

JYX



**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Kuronen, Mikko; Tergujeff, Elina

**Title:** Second Language Prosody and its Development : Connection between Different Aspects

**Year:** 2020

**Version:** Accepted version (Final draft)

**Copyright:** © 2018 Association for Language Learning

**Rights:** CC BY-NC-ND 4.0

**Rights url:** <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**Please cite the original version:**

Kuronen, M., & Tergujeff, E. (2020). Second Language Prosody and its Development : Connection between Different Aspects. *Language Learning Journal*, 48(6), 685-699.  
<https://doi.org/10.1080/09571736.2018.1434228>

## **Second language prosody and its development: Connections between different aspects**

### **Abstract**

This study focuses on the development of L2 prosody and in particular, whether different aspects can affect each other. Finnish-speaking learners of Swedish took part in a pronunciation and oral skills course and were recorded for various speaking tasks before and after the experimental intervention. Read-aloud declarative sentences (statements) from nine learners were acoustically analysed for several tonal and temporal aspects, focussing on the marking of primary stress. The results reveal that the learning of Swedish tonal word accent 2 (H\*LH) facilitates other tonal developments towards native-like utterance intonation. A link between tonal and temporal developments in the marking of primary stress was however not found, but the duration measurements suggest that articulation rate and tonal developments may be connected. The study's implications for teaching and future research regarding any L2 – not only Swedish – include that aspects of learning L2 prosody should not be treated as separate, and that more emphasis should be placed on the possible connections between the learning of different aspects.

**Keywords: second language pronunciation; prosody; intonation; rhythm; Swedish.**

### **Introduction**

This paper presents a study on learners' development in second language (L2) prosody. Pronunciation is an important component of speech, consisting of various features beyond individual sounds. Indeed, the appropriate use of prosodic features such as stress and intonation has been found to be more crucial for intelligibility than accurate production of individual segments, or sounds. Even though there is a need for the teaching and practising of sounds as well, numerous studies suggest that more emphasis should be put on prosody. For example, when a narrow teaching approach, concentrating on sounds, was compared to a broader, prosody-oriented one, the broader approach led to better learning results as judged by native-speaker listeners, who rated the learners on comprehensibility, accentedness and fluency before and after a training period (Derwing et al. 1998). Despite the importance of prosody, theoretical models of L2 learning have a bias towards description of how segments are learnt. In order to further develop these models, more knowledge on the acquisition of prosody is required.

The focus of the present study is on the realisation of primary stress (in both its tonal and temporal aspects) and on overall utterance intonation. Learning of both tonal and temporal patterns has seldom been examined in the same study. Overall, research on L2 prosody has tended to focus on one feature at a time, as pointed out in Trofimovich and Baker (2006). The present study has a specific reason for considering several aspects: we set out to explore

whether there is a connection (a) between the learning of tonal and temporal aspects of utterance prosody, and (b) between the tonal development of L2 primary stress and development of other tonal features such as the overall intonation contour, mean  $f_0$ , and pitch range.

Although we focus on these theoretically interesting questions, the present study also has a practical inspiration. It arises from uncertainty as to whether or not to teach L2 learners tonal word accents (= tone accents) in a language (Swedish) where they characterise first language (L1) prosody but are rarely crucial for intelligibility (Abelin and Thorén 2015). In addition, lack of, or even mixing up, tonal word accents does not seem to play a great role in foreign accentedness ratings (Author A, Author A and B, Cunningham-Andersson and Engstrand 1989). In this light, one is tempted to disregard this learning goal even with advanced learners and focus on features that contribute more to intelligibility (and low accentedness). However, if a connection between tone accent and other tonal features exists, and if learning of tonal aspects facilitates learning of temporal aspects, teaching tone accents in L2 pronunciation may be worthwhile after all.

### **Learning second language prosody**

The most commonly cited theoretical models on second language pronunciation focus on the learning of sounds, not prosody. However, processes connected to the learning of sounds may be applicable to the learning of prosody as well. In the following, we will explore the prevailing models and in particular, whether they suggest any possibility for different features to develop in connection with each other.

According to the Speech Learning Model (SLM, Flege 1995), a clear perceived dissimilarity between L1 and L2 sounds facilitates the building of a L2 sound category. Thus, L2 sounds that are clearly different from L1 sounds are easier to attain (or at least to perceive, if not produce), whereas L1 and L2 sounds with similar qualities can easily assimilate and fall into a merged category. Similarly, the Perceptual Assimilation Model for L2 (PAM-L2, Best and Tyler 2007) suggests that when a naive listener hears an unfamiliar nonnative sound, they assimilate it with the closest equivalent in their L1 – unless its quality is too far from any L1 sound. In contrast to SLM, however, PAM-L2 does not address equivalence merely at the phonetic level but also at the phonological level. Both SLM and PAM-L2 are based on the idea that L2 production develops gradually and is tied to exposure to the target language. These processes have been found to apply to the development of prosody as well (Trofimovich and Baker 2006). SLM and PAM-L2 do not touch on the possibility of features developing in connection with each other.

Eckman's (1977) Markedness Differential Hypothesis (MDH) proposes that typologically marked language features correspond to degree of difficulty in L2 learning; in other words, features that are universally common to languages are easier to learn. Connections between the development of different features are not discussed in this model either. Markedness plays a role in other theories such as Prince and Molensky's (1993) Optimality Theory (OT) and a recent attempt to model the learning of L2 prosody at word level, Özçelik's (2016) Prosodic

Acquisition Path Hypothesis (PAPH), which builds upon Dresher and Kaye (1990). Özçelik's model assumes that word-level prosodic parameters are hierarchically organised into a tree, with some parameters embedded under others, and that learning proceeds in well-defined stages. However, the model is limited to the learning of word-stress patterns, and again, does not consider cross-feature connections such as word-stress pattern being linked to utterance intonation.

These theories regarding second language pronunciation learning do not consider connections between features, although a hierarchical relationship of word-level prosodic parameters is assumed by Özçelik (2016). As pointed out by Trofimovich and Baker (2006), learning of prosodic features has often been investigated one feature at a time, and possible connections between features are thus disregarded. When Trofimovich and Baker (ibid.) studied the effect of experience and age of arrival on the learning of several suprasegmental features, the results showed that suprasegmental development happens gradually over time, as does segmental development, and that suprasegmental learning varies across features. The authors conclude that some features (pause frequency, pause duration) may be learnt relatively early in the studies/exposure, whereas other features (stress timing) require more exposure and practice to develop. Yet, there may be aspects of L2 prosody (speech rate, tonal peak alignment) that pose a significant challenge to the learner even after a decade of exposure. This could suggest that features develop independently, but also does not exclude the possibility of certain features developing hand in hand.

The present study is not bound by any of the abovementioned theories. Our ambition is to describe the learning of tonal and temporal features at both phonological (= tone accent and duration as a stress correlate) and phonetic (= overall intonation contour, mean  $f_0$ , and pitch range) levels. This enables us to explore possible connections between different parameters and prosodic levels, which is rare in previous research (Gut 2007) and as we have pointed out, has been overlooked in the learning theories.

### **Context of the present study**

The context of the present study is advanced L1 Finnish-speaking learners studying Swedish (in this study the standard pronunciation of Sweden-Swedish; Riad 2014). This context was chosen because there are considerable prosodic differences between these two languages, and the learning scenario is thus theoretically interesting. The differences are phonological and phonetic, and they concern both rhythm and intonation. To give a background to our research questions, we will briefly describe the main differences in prosody between the two languages. The description focuses on prosody in declarative sentences (statements), which is the speech material used in this study.

The default intonation in declarative sentences in Finnish is declining (Iivonen 2009). Fundamental frequency ( $f_0$ ) typically has the highest peak in the beginning of the utterance (Figure 1). Lexical stress is fixed on the first syllable of the word, and  $f_0$  is also declining in

the stressed word (Vainio et al. 2006). The phonetic correlates of stress are a rise in  $f_0$  and amplitude. Duration has only a marginal role as a stress correlate due to the extensive use of length for phonological contrasts (Suomi et al. 2008). In Finnish, length is a meaning-distinguishing feature, e.g. *tuli* ‘fire’ vs. *tuuli* ‘wind’. The following  $f_0$  peaks in an utterance are normally lower than the initial peak, even if the primary-stressed word is in the latter part of the utterance (Iivonen 2009). Mean  $f_0$  seems to be lower and pitch range smaller in Finnish than for example in English (Laukkanen et al. 1999), Russian (Ullakonoja 2011), and Swedish (Author A). These features are probably due to the relatively simple intonation system in Finnish.

### (FIGURE 1)

Swedish intonation typically does not fall entirely in declarative sentences (Figure 1, see also Leinonen et al. 1990). Even though there is often an  $f_0$  peak in the first word or in the initial part of an utterance (so-called initiality accent; Myrberg 2010), this first peak is seldom as high as the peak(s) in the primary-stressed word, normally located in the latter part of the utterance. The main correlates of stress are duration and  $f_0$ . In the description of Swedish, three different prominence levels are distinguished: two primary stress levels (focus accent and word accent) and unstressed pronunciation of a word (Riad 2014). The stressed syllable in a primary-stressed word is long (V:C, VC: or VCC), whereas an unstressed word has no long syllables. Both primary stress levels bring about a phonological contrast between tone accents 1 and 2 (Figure 2). As focus accent, tone accent 1 is realised as L\*H (low tone + high tone, *and/-en*, ‘duck’) and tone accent 2 as H\*LH (*ande/-n*, ‘spirit’). As word accent (a lower level of primary stress), tone accent 1 is realised as HL\* and tone accent 2 as H\*L (Bruce 1977, Riad 2014). Swedish is characterised by some degree of destressing and shortening of the phonetically unstressed part of the utterance (Bruce 2010), which either does not exist or is significantly weaker in Finnish. This leads in Swedish to a greater prosodic difference between the stressed and the unstressed part(s) of the utterance than in Finnish.

### (FIGURE 2)

To sum up, the main prosodic differences between Finnish and Swedish are (a) in the general intonation contour of an utterance, (b) in the phonetic correlates of stress, (c) in the intonation of the primary-stressed word, and (d) in rhythmic differences, due to (b). Further, the mean  $f_0$  and  $f_0$  standard deviation (SD) can be affected by tone accents and could thus be considered as learning goals for Finnish speakers of Swedish (Figure 1).

Previous studies on learning of Swedish prosody have focused on tone accents and quantity at word level. The results concerning tone accents show that they are a difficult but a possible learning goal even for adult L2 speakers (e.g. Tronnier and Zetterholm 2013, 2014); also for those with no word tonal opposition in their L1 (Author A). Both formal and informal

learning can lead to acquisition of tones (Author A, Tronnier and Zetterholm 2013). Tone accent 2 seems to be more difficult to attain than tone accent 1 – perhaps due to the markedness of tone accent 2 (Author A, for discussion on markedness of tone accent 2 see Riad 2015). Syllable quantity has been shown to be more important for word recognition than tone accent (Abelin and Thorén 2015), and durational aspects seem to be more important for perception of syllable quantity than spectral qualities (Thorén 2008). Studies on rhythm and intonation at utterance level are very few (cf. a pilot study by Hackman 1977; on learning of L2 prosody in Finland-Swedish, see Kautonen 2017). Thus, empirical knowledge on the connections between different prosodic features in L2 Swedish is practically non-existent.

## **Research design**

### *Aim*

In the present study, production of L2 prosody and its development as a result of training were investigated. The focus was on tonal and temporal features and in particular, possible connections between different features. The following research questions were addressed:

1. How does L2 prosody (intonation, temporal pattern) develop as a result of a six-week training period?
2. Are there any connections between different tonal features (tone accent 2, pitch range, mean  $f_0$ , standard deviation in  $f_0$ ,  $f_0$  difference between stressed and unstressed part of utterance) in the development of L2 prosody? If so, what kind?
3. Is learning of tonal and temporal features of L2 prosody connected? If so, how?

To answer the research questions, a teaching experiment was carried out with L1 Finnish-speaking learners studying Swedish. In the analyses, we concentrated on the development of primary stress, tone accent 2, and general intonation contour. Tone accent 1 was excluded since learning of this feature had been studied earlier (Author A, Author A). For the same reason, the lexical learning of tone accents was excluded (the distribution of tones is morphologically based; Riad 2015).

### *Participants*

The participants were 25 advanced L1 Finnish-speaking learners of Swedish, of whom 9 (4 male, 5 female) were selected for closer examination based on number of utterances produced without major interruptions and hesitation in the recordings. The learners were 21–34 years old. They had studied Swedish for at least six years as part of their secondary education, after which they had entered the university to study Swedish as their major or minor subject. They reported that their previous studies had not included much explicit pronunciation teaching, and any teaching they had received had focused on the Finland-Swedish variety<sup>1</sup>. None of the participants had had extensive exposure to Sweden-Swedish. The learners' overall proficiency level in Swedish was around B1–B2 on the CEFR scale (Council of Europe 2001), as judged by the authors of this paper, both very familiar with the CEFR scale and assessment of oral skills in L2. Reference materials were also collected from four native

speakers of Sweden-Swedish. They were all female, aged 24–45. All resided and had grown up in the Stockholm region.

### *Teaching experiment*

The participants took part in a combined pronunciation and overall oral skills course in Swedish. They were all taught by author A; an L1 Finnish-speaking Senior lecturer in Swedish, specialising in pronunciation teaching. Based on interviews conducted with the students (not dealt with in this paper) and discussions in the classroom, they were highly motivated to improve their pronunciation. The participants had twelve 90-minute sessions of contact teaching over six weeks, at a rate of two sessions per week. The sessions took place in a language lab, but many of the exercises, e.g. group discussions, were undertaken without the lab technology. The course consisted of varied theoretical and practical training in Swedish prosody and segments, including learners recording their speech and listening back to it. Pronunciation-specific listening tasks were also used, and the instructor gave individual feedback on students' pronunciation. The participants were encouraged to practise at home, and many discussions in the classroom were based on home assignments, which ensured that homework was completed conscientiously. The phonetic description of Swedish given in the teaching was based on (Author A), which was also the course literature. Primary stress and tone accents were included in the learning goals. They were dealt with in two of the sessions, and they were practised throughout the course. Because the teacher was not a native speaker of Swedish, he did not act as the only pronunciation model during the course; native-speaker models were also offered through the online Svenska Lexin dictionary (<http://lexin.nada.kth.se/lexin>), through various listening exercises, and Swedish radio.

### *Materials*

Materials used for this research were declarative sentences with primary stress placement towards the end. The sentences included a disyllabic accent 2 word in a position that normally receives primary stress with focus accent (H\*LH). Declarative sentences were chosen as speech material, because all investigated parameters occur in them in controlled and comparable manner. Further, relatively short declaratives are very common in ordinary oral communication. Read-aloud speech was used instead of spontaneous speech in order to have identical tokens (a) when comparing L2 and L1 speakers' pronunciation, and (b) when analysing L2 speakers' development between pre- and posttest (see below). These analyses would have been rather impossible to carry out accurately if using spontaneous speech. The analysed sentences are presented below with English translations. Usual primary stress placement is marked with bold print.

- |  |                                       |
|--|---------------------------------------|
| 1. Jag vill <b>tacka</b> dig.                | I want to thank you.                  |
| 2. Jag vill köpa en <b>kappa</b> till mamma. | I want to buy a coat for (my) mother. |
| 3. Det kan <b>blåsa</b> i kväll.             | It might be windy this evening.       |
| 4. Vi ska till <b>stugan</b> snart.          | We are going to the cottage soon.     |

5. *Han är ganska dålig, vill jag minnas.* He is quite bad, I recall.  
 6. *Han köpte en matta sa pappa.* He bought a carpet, father said.

### **Procedures**

Participants recorded a list of 33 read-aloud sentences and other tasks (not dealt with in this paper) before and after the course. Identical materials were also collected from the four native speakers of Swedish for reference. The six sentences presented above were selected for closer analyses. The sentences were recorded in a language lab. Before the recording, the participants were allowed to take a look at the sentences but not to practise. They were instructed to read any sentence again, if they were not satisfied with their performance, e.g. if they misread a word or felt like they had made an obvious pronunciation error. The six sentences that were chosen for the analyses were produced without hesitation before and after the pronunciation course by all 9 participants. The total amount of analysed Swedish utterances was 132 (learners: 6 utterances x 9 speakers x 2 (pretest and posttest) = 108 utterances; native speakers: 6 utterances x 4 speakers = 24 utterances). In addition, the pronunciation of 4 utterances in Finnish by 7 learners were analysed for reference.

Next, the participants' productions were subjected to a listening test, in which 4 expert listeners identified primary stress placement in the utterances. If the listeners did not find primary stress on the assumed word, that particular sentence was omitted from the data. This option was used for 4 utterances (considered in the sample sizes stated above). As for the native-speaker utterances, the primary stress was determined by studying pitch and placement of tone accent 2. In addition, the authors also listened for the primary stress. Concerning native speakers, it is hardly possible *not* to locate the primary stress correctly based on an acoustic-auditory analysis as described above. One native-speaker utterance was omitted from the data due to exceptional primary stress placement.

### **Analyses**

Participants' productions of the utterances were segmented and annotated into syllables by using Praat (Boersma and Weenink 2015). Syllables were defined following Bruce (2012: 30–31). A standard segmentation procedure was followed, relying on the waveform, spectrogram and the ear. Segmentation was done by author B for all data, and then checked by author A. Having segmented the data into syllables, the next step was to measure syllable duration. This was done using a Praat script by Lennes (2002). Focus was on the word with primary stress. The following calculations were made: (1) duration of word with primary stress out of total duration of the utterance, (2) duration of stressed syllable in a word with primary stress out of total duration of the utterance, and (3) duration of stressed syllable in a word with primary stress in relation to the previous stressed<sup>2</sup> syllable. In addition, we measured (4) utterance duration to study possible changes in articulation rate. These measurements and calculations were also made for the native-speaker materials. Sample size as for the learners was 57, and 23 as for the native speakers. Statistical significance between learners' pretest and posttest (sample size bigger than 30) was tested using a paired-samples *t*-test, and between learners and native speaker (sample size smaller than 30) by using Mann-



Whitney *U*. Effect sizes were calculated using Cohen's *d* that indicates the difference between means in terms of standard deviation units (Cohen 1988). Effect sizes were interpreted following the general guidelines: small effect >0.2, medium effect 0.6–0.8, large effect 0.8–1.3, very large effect >1.3. None of the effect sizes calculated in this study were below 0.4 which has been suggested the criterion threshold for pretest-posttest settings (Hattie 2009).

As for pitch movement, the following 8 measurement points were chosen (see Figure 3):

- (1) beginning of utterance, measured at the beginning of the first vowel;
- (2) f0 peak prior to sentence stress (cf. initiality accent);
- (3) unstressed syllable prior to sentence stress, measured at the middle of the syllable (e.g. in *en*) or vowel (e.g. in *ska*) depending on the voicing in the syllable in question;
- (4–6) intonation of word with primary stress, measured at 3 points (seeking the H\*LH-pattern): (4) peak f0-value in the first syllable, (5) value in the middle of the word (end of the first syllable or beginning of the second), (6) highest f0 in the second syllable;
- (7) unstressed syllable after primary stress, measured at the middle of the vowel; and
- (8) end of utterance.

The eight measurement points were chosen because they capture three important aspects of intonation in the utterance: general contour, relation between the stressed and the unstressed syllables, and contour in the primary-stressed word. Averages of the measures were used to create a general contour for each speaker. In addition, (9) f0 mean, (10–11) pitch range in utterance and in primary-stressed word, and (12) standard deviation (SD) were calculated for each speaker.

### (FIGURE 3)

The f0 measurements were made manually by author A. Mean f0 and standard deviation were analysed with Praat's pitch analysis. Before this was done, clearly incorrect f0 value points and places with creaky voice were devoiced, which is a standard procedure in acoustic analysis of f0. Pitch contours in spectrogram and pitch analysis were compared in order to detect any artifacts in the f0 analysis. The range in the utterance (highest f0 *minus* lowest f0) and in the primary stress were calculated from speaker averages. Values for f0 are given in both Hertz (Hz) and semitones (ST). The conversion from Hz to ST was done by a converter published by Traunmüller (2005). Semitones were used to enable a comparison in f0 range by different speakers and by different genders. We acknowledge that comparing f0 range in males and females is not completely unproblematic, as some gender differences can exist. However, they seem to be very moderate; clearly 'less sex-linked than F0 values' (Honorof and Whalen 2010: 3097, see also Honorof and Whalen 2005, Henton 1995 and Pépiot 2014). Further, the focus of our analysis is not on gender differences but rather on the individual L2 speaker development. Statistical significance between learners' pre- and posttest was tested

by using a paired-samples *t*-test (sample sizes of 30) and a Wilcoxon Signed Ranks Test (sample sizes smaller than 30). For testing the difference between learners and native speakers, a Mann-Whitney *U* was used. Effect sizes were calculated by using Cohen's *d*.

## Results

### *Acquisition of tonal features*

Seven out of nine L2 speakers produced native-like (= Swedish-like) utterance contours after the course. Two L2 speakers were not judged as more native-like, even though their pronunciation underwent tonal changes. In the following, these two learner groups are described. The names of the learners have been changed.

All seven learners with native-like utterance contours in the posttest also acquired tone accent. Acquisition of tone accent seemed to facilitate three other changes in utterance intonation: (a) the primary stress gets a clear tonal correlate, (b) the utterance contour is non-descending due to a higher *f*<sub>0</sub> in primary stress than in the beginning of the utterance, and (c) the *f*<sub>0</sub> fall after the latter *f*<sub>0</sub> peak in primary stress is deeper than without tone accent. Thus, there is evidence here of a connection between acquisition of different tonal features; more specifically, between acquisition of tone accent and other features in utterance intonation (Figures 4–6). Effects of learned tone accent on utterance intonation are logical but not self-evident. Utterance contour could remain descending even if the tone accent was learned, if the speaker started the utterance with a pitch high enough. Still, in our material, the connection between learned tone accent and other tonal changes towards native-likeness seems sequential, not cyclic: tone accent is followed by other tonal changes, not vice versa (cf. Figure 7).

The tonal development and the connection between different features are most obvious with the L2 speaker we call Hilma (Figure 4). This speaker is native-like in the posttest concerning all tonal parameters examined in our study: overall contour, mean *f*<sub>0</sub>, *f*<sub>0</sub> range, and difference between unstressed and stressed part of the utterance.

**(FIGURE 4)**

**(FIGURE 5)**

**(FIGURE 6)**

As can be seen from Figures 4-6, acquisition of tone accent significantly affects the shape of the utterance contour. In addition, it also facilitates acquisition of other tonal features: higher

f0 standard deviation (SD), higher mean f0, and greater pitch range. All seven learners with tone accent and native-like utterance contour in the posttest, also had higher mean f0, greater SD, wider pitch range (both in utterance and in primary-stressed word), and greater f0 difference between the unstressed syllable preceding primary stress and the peak in primary stress (see Appendix). Especially among female learners, range in utterance was much greater and thus more native-like in the posttest. Learners' mean f0 in their Finnish and Swedish before training were lower than in their Swedish after training; e.g. 205 Hz for female speakers in both Finnish and in Swedish before training, 221 Hz in Swedish after training (see Appendix). This difference in mean f0 before and after training was statistically significant ( $p=0.000$ ), and the effect size was large ( $d=0.91$ ). Further, the difference in range in primary stress before and after training was statistically significant ( $p=0.020$ ), as was the case in pitch range in utterance before and after training ( $p=0.016$ ). In both cases, effect size was found to be equally large ( $d=0.90$ ). As for SD, the difference between pre- and posttest was not statistically significant ( $p=0.170$ ). Mean f0, SD and pitch range were all somewhat higher/greater in native speakers' averages than in learners' averages (only female speakers compared; see last two rows in Appendix). When compared to native speakers, there was a statistically significant difference to the female learners' mean f0 pretest values ( $p=0.000$ ), with a large effect size ( $d=1.28$ ), but not to posttest values ( $p=0.082$ ). As for SD, a statistically significant difference was found between the groups both in pretest ( $p=0.001$ ) and posttest ( $p=0.038$ ). Effect size was large as for the pretest ( $d=0.89$ ) and medium as for the posttest ( $d=0.62$ ).

The four best L2 speakers had values very close or identical to native-speaker averages for almost all of the eight tonal measurements in the posttest (see data for Eppu, Julia, Heli, and Hilma in Appendix). For example with Heli, the development towards native-speaker averages was striking: range in utterance and in primary stress, and difference between the preceding unstressed syllable and primary stress were native-like in the posttest (Figure 5 and Appendix).

Those two learners who did not acquire tone accent, did not acquire native-like utterance contour either (cf. Tommi and Tarmo in Figure 7 and Appendix). This finding supports the idea of tone accent being crucial for the acquisition of native-like utterance contour. Further, learners who did not acquire tone accent and native-like utterance contour, had higher mean f0, greater SD, and bigger f0 range in utterance in posttest than in pretest. Although these changes are apparently steps towards a more native-like intonation, they were produced in a manner that differed from native speakers' production. This can be seen e.g. as a steep f0 fall in Tommi's posttest (Figure 7) – an interlanguage feature found neither in the speaker's L1 nor in the target language.

**(FIGURE 7)**

***Syllable duration***

Results of the duration measurements are summarised in Table 1 (speaker averages). The measurements are numbered according to the order in which they were introduced in Analyses above.

**(TABLE 1)**

Overall, many of the learners already produced the primary-stressed word in a temporally native-like manner before training, as shown by measurements 1 and 2. However, all learners made their stressed syllables longer after training, and the difference between pretest and posttest was statistically significant ( $p=0.000$ ) in both measurements with a medium effect size ( $d=0.594$  and  $0.681$  respectively). On average, the stressed syllable was found to be considerably longer in some learners than in native-speakers. When native-speaker values were compared with learner values, no statistically significant differences were found at the .05 level. This applied to both pre- and posttest values ( $p=0.941$  vs.  $p=0.149$  in measurement 1,  $p=0.933$  vs.  $p=0.051$  in measurement 2). The  $p$ -value was very close to significance level in measurement 2 posttest, which we find a possible indication of exaggeration of the syllable length in learners after training.

Measurement 3 compared the duration of the stressed syllable in a word with primary stress to the previous phonetically unstressed syllable. This measure was found to be the most relevant measurement for primary stress production. In the pretest, there was great variation between the learners. Learners' values were typically substantially lower than those of the native speakers, whereas in the posttest the values were much closer to native-speaker values, or even longer. The difference between pretest and posttest was statistically significant ( $p=0.000$ ), with a medium effect size ( $d=0.714$ ). The difference between learners and native speakers was not statistically significant in the pretest ( $p=0.061$ ) or the posttest ( $p=0.650$ ).

Total duration of the utterance (measurement 4) shows an increase in learners' articulation rate from pretest to posttest. The difference between pre- and posttest was statistically significant ( $p=0.002$ ), with a small effect size ( $d=0.446$ ). For this measurement the variation in development was not as straightforward as for the other measurements. Two learners (Tommi and Tarmo) in fact slowed down or maintained approximately the same rate between pretest and posttest. When the learners were compared with the native speakers, no statistically significant difference was found between the groups in the pretest ( $p=0.085$ ) or posttest ( $p=0.513$ ).

The measurements show that the learners learned to mark primary stress with duration, a stress correlate not used in their L1. This includes not only lengthening the stressed syllable in the primary-stressed word, but also shortening and destressing the phonologically stressed (but phonetically unstressed) syllable in the previous word. Learners who developed the least tonally – Tommi and Tarmo – nevertheless made progress in the temporal marking of primary stress just like the learners who developed well on the tonal side. Hence, we cannot

conclude that tonal and temporal marking of primary stress go hand in hand. This conclusion is also supported by the fact that one learner (Minna) made good progress on the tonal side but was the only learner who did not reach native-like values in the syllable duration measurements even after the training (see Table 1). There is, however, one aspect that distinguishes the weak tonal learners from the others also in the temporal measurements: articulation rate. Whereas the learners who made progress tonally also speeded up their articulation, Tommi's and Tarmo's development was minimal on the tonal side and they also slowed down their articulation or maintained approximately the same rate.

### **Discussion and concluding remarks**

This paper presented a study on L1 Finnish-speakers' ( $n=9$ ) development in Swedish prosody after a six-week training period. Learners' pronunciation of isolated utterances (declarative sentences) was compared with native speakers' ( $n=4$ ) pronunciation. The results show that learners already had quite a good temporal pattern before training. Still, they learned to prolong the primary-stressed syllable in proportion to the preceding unstressed syllable, indicating that they were able to reset the duration parameter in their L1 Finnish and use it as a stress correlate. With regards to intonation, most of the learners were far from native-like patterns before training, but they were able to develop substantially after training. Seven out of nine learners produced fairly stable tone accent 2 in the posttest. These learners also acquired a native-like utterance contour. Seven learners had a higher mean  $f_0$  in the posttest than pretest (statistically significant difference to native speakers in pretest but not in posttest), and 8 learners had a greater pitch range at posttest both in utterance and in primary-stressed word. It is rather encouraging from the perspective of formal instruction that such a clear improvement in intonation occurred during a relatively short training period. Previous studies have shown that tone accents are a realistic but difficult learning goal for adult learners of Swedish (e.g. Author A). The new information gained in the present study includes the fact that other, more phonetic aspects of intonation can also be learned, and that formal instruction seems to be quite effective in achieving these learning goals.

The results show that learners acquired both phonological and phonetic aspects of the target language. The learning of a phonological aspect – tone accent 2 – also seemed to support the acquisition of phonetic-prosodic features in the target language: native-like utterance contour, higher  $f_0$  SD, higher mean  $f_0$ , and greater pitch range. This demonstrates a connection between the development of different features, a possibility that has been overlooked in the prevailing theoretical models of L2 pronunciation learning. Even though these latter features are not necessary for communicative, well-functioning pronunciation, they should not be underestimated as learning goals for advanced L2 speakers, because the cumulative effect of even slightly inappropriate tonal traits can be noticeable for L1 listeners (Mennen 2007: 54). The connection between the acquisition of tone accent and other, more phonetic aspects of L2 intonation shows that acquisition of one aspect of L2 prosody can facilitate acquisition of other aspects. Thus, we suggest that future studies should not concentrate on L2 prosody as a phenomenon consisting of separate, non-related entities but rather on acquisition of a system where development of one trait supports development of other traits. The connections

between the learning of different prosodic features should also be given attention in theoretical models of L2 pronunciation.

The temporal pattern in Swedish seemed to be a somewhat easier learning goal for these L1 Finnish learners than the tonal one. This may be due to the participants' high proficiency level in general, or the fact that they were used to altering sound durations in their L1. Hence, this finding may not be generalizable to learners with L1s other than Finnish. To further explore the influence of general language proficiency vs. L1 on the temporal pattern, similar research to ours should be conducted on learners with L1s other than Finnish. The results of the present study do not indicate a connection between the learning of tonal and temporal features. However, the weakest learners on the tonal side were distinguished from the other learners also in that they did not increase their articulation rate. Hence, there may potentially be a link of some kind between tonal and temporal development. Here, however, the explanation may just as well lie in individual differences in learning ability as in phonetically-based factors.

The results in Abelin and Thorén (2015) show that syllable length is more important than tone accent for correct word recognition in Swedish. We acknowledge this finding, but wish to point out that in the teaching of pronunciation to advanced learners as in this study, the phonological aspects of language are not the only learning goals. Whether or not to teach tone accents is of course dependent on the learner group and its goals. Naturally, we do not suggest tone accents in Swedish are crucial – or even particularly important – when teaching e.g. immigrants with low target language proficiency and still an evident need to work on intelligibility. However, when dealing with learners who wish to attain (near) native-like pronunciation, there may be a place for the teaching of tone accents considering their influence on acquisition of other tonal aspects.

## Notes

1. Finland-Swedish is a distinct variety of Swedish spoken in Finland. It is influenced by Finnish especially in regards to phonetics, and differs in many prosodic respects from Sweden-Swedish (e.g. Helgason et al., 2013).
2. We refer here to the previous stressed syllable at word level: *vill, köp-, kan, ska, gan-, köp-* in sentences 1–6, respectively. Notice however, that stress relations are neutralised outside the primary-stressed word, i.e. at sentence level, and the syllable is phonetically more or less unstressed.

## References

- Abelin, Å. and B. Thorén. 2015. What affects recognition most – wrong word stress or wrong word accent? In *Working Papers 55. Proceedings from Fonetik 2015 Lund, June 8–10, 2015* ed. M. Svensson Lundmark, G. Ambrazaitis and J. van de Weijer, 7–10. Lund: Lund University.
- Bannert, R. 2004. *På väg mot svenskt uttal*. Lund: Studentlitteratur.
- Best, C.T. and M.D. Tyler 2007. Nonnative and second-language speech perception: commonalities and complementarities. In *Language Experience and Second Language Speech Learning: In*

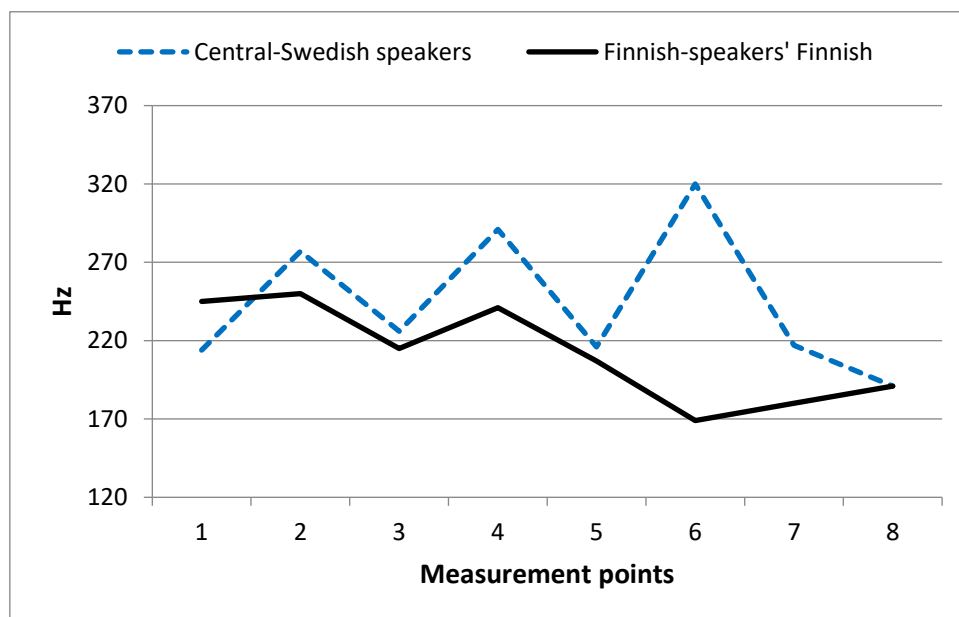
- Honor of James Emil Flege*, ed. O.S. Bohn and M.J. Munro, 13–34. Philadelphia, PA: John Benjamins.
- Boersma, P. and D. Weenink. 2015. PRAAT: Doing phonetics by computer. <http://www.praat.org>  
Amsterdam: Phonetic Sciences, University of Amsterdam.
- Bruce, G. 1977. Swedish word accents in sentence perspective. PhD thesis. Lund:  
LiberLäromedel/Gleerup.
- Bruce, G. 2010. *Vår fonetiska geografi. Om svenskans accenter, melodi och uttal*. Lund:  
Studentlitteratur.
- Bruce, G. 2012. *Allmän och svensk prosodi*. Lund: Studentlitteratur.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence  
Erlbaum.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning,  
Teaching, Assessment*. Cambridge: Cambridge University Press
- Cunningham-Andersson, U. and O. Engstrand. 1989. Perceived strength and identity of foreign accent  
in *Perilus X. Experiments in Speech Processes*, ed. O. Engstrand, M. Dufberg and C. Kylander,  
65–86. Stockholm: Swedish Institute of Linguistics, University of Stockholm.
- Derwing, T.M., M.J. Munro and G. Wiebe. 1998. Evidence in favor of a broad framework for  
pronunciation instruction. *Language Learning* 48, no.3: 393–410.
- Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language learning* 27, no.2:  
315–330.
- Flege, J.E. 1995. Second language speech learning: theory, findings, and problems. In *Speech  
Perception and Linguistic Experience: Issues in Cross-language Research* ed. W. Strange, 233–  
76. Timonium, MD: York Press.
- Gut, U. 2007. Learner corpora in second language prosody research and teaching. In *Non-native  
Prosody: Phonetic Descriptions and Teaching Practice*, ed. J. Trouvain and U. Gut, 145–167.  
Berlin: Mouton De Gruyter.
- Hackman, D.J. 1976. Prosodic deviance in a Finnish foreigner's Swedish speech. In *Department of  
Phonetics at Umeå University: Publication 10*, ed. C.C. Elert and T. Wictorsson, 38–43. Umeå,  
Sweden: Umeå University.
- Hattie, J. 2009. *Visible Learning – a Synthesis of over 800 Meta-analyses relating to Achievement*.  
London: Routledge.
- Helgason, P., C. Ringen, and K. Suomi 2013. Swedish quantity: central standard Swedish and Fenno-  
Swedish. *Journal of Phonetics* 41: 534–545.
- Henton, C. 1995. Pitch dynamism in female and male speech. *Language & Communication* 15, no.1:  
43–61.
- Honorof, D.N. and D.H. Whalen. 2005. Perception of pitch location within a speaker's F0 range. *The  
Journal of the Acoustical Society of America* 117: 2193–2200.
- Honorof, D.N. and D.H. Whalen. 2010. Identification of speaker sex from one vowel across a range of  
fundamental frequencies. *The Journal of the Acoustical Society of America* 128: 3095–3104.
- Iivonen, A. 2009. Finnish sentence accent and intonation. In *Phonetics of Russian and Finnish.  
General Description of Phonetic Systems. Experimental Studies on Spontaneous and Read-aloud  
Speech*, ed. V. de Silva and R. Ullakonoja, 67–73. Frankfurt am Main: Peter Lang.
- Kautonen, M. 2017. Finskspråkiga talares intonation av finlandssvenska i påståendeyttranden i fritt  
tal. *Folkmålsstudier* 55: 31–60.
- Laukkanen, A-M., E. Mäki, J. Pukander and I. Anttila 1999. Vertical laryngeal size and the lowest  
tone in the evaluation of the average fundamental frequency (F0) of Finnish speakers.  
*Logopedics, Phoniatrics, Vocology* 24: 170–177.

- Leinonen, K., A.J. Pitkänen and V.V. Vihanta. 1990. Om rytmen i finlandssvenska och sverigesvenska – ett bidrag till forskningen om finlandssvenska. In *Svenskan i Finland 8, Tampere, Finland, 12–13 October 1989*, ed. K. Nikula and A.J. Pitkänen, 49–100. Tampere: University of Tampere.
- Lennes, M. 2002. Praat script calculate\_segment\_durations.praat. Retrieved from <http://www.helsinki.fi/~lennes/praat-scripts>.
- Myrberg, S. 2010. The intonational phonology of Stockholm Swedish. PhD thesis, University of Stockholm.
- Pépiot, E. 2014. Male and female speech: a study of mean f<sub>0</sub>, f<sub>0</sub> range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody 7, May 2014, Dublin, Ireland*. 305–309.
- Riad, T. 2014. *The Phonology of Swedish*. Oxford: Oxford University Press.
- Riad, T. 2015. *Prosodin i svenskans morfologi*. Stockholm: Morfem.
- Suomi, K., J. Toivanen and R. Ylitalo 2008. *Finnish Sound Structure. Phonetics, Phonology, Phonotactics and Prosody*. Oulu, Finland: University of Oulu.
- Thorén, B. 2008. The priority of temporal aspects in L2-Swedish prosody. Studies in perception and production. PhD thesis, University of Stockholm.
- Trautmüller, H. 2005. Conversion between acoustic and auditory units of pitch and calculation of auditory intervals and distances. Retrieved from <http://www2.ling.su.se/staff/hartmut/umrechnung.htm>.
- Trofimovich, P. and W. Baker 2006. Learning second-language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28, no.1: 1–30.
- Tronnier, M., and E. Zetterholm 2013. Appropriate tone accent production in L2-Swedish by L1-speakers of Somali? *Concordia Working Papers in Applied Linguistics (COPAL)* 5: 722–736.
- Tronnier, M. and E. Zetterholm. 2014. Swedish word accent production by L2-speakers with different tonal L1s. *4th International Symposium on Tonal Aspects of Languages (TAL-2014)*, 59–62. Nijmegen, The Netherlands, May 13-16, 2014.
- Ullakonoja, R. 2011. *Da. Eto Vopros! Prosodic Development of Finnish students' Read-aloud Russian during Study in Russia*. Jyväskylä Studies in Humanities 151. Jyväskylä: University of Jyväskylä.
- Vainio, M., D. Aalto, J. Järvikivi and A. Suni. 2006. Quantity and tone in Finnish lexically stressed syllables. In *Proceedings of the Second International Symposium on Tonal Aspects of Languages 2006*, ed. A. Belotel-Grenié and M. Grenié, 121–124. La Rochelle: University of La Rochelle.
- Özçelik, Ö. 2016. The prosodic acquisition path hypothesis: towards explaining variability in L2 acquisition of phonology. *Glossa: A Journal of General Linguistics* 1, no.1. 28. 1–48. DOI: <http://doi.org/10.5334/gjgl.47>

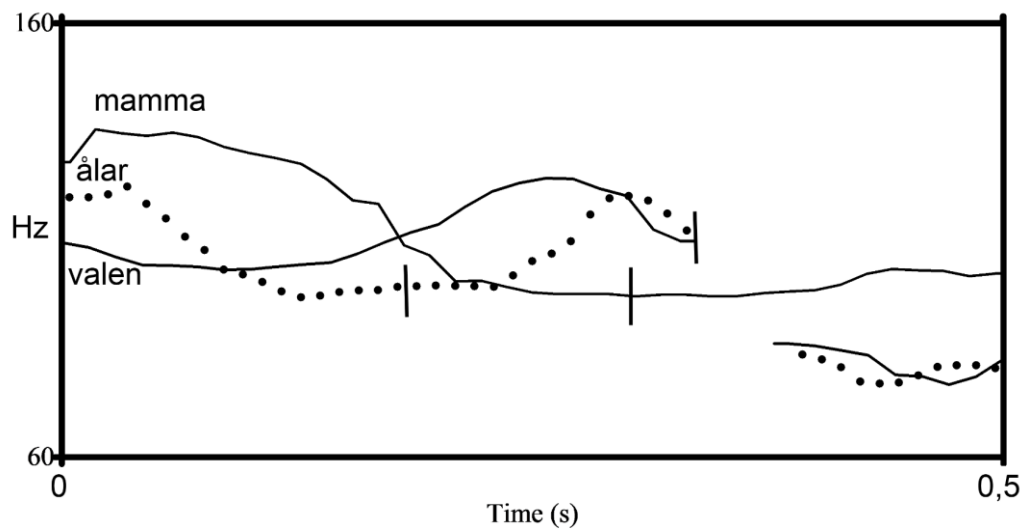
## Appendix



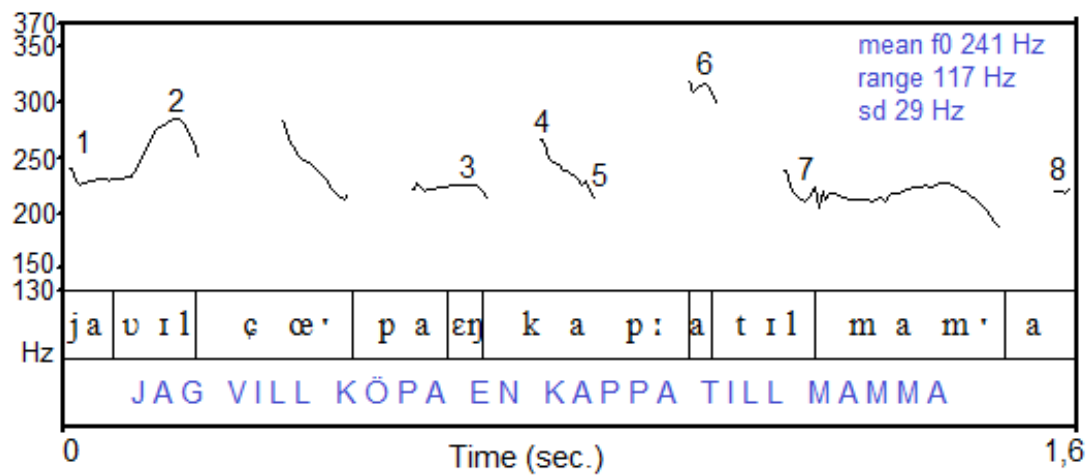
**Figure 1.** Finnish-speakers' (4 subjects' average of 4 utterances; from this study) and Swedish speakers' (4 subjects' average of 6 utterances; from this study) typical intonation in declarative sentences. Utterances had similar length (5–10 syllables) and similar information and prosodic structure, e.g. in Swedish *Vi ska till **stugan** snart* 'We are going to **the cottage** soon' and in Finnish *Hän pesi **pyykkiä** eilen* 'He washed some **clothes** yesterday'. The disyllabic primary-stressed word, marked with bold print in the example utterances, was in non-final position (i.e. rhematic accent; measurement points 4–6 in both languages). The eight measurement points used in this study for f0 are described in detail in Analyses and Figure 3.



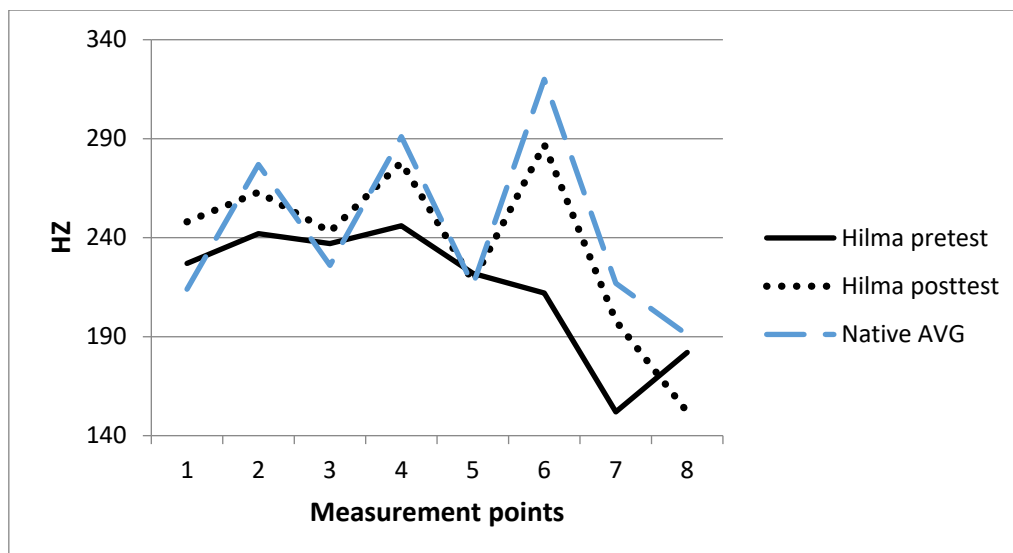
**Figure 2.** Typical f<sub>0</sub> contours in Swedish for tone accent 1 in *valen* ‘the whale’ and tone accent 2 in *ålar* ‘eels’ (the dotted line) and *mamma* ‘mother’. The vertical lines in the contours show the syllable boundaries in the words: *ål–ar*, *val–en* and *mam–ma*. Tone accent 1 (*valen*) has an L\*H contour as focus accent, while tone accent 2 (*ålar*) has an H\*LH contour as focus accent. As word accent, tone accent 2 (*mamma*) is pronounced with an H\*LL contour. The figure is from XXX.



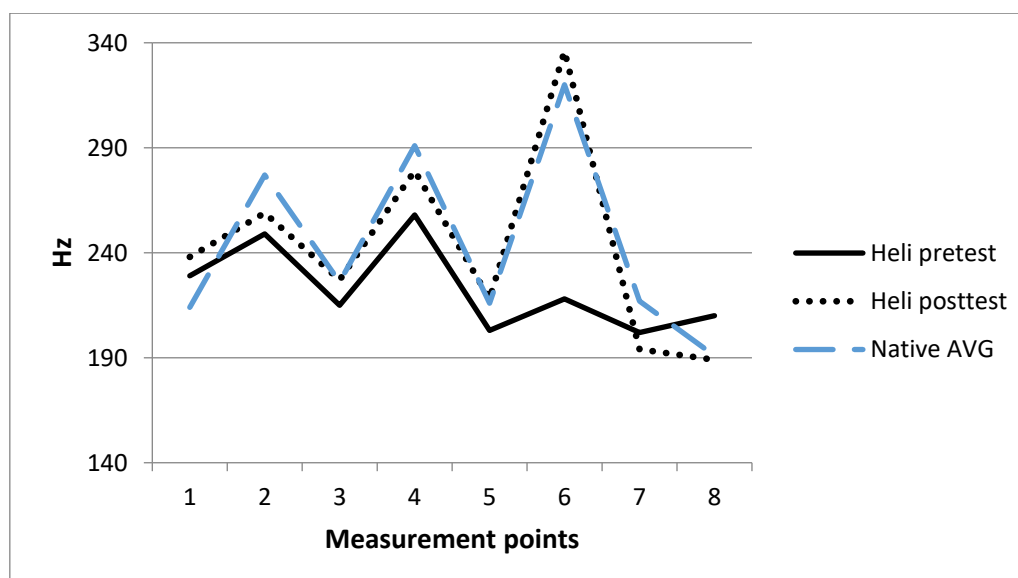
**Figure 3.** L2 speaker Hilma’s pronunciation of *Jag vill köpa en kappa till mamma* ‘I want to buy a **coat** for (my) mother’ in the posttest. The analysis shows the eight measurement points for f0 in this study. The word *kappa* ‘coat’ has tone accent 2 (H\*LH).



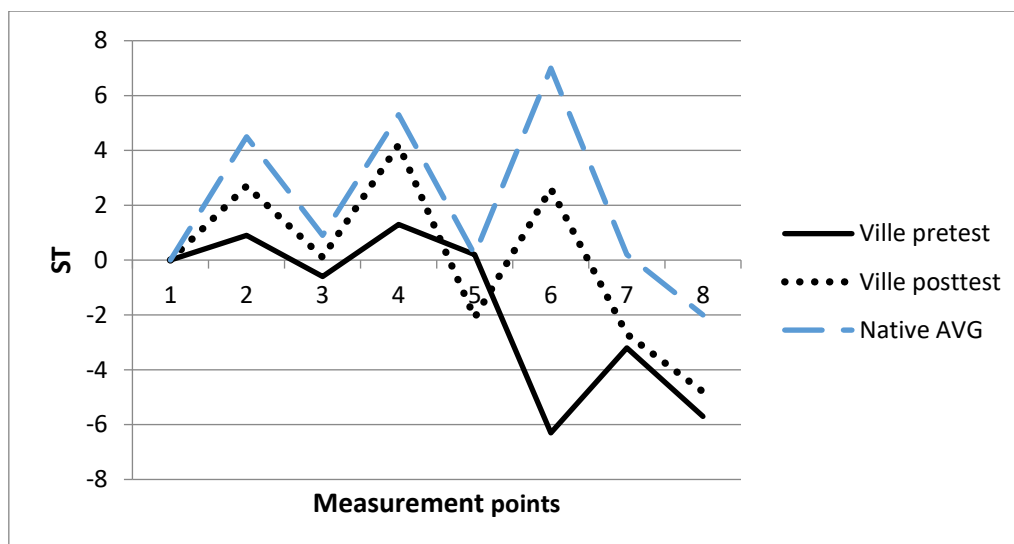
**Figure 4.** Hilma's contour (average of six utterances) in pre- and posttest compared with native-speaker average.



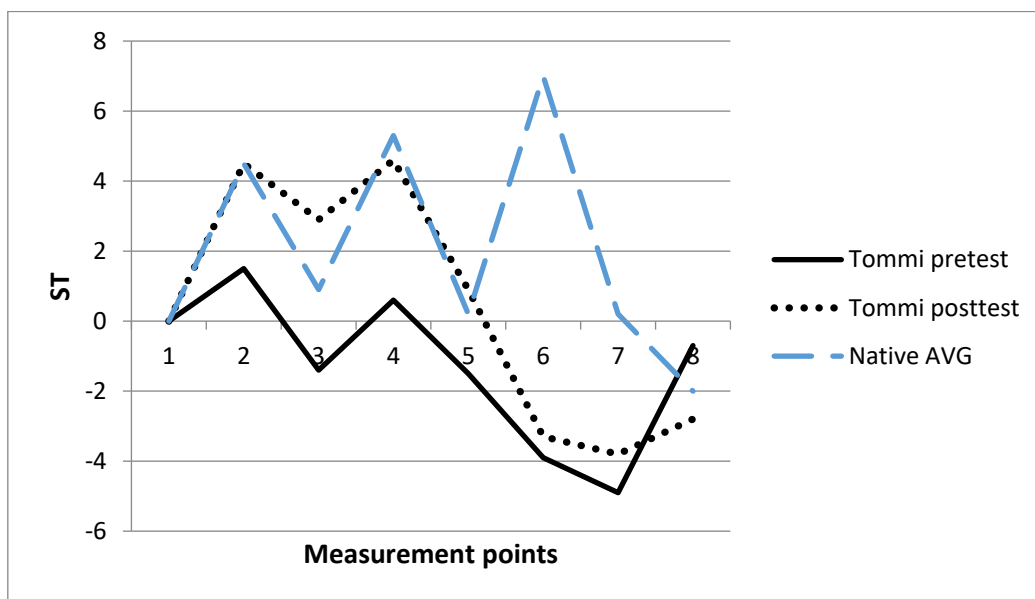
**Figure 5.** Heli's contour (average of six utterances) in pre- and posttest compared with native-speaker average. Scale in Hz.



**Figure 6.** Ville's contour (average of six utterances) in pre- and posttest compared with native-speaker average. To normalize f0-differences depending on gender, values are given in semitones in relation to the reference value 0 in the beginning of utterance (scale +8 to -4 ST). The eight measurement points are given in the middle of the figure.



**Figure 7.** Tommi's contour (average of six utterances) in pre- and posttest compared with native-speaker average. To normalize f0 differences depending on gender, values are given in semitones in relation to the reference value 0 in the beginning of utterance (scale +8 to -4 ST). The eight measurement points are given in the middle of the figure.



**Table 1.** Results of the duration measurements, indicating speaker averages (6 utterances).

speaker	(1) primary stress out of utterance		(2) stressed syllable out of utterance		(3) stressed syllable in relation to previous unstressed syllable		(4) total duration of utterance	
	before	after	before	after	before	after	before	after
Eppu	34.5%	36.3%	23.9%	28.0%	+76%	+139%	1.62 s.	1.43 s.
Minna	28.9%	29.8%	21.0%	21.6%	+39%	+50%	1.53 s.	1.42 s.
Tommi	31.0%	34.2%	22.0%	24.8%	+33%	+90%	1.44 s.	1.60 s.
Ville	30.7%	35.5%	23.3%	25.1%	+67%	+82%	1.52 s.	1.41 s.
Hilma	25.0%	31.6%	17.1%	21.2%	+21%	+62%	1.95 s.	1.51 s.
Heli	30.7%	36.3%	20.8%	27.8%	+43%	+93%	2.08 s.	1.73 s.
Julia	32.1%	35.1%	24.8%	26.9%	+67%	+107%	1.69 s.	1.58 s.
Tarmo	32.7%	35.8%	23.8%	26.4%	+53%	+84%	1.71 s.	1.75 s.
Vilma	29.6%	33.8%	20.8%	24.4%	+42%	+70%	1.75 s.	1.44 s.
Learner average	30.6%	34.3%	21.9%	25.1%	+49%	+86%	1.70 s.	1.54 s.
Native speakers (n=4)	31.0%		22.1%		+75%		1.49 s.	

**Appendix.** Mean f0, standard deviation (SD), range in utterance and in primary-stressed word and difference between the unstressed and the stressed part of the utterance. All values are given in Hertz, and when meaningful also in semitones (in brackets). Individual values for mean f0 and SD for native speakers are given in square brackets.

speaker	mean f0		stand. dev. (SD)		range in utterance (Hz, ST)		range in prim. stress (Hz, ST)		diff. between stress and unstr. (Hz, ST)	
	bef.	after	bef.	after	bef.	after	bef.	after	bef.	after
Eppu	115	110	23	21	65 (9.7)	67 (10.3)	24 (3.0)	36 (4.8)	31 (3.9)	38 (5.1)
Minna	191	215	28	39	56 (5.1)	109 (9.0)	20 (1.7)	45 (3.5)	8 (0.7)	7 (0.5)
Tommi	110	133	13	32	40 (6.4)	66 (8.4)	28 (4.5)	63 (7.9)	13 (2.0)	16 (1.7)
Ville	107	123	12	20	44 (7.7)	64 (9.1)	44 (7.7)	48 (6.3)	13 (1.9)	33 (4.1)
Hilma	211	226	54	53	94 (8.3)	135 (11)	34 (2.6)	69 (4.8)	9 (0.6)	35 (2.3)
Heli	215	238	25	48	56 (4.2)	147 (10)	56 (4.2)	118 (7.5)	43 (3.2)	109 (6.8)
Julia	197	216	49	47	121 (9.6)	144 (10.7)	70 (4.9)	109 (7.4)	73 (5.1)	87 (5.7)
Tarmo	113	119	12	20	36 (5.5)	54 (7.7)	36 (5.5)	46 (6.3)	15 (2.1)	31 (4.0)
Vilma	210	212	34	29	88 (7.1)	81 (6.4)	64 (4.8)	42 (3.0)	32 (2.2)	39 (2.8)
Learner average male (n=5)	111	121	15	23	46 (6.7)	63 (7.9)	33 (4.5)	48 (5.7)	18 (2.4)	30 (3.5)
Learner average female (n=5)	205	221	38	43	83 (6.9)	123 (9.4)	49 (3.6)	77 (5.2)	33 (2.4)	57 (3.7)



Native speakers ( <i>n</i> =4)	233 [251, 242, 234, 205]	53 [62, 57, 47, 47]	134 (9.5)	107 (7.0)	97 (6.4)
-----------------------------------	--------------------------	---------------------	-----------	-----------	----------