F1/F2 TARGETS FOR FINNISH SINGLE VS. DOUBLE VOWELS

Satsuki Nakai*†, Kari Suomi** and Alan Wrench‡†

*University of Glasgow, UK, †Queen Margaret University, UK, **University of Oulu, Finland, ‡Articulate Instruments Ltd.
satsuki@ovod.net, kari.suomi@oulu.fi, awrench@articulateinstruments.com

ABSTRACT

This paper explores the reason why Finnish single (short) vowels tend to occupy less peripheral positions in the F1/F2 vowel space compared to their double (long) counterparts. The results of two production studies suggest that the less extreme vowel quality of single vowels is best described as arising from undershoot of articulatory/acoustic targets due to their short durations, assuming single, context-free targets for phonemes.

Keywords: Finnish, vowel, quantity, quality, undershoot

1. INTRODUCTION

Finnish has eight vowel phonemes /i, y, e, ø, æ, ɑ, o, u/, each of which can occur as a single vowel (with short duration) or a double vowel (with long duration) [4, 13]. Acoustic studies indicate that, in addition to being shorter in duration, single vowels generally occupy less peripheral positions in the F1/F2 vowel space, as compared to their double counterparts [15]. Yet, the quality difference between single vs. double vowels has long been said to have no phonological significance, and to play little role in the perception of the single vs. double opposition [5, 6].

This view was challenged by O'Dell [11], who conducted perception experiments using two stimulus continua, each spanning from *tuli* to *tuuli*. The two stimulus continua were each created from spoken *tuli* and *tuuli*, by temporally manipulating them in such a way that the segmental durations of the two stimulus continua matched. Despite the matching segmental durations, Finnish listeners heard *tuli* more often in the stimulus continuum created from *tuli* and heard *tuuli* more often in the stimulus continuum created from *tuuli*. As the target vowel in the spoken *tuli* was more centralised in quality than the target vowel in the spoken *tuuli*, O'Dell suggested that quality information is relevant to the perception of single vs. double vowels, along with F_0 information that also differed between the two stimulus continua. This raises a possibility that Finnish single vs. double vowels have distinct articulatory/acoustic targets not just for duration but also for quality.

On a view that context-free articulatory/acoustic targets exist for phonemes [8], however, the observed quality difference between Finnish single vs. double vowels can be considered to result from greater target "undershoot" in single vowels, as single vowels are likely to be more affected by coarticulation due to their short durations [13]. For example, Hirata and Tsukada [3] reported that Japanese short vowels were generally realised with less extreme F1 and F2 compared to their long counterparts, but that the quality difference between short and long vowels diminished considerably in slow speech, where short vowels presumably had additional time to move closer to their targets. Conceivably, the centralised vowel quality of Finnish single vowels results from a similar mechanism.

In this paper we ask whether Finnish single vs. double vowels can be considered to have different quality targets (as defined by their positions in the F1/F2 acoustic space) or have the same target but their realisations differ because of undershoot in single vowels. To answer the question, we first examine the quality difference between some Finnish single vs. double vowels produced in tightly controlled laboratory materials (Study 1). If single vs. double vowels have distinct quality targets, the distributions of F1 and F2 values for each pair of single vs. double vowels should be largely separated, when the context is rigorously controlled. We then look at the relationship between vowel duration and quality, by exploiting durational variation in single and double vowels reflecting their position in word and/or word structures in which they occur (Study 2). If the more centralised quality of single vowels arises from target undershoot, with increase in duration the quality of the single vowel should approach that of the double vowel.

2. STUDY 1: QUALITY OF SINGLE VS. DOUBLE VOWELS IN MINIAL-PAIR WORDS

2.1. Methods

2.1.1. Data

We examined a subset of speech data originally collected for Nakai et al. [10]. Test vowels were single and double vowels (underlined) in the following four minimal-pair Finnish words: *kataa* - *kaataa* (/ɑ/ vs. /ɑɑ/), *pesä* - *pesää* (/æ/ vs. /ææ/), *siisti* - *siistii* (/i/ vs. /ii/) and *puskaan* - *puuskaan* (/u/ vs. /uu/). Two of the four pairs of vowels (/ɑ/ - /ɑɑ/ and /u/ - /uu/) carried primary lexical stress, which is fixed on the first syllable of the word in Finnish, e.g. [5, 13]. Each of these words was produced in an utterance frame in response to a pre-recorded precursor question, designed to elicit phrasal accent before the test word, and the test word itself without phrasal accent. For instance, the test word *kaataa* ("overturn") was embedded in the following question (Q) and answer (A):

(Pre-recorded) Q: *Kenen mielestä kaataa sopii numeroon yksitoista*?
(Who thinks "overturn" fits number eleven?)
A: *HÄNESTÄ kaataa sopii numeroon yksitoista.*
(S/HE THINKS "overturn" fits number eleven.)

The final word *yksitoista* ("eleven") in the example precursor question and the answer was altered depending on the test word, so that the total length of the answer was constant in syllables as well as moras regardless of test word.

Each test word was produced in the utterance frame by eight female speakers of Northern Finnish (age range: 21-35, *M* = 24) three times in separate blocks.

2.1.2. Duration Measurement

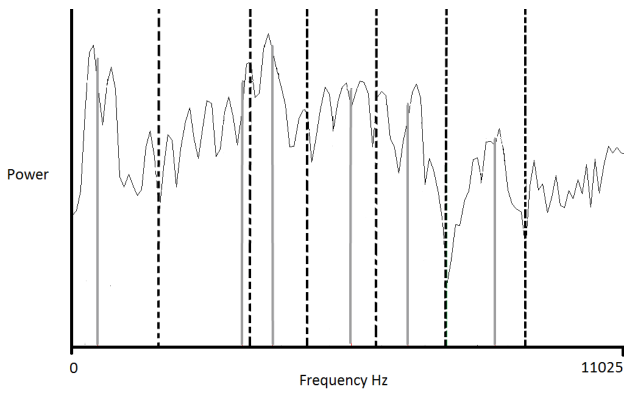
The duration of each test vowel was measured using a supralaryngeal criterion [14]. The voice onset time (VOT) following the release of an oral stop phoneme was included in the vowel interval.

2.1.3. Formant Estimation

The most common approach to estimating formants is to apply an all-pole model (linear predictive coding, or LPC) to the speech signal, and peak-pick the resulting spectral envelope [9]. In that approach, the formant estimates can be biased towards the harmonic peak when speech does not comply with the assumptions of the all-pole model such as lateral or nasalised segments, or when the speaker's F_0 is high, and the harmonics are consequently widely spaced. We therefore used a multiple centroid approach, which has been shown to reduce harmonic bias [16].

Multiple centroid analysis simply divides the spectrum into multiple partitions and calculates the centroid of each partition (see Fig. 1). How many partitions to have and how to determine the boundaries are determined in two stages. First, the researcher looks at data from a given speaker and estimates the maximum and minimum frequency that each formant can have across all speech samples. This also determines the number of partitions. The ranges for consecutive formants will necessarily overlap. The region of overlap gives the range of frequencies where each partition boundary can lie. Second, for each possible set of boundaries the centroid frequency is calculated for each of the resulting contiguous partitions. The calculation produces a minimum squared error value. This error is noted for every possible partitioning, and the set of centroid frequencies with the minimum error is selected.

Figure 1: Power spectrum divided into six partitions (dashed lines). Within each partition a centroid frequency (grey lines) is calculated. (Note: upper boundary of the last frequency partition is less than the maximum Nyquist frequency.)



Using the multiple centroid approach implemented in AAA software [1], F1 and F2 were estimated for each test vowel at the F1 turning point (i.e. point furthest away from the coarticulatory effects of the flanking consonants), determined by averaging time-normalised F1 trajectories of all instances of each target vowel. Where a clear F1 turning point was absent, F1 and F2 were estimated at the midpoint of the vowel interval.

2.2. Results

Fig. 2 shows the durations of the four pairs of single vs. double vowels. Consistent with previous studies [6, 13, 15], double vowels were roughly twice as long as their single counterparts, and the two categories were generally well separated in duration. According to paired t-tests performed on mean duration of each target vowel calculated for each speaker, all single vowels were significantly shorter than their double counterparts (all *ps* < .001, Bonferroni-corrected).

Figure 2: Durations of single vs. double vowels. Error bars represent ± *SD*.

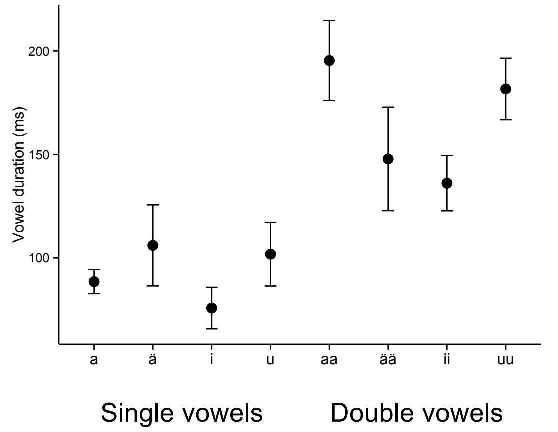
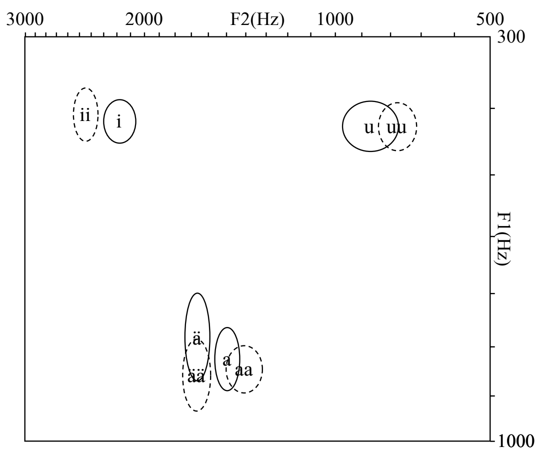


Fig. 3 shows estimated F1 and F2 of single vs. double vowels. Also consistent with previous studies [11, 15], single vowels were on the whole less extreme than their double counterparts. According to paired t-tests performed on mean F1 and F2 values of each target vowel calculated for each speaker, single vowels were significantly less extreme than double vowels along F1 for /æ/ vs. /ææ/ (*pesä* vs. *pesää*, $p < .002$), and along F2 for /ɑ/ vs. /ɑɑ/ (*kataa* vs. *kaataa*, $p < .002$), /i/ vs. /ii/ (*siisti* vs. *siistii*, $p < .001$) and /u/ vs. /uu/ (*puskaan* vs. *puuskaan*, $p < .05$; all *p*-values are Bonferroni-corrected).

Figure 3: Estimated F1 and F2 values of single vs. double vowels. Ellipses cover ± *SD*.



Unlike duration, however, there was non-negligible overlap in the positions of single vs. double vowels in the F1/F2 space, except those of /i/-/ii/ in *siisti* vs. *siistii* (see Fig. 3). This was the case, despite the fact that the target vowels were produced in tightly controlled contexts by a homogenous group of speakers. Thus, in Finnish vowel quality does not appear to be as reliable a cue as duration to the single vs. double vowel opposition. Finnish linguists' intuition that quality is not relevant to the perception of single vs. double vowels most probably is a reflection of the difference in the reliability of these cues. Given observations that undershoot of vowel formants is greater in vowels of shorter durations [3, 8], the greater separation of /i/-/ii/ in *siisti* vs. *siistii* in the F1/F2 space may be attributed to the duration of /i/ in *siisti*, which was much shorter than other single vowels (see Fig. 2).

3. STUDY 2: RELATIONSHIP BETWEEN VOWEL DURATION AND QUALITY

Within each category, Finnish single vs. double vowels vary systematically in duration, depending on their position in word as well as the structure of the word in which they occur [6, 13, 15]. For example, in many varieties of Finnish (including Northern Finnish), the single vowel V_2 is much longer in duration in $CV.CV_2$ than in $CVV.CV_2$. In Study 2 we exploit the systematic durational variation in single and double vowels to more directly examine the relationship between vowel duration and quality.

3.1. Methods

This study used four nonwords collected for Nakai et al. [10] as test words: *sasa*, *saasa*, *sasaa* and *saasaa* (/ˈsɑsɑ/, /ˈsɑɑsɑ/, /ˈsɑsɑɑ/, /ˈsɑɑsɑɑ/). We examined vowels in phrasally unaccented rendition of each test word, elicited using the same utterance frames and precursor questions as in Study 1 (see 2.1.1). In order to include single vs. double vowels of more varied durations, phrasally accented test words were also used. These were elicited in the same utterance frames as phrasally unaccented test words, by presenting different precursor questions. For *saasaa*, for example, the following precursor question (Q) and answer (A) were used:

(Pre-recorded) Q: *Onko vastaus numeroon yksitoista läikkyä*?
(Is "spill" the answer to number eleven?)
A: *Hänestä SAASAA sopii numeroon yksitoista*.
(S/he thinks "*SAASAA"* fits number eleven.)

Each test word in each condition (phrasally accented vs. unaccented) was produced twice in separate blocks by the same eight speakers as in

Study 1. F1, F2 and duration of all vowels in the test words were measured. Other aspects of the methods were identical to Study 1.

3.2. Results

Mean F1, F2 and duration of /ɑ/ and /ɑɑ/ in the test words were calculated separately for each position of each word and condition. Figs. 4 and 5 summarise the results.

Figure 4: Mean F1 and duration of each vowel in each test word (▼: /ɑ/ in phrasally unaccented word; ●: /ɑ/ in phrasally accented word; ▼: /ɑɑ/ in phrasally unaccented word; ●: /ɑɑ/ in phrasally accented word).

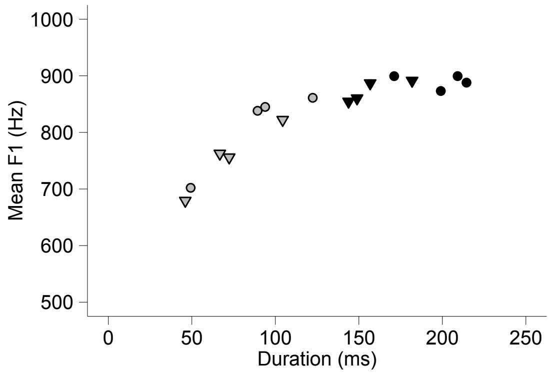
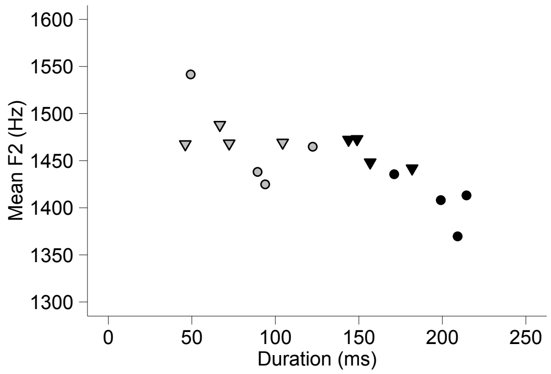


Figure 5: Mean F2 and duration of vowels in each vowel in each test word (▼: /ɑ/ in phrasally unaccented word; ●: /ɑ/ in phrasally accented word; ▼: /ɑɑ/ in phrasally unaccented word; ●: /ɑɑ/ in phrasally accented word).



As is evident in Fig. 4, there was a positive correlation between F1 values achieved by the vowels and their durations, with F1 values reaching a plateau when the duration of the vowel exceeded c. 150 ms. Fig. 5 suggests a negative correlation between F2 values achieved by the vowels and their durations. Results from hierarchical regression models indicated that vowel duration and its interaction with vowel quantity significantly affected F1 ($p < .001$ in both cases), over and above the effect of vowel quantity. Vowel duration also significantly affected F2 ($p < .003$), over and above the effect of vowel quantity. As /ɑ/ and /ɑɑ/ are low back vowels, these correlations indicate that the longer their durations, the more extreme positions in the F1/F2 space the vowels occupied. In other words, the degrees of centralisation of single (and double) vowels are tied to their durations. Notice also that mean F1 and F2 of longest single vowels are similar to those of shortest double vowels. Together, the results suggest that the less extreme quality of single vowels arises from their generally short durations, consistent with the undershoot view.

4. CONCLUSION

In this paper, we asked whether centralised vowel quality reported for Finnish single vowels (as compared to double vowels) arises from different quality targets between single and double vowels, or from target undershoot in single vowels. In Study 1, we found that single vs. double vowel pairs produced in tightly controlled laboratory materials were more distinct in duration than in quality, although single vowels on the whole occupied more central positions in the F1/F2 space. In Study 2, we found that physical durations of a single vowel /ɑ/ and a double vowel /ɑɑ/ affected the vowels' F1 and F2, over and above the effect of vowel quantity.

In conclusion then, the general tendency for single vowels to occupy more central positions in the F1/F2 vowel space appears not to result from different quality targets for single vs. double vowels but undershoot of targets in single vowels, which do not always have enough physical durations to achieve their F1/F2 targets.

This is not to say that quality information plays no role in the perception of quantity in Finnish. Given the observed correlation between vowel quality and duration, a primary cue to vowel quantity, quality differences can be useful additional information to the listeners, as found by O'Dell [11]. We also note that the above conclusion rests on an assumption that a language has a single, context-free target for each phoneme. This assumption can be considered to have some psychological reality, as evidenced in, for example, writing systems of many languages (including Finnish) that are phoneme based. At the same time, however, some researchers have suggested the role of units larger than phonemes in speech production [2, 7, 12]. If we view context-sensitive allophones as targets, we could argue that at least in some contexts (e.g. /i/ vs. /ii/ in *siisti* vs. *siistii*, as we saw in Study 1) Finnish single and double vowels have clearly distinct quality targets. Either way, though, duration enjoys the status of a primary cue to the single vs. double vowel opposition in Finnish.

5. ACKNOWLEDGEMENTS

Data collection for this work was funded by a Leverhulme Research Grant (F/00158/AF) and a British Academy Small Research Grant (SG-39167). We thank Alice Turk, Sari Kunnari and Riikka Ylitalo for their involvement in an earlier project, for which the data were originally collected.

6. REFERENCES

[1] Articulate Instruments Ltd. 2011. Articulate Assistant Advanced User Guide: Version 2.13. Edinburgh: Articulate Instruments Ltd.

[2] Cholin, J., Dell, G. S., Levelt, W. J. 2011. Planning and articulation in incremental word production: Syllable-frequency effects in English. *J. Exp. Psychol. Learn.* 37, 109–122.

[3] Hirata, Y., Tsukada, K. 2009. Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* 66, 129–149.

[4] Karlsson, F. 1969. Suomen yleiskielen segmentaalifoneemien paradigma [The paradigm of standard Finnish segmental phonemes]. *Virittäjä* 73, 351–362.

[5] Karlsson, F. 1987. *Finnish Grammar*. Porvoo: WSOY.

[6] Lehtonen, J. 1970. Aspects of quantity in Standard Finnish. PhD dissertation. University of Jyväskylä.

[7] Levelt, W. J., Wheeldon, L. 1994. Do speakers have access to a mental syllabary? *Cognition* 50, 239–269.

[8] Lindblom, B. 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am*. 35, 1773–1781.

[9] McCandless, S. 1974. An algorithm for automatic formant extraction using linear prediction spectra. *IEEE Trans*. ASSP-22, 135–141.

[10] Nakai, S., Turk, A., Suomi, K., Granlund, S., Ylitalo, R., Kunnari, S. 2012. Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *J. Phonetics.* 40, 796–807.

[11] O'Dell, M. 2003. Intrinsic timing and quantity in Finnish. PhD dissertation. University of Tampere.

[12] Pierrehumbert, J. 2002. Word-specific phonetics. In: C. Gussenhoven, T. Rietveld, N. Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 101–139.

[13] Suomi, K., Toivanen, J., Ylitalo, R. 2008. *Finnish Sound Structure: Phonetics, Phonology, Phonotactics and Prosody.* Oulu: Oulu University Press.

[14] Turk, A. E., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: a practical guide. In: S. Sudhoff, et al. (eds.), *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 1–27.

[15] Wiik, K. 1965. Finnish and English vowels: A comparison with special reference to the learning problems met by native speakers of Finnish learning English. PhD dissertation. University of Turku.

[16] Wrench, A. A., Watson, J. M. M., Soutar, D. S., Robertson, A. G., Laver, J. 1994. Fast formant estimation of children's speech. *Proc. ICSLP*94 3, 1651–1654.