, the frequency dictionary did in some cases contain several structurally identical occurrences of a word form. Moreover, it is to be noted that the dictionary is too small to represent all the lexical properties of spoken Finnish. However, it can be used to estimate the distribution of vowel types for common vs. less frequent word forms. For this purpose, the frequency dictionary was divided into four categories: category I (20 most common word forms,

covering 10619 word tokens), category II (word forms with frequency ranks from 21st to 200th; 8455 tokens), category III (201st to 2000th; 6259), and category IV (2001st to 5769th; 3769). Category IV contained only forms that occurred exactly once in the five dialogues, whereas forms in category I typically occurred hundreds of times and were used by all speakers.

For two female (S1 and S2) and two male speakers (S3 and S4), fragments of the recorded material were phonetically segmented and labelled. The boundaries of each vowel phone were marked so that the quality of the adjacent consonants could no longer be perceived, thus excluding the strongest transition phases. This method was preferred since the duration of consonant-vowel transitions typically varies for different consonants (e.g., for approximants vs. stops).

For the four speakers, all phonetically labeled *vowel segments* were analyzed. In this study, a vowel segment refers to the combination of any contiguous phone segments that had been labeled as vowels within one syllable. A phonetic vowel segment may contain one or more vowel phones of any duration, but a vowel sequence crossing a syllable boundary would be considered as two vowel segments. In order to link each vowel segment with a specific word token, syllables and words were also marked and orthographically labelled. Since the Finnish orthography has a nearly one-to-one correspondence with phonemic structure, the vowel phoneme(s) in each syllable could be roughly derived from the transliteration. This vowel phase could then be linked with the phonetic vowel segment that was actually produced, regardless of the number of phonemic or phonetic vowel types. Here, the term *vowel phoneme* refers to the vowel(s) derived from the orthographic transcription of a syllable.

The frequencies of two lowest formants (F1 and F2) were calculated at the temporal midpoint of each vowel segment using Praat [2] (Burg algorithm; parameters were adjusted for male and female speakers accordingly). Those vowel segments that did not yield acceptable values for both F1 and F2 were discarded.

### 3 RESULTS

Table 1 shows the relative numbers of different vowel phonemes that were used by four speakers in the four word form frequency categories. The vowel set appears to be quite limited for the most frequent word forms (category I), whereas a much richer vowel inventory is used in categories III and IV. Since all 20 word forms in category I were either mono- or bisyllabic, only three monophthongs are seen for the non-initial syllables in this category. For less frequent, often multisyllabic forms, the number of vowel types in non-initial syllables increases.

Formant frequencies from the temporal midpoints of actual vowel segments are shown in Table 2 for two speakers (S1 and S4; very similar charts were obtained for the two other speakers as well). To avoid a bias caused by restricted syllable structure (cf. table 1), only data from word-initial syllables are shown. Circles represent mean formant values for tokens of vowel phonemes, and values for individual vowel segments are plotted as dots. As predicted, within frequent words, there is large variability for individual vowel segments, and also the phoneme means are bundled up and “centralized”. For rare words with a more complex vowel phoneme inventory, the mean formant values disperse towards the periphery, and even the individual measurements appear to organize into a v-shaped formation.

Since diphthongs are produced by a gliding movement, it was expected that the temporal mid points of diphthong segments would tend to fall between the mid points of monophthongs. This was indeed the case. In Table 2, diphthong segments are plotted in grey colour. The mean values of diphthongs (see the grey circles) are between those of monophthongs (black circles) in all four frequency categories.

## 4 CONCLUSIONS

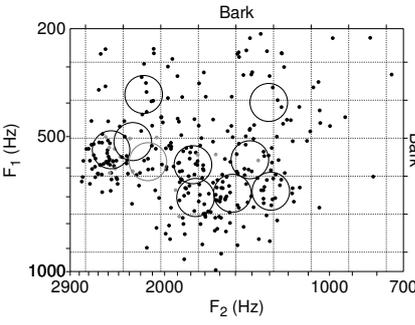
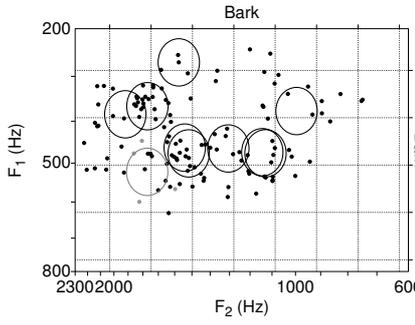
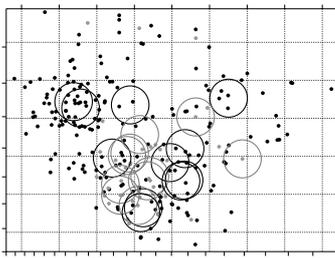
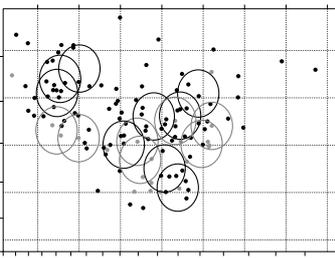
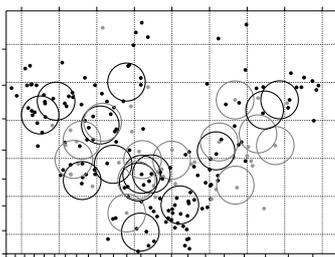
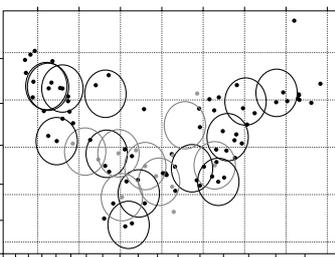
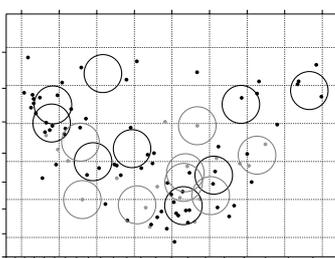
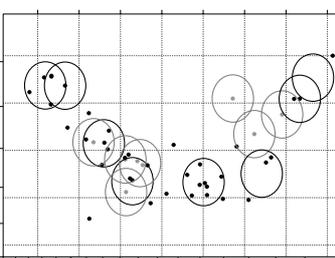
Rare words are more likely to be prominent or accented in speech than frequent and predictable words, which affects the phonetic vowel qualities that can be expected to occur. In this preliminary study, fragments of Finnish informal dialogues were analyzed. According to the frequencies of first two formants, vowel segments within tokens of frequent word forms were found to exhibit more phonetic variability than vowels in rare words. It is likely that people in fact take advantage of such tendencies in the production and comprehension of everyday speech.

## ACKNOWLEDGMENTS

This study was partly funded by Academy of Finland (project *Integrated Resources for Speech Technology and Spoken Language Research in Finland*) and INTAS (project 915).

## References

- [1] Dick van Bergem, *Acoustic and lexical vowel reduction*, Ph.D. thesis, University of Amsterdam, 1995.
- [2] Paul Boersma, “Praat: doing phonetics by computer,” 1992–2003, available at: <http://www.praat.org/>.

Category	Word form frequency ranks	Formant values at vowel centre	
		Speaker 1 (female)	Speaker 4 (male)
I	1–20	 <p>/i: i y e i e a æ o u o:/</p>	 <p>/i: i e i y e a o: o æ u/</p>
II	21–200	 <p>/i i: e iæ y ai ie oi æ: æ ei o: a a: o ui u ou/</p>	 <p>/i: ei i y ie e ai æ a a: o u o: ua/</p>
III	201–2000	 <p>/i: i iæ ie e: y: ei e y æi æ æ: yø ø: a: oi a o uo au ui ua u: ou u/</p>	 <p>/i e: i: ei y ia e i æy æ: æ eu ai a oi a au a: o u: u/</p>
IV	2001–5769	 <p>/i: i ei ie e y æ ai a oi uo ui au o u ou u:/</p>	 <p>/i y: ie e ia æi æ æy a ui ua o uo u: u/</p>

**Table 2:** Formant frequencies at the temporal midpoints of vowels in word-initial syllables vs. the frequency ranks of the corresponding word forms. Each dot corresponds to the vowel segment of a syllable. Circles represent the average formant values for vowel phonemes. Values for diphthongs are plotted in grey colour. Vowel phoneme labels from left to right are indicated below each figure. The grids in the background as well as the size of the circles indicate perceptual “critical bandwidths” of one Bark. Note how the phoneme circles are bundled up in common word forms, whereas in rare words the vowel categories disperse.