AUTHOR
Antti Saloranta, Leena Maria Heikkola

TITLE Acquisition of non-native vowel duration contrasts through classroom education: Perception and production affected differently

YEAR 2022

DOI
https://doi.org/10.1075/jslp.20040.sal

VERSION
Final Draft

CITATION
Saloranta, A., Heikkola, L.M., 2022. Acquisition of non-native vowel duration contrasts through classroom education: Perception and production affected differently: Journal of Second Language Pronunciation. https://doi.org/10.1075/jslp.20040.sal

Acquisition of non-native vowel duration contrasts through classroom education: perception and production affected differently

Antti Saloranta[1], Leena Maria Heikkola[2]

1) Phonetics and Learning, Age and Bilingualism Laboratory, Department of Computing, Faculty of Technology, University of Turku, Finland

2) The Department of Psychology and Speech and Language Pathology, The Faculty of Arts, Psychology and Theology, Åbo Akademi University

Abstract

In quantity languages, the durations of segments affect the meanings of words. This can present problems for second language (L2) learners who do not already have this feature in their native language. This study examines the effects of an intensive, four-week language course with a communicative focus on the perception and production of non-native vowel duration contrasts. A total of 68 students of Finnish, divided into speakers of quantity or non-quantity languages, took part in identification and production tests before and after taking part in the course. The course produced a significant improvement on identification, but not production. Furthermore, a slight advantage was found for speakers of quantity languages in the identification task. Comparison to native control groups revealed significant differences between groups in both tasks. The results are discussed in relation to the interaction of perception and production, L2 learning models and relevance to L2 teaching.

Keywords: L2 acquisition; L2 teaching; classroom education; quantity language; vowel duration; phonological length

1 Introduction

All languages exhibit some variation in the duration of individual speech segments, but the importance of its role varies greatly. Duration can be affected by speech rate, stress, and phonetic context, for example, and in these cases its variation is typically a secondary feature. In languages such as Finnish and Japanese (Isei-Jaakkola, 2004), however, variation of segment duration affects the meanings of words, so that the Finnish word *tuli* (“fire”) becomes *tuuli* (“wind”) when the /u/ vowel is lengthened. In these languages, duration is a phonological feature, and they are typically known as *quantity languages*. Both vowels and consonants can exhibit duration changes, which varies between different languages. In Estonian, for example, vowel duration is distinctive only in stressed syllables, whereas in Japanese and Finnish it can occur in any syllable, regardless of stress (Meister, Nemoto & Meister, 2015).

Because the purpose of this study is to examine the learning of the perception and production of Finnish vowel duration contrasts, we will present the quantity system of Finnish in more detail.Finnish has eight vowels, typically represented with the IPA symbols /i/, /e/, /ʏ/, /ø/, /æ/, /ɑ/, /o/ and /u/, and thirteen consonants /p/, /t/, /d/, /k/, /s/, /m/, /n/, /ŋ/, /l/, /r/, /ʋ/, /j/ and /h/. Loanwords may also contain /f/, /b/, /g/ and /ʃ/. In Finnish, the quantity system is extensive, in that both vowels and consonants have contrasting durations, and the duration systems are independent of each other and stress (Suomi, Toivanen & Ylitalo, 2008, p. 39). On average, long vowels in the first syllable of Finnish words are 2.2–2.4 (Wiik, 1965, p. 60) or 2.2 (Lehtonen, 1970, p. 89) times longer than short ones in the same position. Vowel durations in following syllables are somewhat modified by the word structure: the relative duration of both short and long vowels changes depending on the phonological complexity of the word. For example, in disyllabic words containing both a long and a short vowel (i.e. CV:CV or CVCV:) the absolute duration of the short vowel is on average shorter than it would be in a word where

both vowels are short. Suomi et al. (2008) identified four different grades of short vowels and three grades of long vowels when grouping vowels based on their position in various morphological structures. None of these grades, however, have been found to be contrastive in Finnish. Vowel quality also changes somewhat depending on the duration of the vowel, with long vowels exhibiting more extreme formant values than short ones (Wiik, 1965, p. 60). No evidence exists, however, that these spectral differences are perceptually relevant.

Native (L1) speakers of languages with length contrasts have been shown to process duration differently than speakers of non-quantity languages or languages where the quantity system is not quite as extensive: speakers of German, for example, are less sensitive to segment duration than Finnish speakers (Kirmse et al., 2008), and Finnish speakers seem to exhibit a category boundary effect for duration differences, unlike Russian speakers (Ylinen, Shestakova, Alku, & Huotilainen, 2005). Phonological duration can therefore present a major obstacle for L2 learners of quantity languages. McAllister, Flege, and Piske (2002) found that speakers of Estonian, a quantity language, outperformed speakers of English and Spanish in the acquisition of a Swedish quantity contrast, and Tsukada (2012) found that native Japanese speakers clearly outperformed Australian English speakers and advanced English L2 learners of Japanese in the discrimination of Japanese vowel duration contrasts.

Several studies on the perception and production of non-native duration contrasts have been conducted using laboratory training. It typically consists of perceptual methods, such as identification and discrimination training, and many studies (e.g. Hirata, 2004; Hirata, Whitehurst, & Cullings, 2007; Okuno, 2014; Okuno & Hardison, 2016; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008) have reported improvements on the perception of both vowel and consonant duration using perceptual methods. Okuno and Hardison (2016) also

found improvements in production accuracy of vowel quantity differences in pseudowords, based on improved rating scores by native speakers of Japanese. In addition to perceptual methods, production training has also been used for non-native duration contrasts: Saloranta et al. (2017) found improvements in behavioral discrimination accuracy and changes in production for a novel Finnish vowel duration contrast with a multilingual group of participants. Saloranta et al. (2020) saw both psychophysiological and behavioral discrimination improvements in the perception of a trained novel vowel contrast, with some generalization to an untrained contrast. Both of these studies focused on young adult participants with no phonological duration contrasts in their native languages (Spanish, English, Russian, French, Lithuanian and Nepali in the former study and French, Spanish, English, Russian, Lithuanian, Mandarin and Nepali in the latter).

While improvements have been found with laboratory training, similar methods are typically not used in real-world L2 learning situations. Improvements in L2 duration processing, however, have also been found in classroom studies. One study (Hardison & Motohashi-Saigo, 2010) of adult L1 English students of L2 Japanese found that the more advanced the students were, the better they were at identifying consonant duration differences in different acoustic conditions. The students were between 19–22 years old, and had studied Japanese between 1–7 semesters, being taught by native Japanese speakers. The teaching included “communicative activities” with “no specific instruction in or practice involving the perception or production of geminates”, although the students were “made aware of geminates because of their contrastive role in the language” (Hardison & Motohashi-Saigo, 2010, p. 86). Tsukada (2012) showed that Australian English L2 learners of Japanese outperformed monolingual Australian English speakers in discriminating Japanese vowel duration contrasts. The L2 learners had undergone 160–320 hours of Japanese instruction or were enrolled in teacher training at

university to become Japanese teachers. No specific information was provided on the nature of the instruction they had received.

The purpose of this study is to examine the effects of an intensive, four-week language course on the perception and production of Finnish vowel duration contrasts. To this end, three areas of inquiry can be identified. First, improvement in the perception of non-native duration contrasts has been achieved with intensive laboratory training, in some studies fairly rapidly. Laboratory training, however, has several limitations compared to classroom education, the most obvious being the lack of communication situations between learners and native speakers of the language being studied. It is therefore of interest to examine the effects of language learning in the classroom and reflect on differences from and similarities to more specific laboratory training. This type of study can provide some ecological validity that is missing from laboratory-based approaches, as it presents a more realistic learning and language use situation that is not limited to specific features. Laboratory studies provide information about what *can* be learned, whereas studies such as the current one reflect more on what students actually learn in the real world.

Second, it has been shown in previous studies (e.g. McAllister et al., 2002; Meister et al., 2015) that native quantity differences make it easier to perceive and produce non-native length contrasts. On the other hand, Bohn's Desensitization Hypothesis (Bohn, 1995) posits that vowel duration in particular is somewhat salient even when it is not used distinctively in the listeners' native language. In the current study, the participants were split into quantity and non-quantity language speaker groups in order to find out if being a speaker of a quantity language provides an advantage over speakers of non-quantity languages in the perception and production of

vowel duration differences, or if the differences between the groups are balanced out by the desensitization effect.

Third, it is often stated that in order for something to be produced correctly, it must first be perceived correctly, as perceptual representations may be used as the basis for production (e.g. earlier versions of Flege's Speech Learning Model (Flege, 1995)). This would mean that if learning occurs, initial effects should emerge in perception, rather than production, as the correct perceptual representations need to be formed. However, some studies have found that this relationship between the two faculties is not always straightforward (Baese-Berk, 2019) and that improvement in non-native production skills can in some cases even precede improvement in perception (e.g. Sheldon & Strange, 1982). This is also reflected in the newest version of Flege's Speech Learning Model (Flege & Bohn, 2021), that suggests that speech perception and production develop with no specific precedence for either faculty. In this study, predictions are difficult to form due to the different native languages of the participants: it may be that speakers of quantity languages already possess some of the perceptual and production skills necessary for production changes to happen.

Based on the outlined literature and the issues discussed in the previous paragraphs, the following research questions were formed:

1. Does an intensive language course with a communicative focus affect perception and production of non-native vowel contrasts?
2. Does being a speaker of a quantity language offer an advantage?
3. Do perception and production develop in different ways?

2 Methodology

2.1 Participants

A total of 71 students of Finnish as an L2 took part in the study. They had arrived in Finland to take part in intensive language courses and had previously had some exposure to native Finnish speakers outside of Finland, most often through native Finnish language teachers at their home universities. The students had completed 6 to 24 months of Finnish studies (mean 15 months) before arriving in Finland. Three students were excluded from the final groups, as they missed some parts of the experiment due to absence; altogether 68 students participated in both the pre- and posttest in both parts of the study. The participants were divided into two groups, one with all speakers of quantity languages, and one with the rest of the participants. From now on, these will be referred to as the "quantity" (Q) and "non-quantity" (NQ) groups, respectively. Independent samples t-tests did not reveal significant differences between the groups' background factors, but there was a significant difference in identification performance at pretest for all vowels and short vowels (Table 1). For the purposes of this study, a quantity language was defined as a language that has phonological duration contrasts in vowels, consonants or both. The native languages of all participant groups can be seen in Table 2.

Table 1

Participants of the study

		Age (yrs) *M* (*SD*)	Gender, *n* (M/F)	*LoR* in Finland (mos.) *M* (*SD*)	*Lo* Finnish studies *M* (*SD*)	ID error rate, all vowels, pretest	ID error rate, short vowels, pretest	ID error rate, long vowels, pretest
All participants (*N* = 68)		22.1 (3.6)	15/53	1.3 (2.3)	14.9 (7.1)			
Group	Non-quantity language group (*n* = 39)	21.4 (3.5)	13/55	1.2 (1.0)	14.9 (6.7)	0.06 (0.08)	0.05 (0.07)	0.07 (0.13)
	Quantity language group (*n* = 29)	23.1 (3.6)	19/49	1.6 (3.3)	14.8 (7.7)	0.02 (0.04)	0.02 (0.02)	0.03 (0.06)
	Independent samples *t*-test	*t* = -2.0 *df* = 66 *p* = 0.054	*t* = 0.9 *df* = 66 *p* = 0.35	*t* = -0.7 *df* = 66 *p* = 0.46	*t* = 0.5 *df* = 65 *p* = 0.96	*t* = 2.03 *df* = 65 *p* = 0.047	*t* = 2.03 *df* = 65 *p* = 0.047	*t* = 0.1 *df* = 66 *p* = 0.13

Note. The independent samples *t*-tests refer to the two cells immediately above them, e.g. in the Age column the test is comparing the ages of the two participant groups. *M* = mean, *SD* = standard deviation, *LoR* = Length of Residence, *Lo* = Length of.

Table 2

Native languages of the participants

Non-quantity group (NQ)	Russian (10), English (5), Polish (5), French (5), Chinese (4), Spanish (2), Romanian (2), Ukrainian (2), Udmurt & Russian (2), Tatar & Russian (1), Georgian (1)
Quantity group (Q)	German (10), Hungarian (6), Italian (5), Czech (3), Japanese (2), Korean (1), Latvian (1), Estonian & Russian (1)

Note: In the case of strongly bilingual participants, dominant language is listed first.

In addition to the non-native participants, two control groups of 10 monolingual native speakers of Finnish (Group 1: 9 female, mean age 26.1 years; Group 2: 6 female, mean age 22.5 years) were recruited for the identification task and the production task, respectively. All of them were university students of Finnish (studying to become Finnish teachers) or general linguistics and will be referred to as the native Finnish speakers (NF) from now on in their respective tasks. As neither native control group took part in the intensive language course, and therefore were not expected to change over time, they only did the identification test and produced the sentences once, as opposed to the non-native participants of the study. All participants in the study volunteered to take part with no compensation.

The non-native participants of the study took part in two intensive summer courses in Finnish language and culture organized by a Finnish university, paid by the National Agency for Education. The courses lasted approximately four weeks. The courses included 80 hours of instruction and focused on spoken language. Similarly to Hardison and Motohashi-Saigo (2010), the instruction on the courses was functional and communicative and was organized as three workshops: vocabulary and grammar, reading comprehension and interaction. In addition, the participants wrote a written assignment and completed oral presentations in groups.

2.2 Identification task and stimuli

For the identification task, a list of minimal pairs was initially created. All pairs followed a CV(ː)CːV structure and consisted of a short and long member with a short and long first syllable vowel, respectively. The vowels used were /y/, /æ/ and /ø/, which were chosen as they are considered quite difficult for many Finnish learners. /y/ and /ø/ are in roundness opposition to the much more common /i/ and /e/, respectively, and may be particularly problematic to those learners with mostly unrounded front vowels in their native languages, such as English, Spanish

or Russian. The word-medial consonants were the stops /t/, /k/ or /p/, and the initial consonants were /l/, /m/, /n/, /r/, /s/ or /t/. The pairs were formed so that each vowel was inserted into each possible consonant combination, resulting in 74 minimal pairs, differing in the length of the first syllable vowel (for example /*syp*ː*y*/ - /*sy*ː*p*ː*y*/). Pairs containing real Finnish words were then removed in order to eliminate any possible effects caused by word recognition. This resulted in the final 50 pairs that were recorded as the identification stimuli (all these pairs are presented in Appendix A). /æ/ and /ø/ were the vowel in 17 pairs each, and /y/ in 16. The stimuli were recorded by a 33-year-old male native Finnish speaker in a sound attenuated booth with a Beyerdynamic MMX 300 headset connected to a laptop running Audacity (version 2.3.2). The speaker was instructed to say the words naturally, with a neutral intonation. The average amplitude of each stimulus file was normalized to 65 dB, but no other modifications were made.

Each stimulus pair was presented four times throughout the experiment, with both the long and short member of the pairs acting as the identification target twice, resulting in a total of 200 tokens. The stimuli were presented in a pseudorandomized order, so that no members of the same pair appeared consecutively, in 4 blocks of 50 with an interstimulus interval of 3 seconds. They were presented through loudspeakers in a lecture hall to all the participants at once. The task itself consisted of the participants listening to the stimulus blocks and circling the correct answer on a form. The form contained the instruction “Ympyröi se sana jonka kuulet/Circle the word you heard”, followed by a list with the answer options for each sound (e.g. “**syppy** or **syyppy**”). Mean values of the identification stimuli can be seen in Table 3; overall, they are representative of the typical Finnish short-long ratios for vowels, as described in the Introduction section.

Table 3

Mean values of the identification stimuli

	Short words	Long words
Word duration (ms) (SD)	458 (30)	491 (35)
Vowel duration (ms) (SD)	81 (12)	174 (15)
Long/short ratio, word (SD)	1.07 (0.06)	
Long/short ratio, vowel (SD)	2.18 (0.28)	

Note. SD = standard deviation. Short words = stimuli with short vowels. Long words = stimuli with long vowels.

2.3 Production task

For the production task, each of the participants read 60 declarative sentences out loud. The sentences were simple, semantically meaningful three-word statements, which each contained either a short or a long exemplar of either /y/, /æ/ or /ø/ in an initial syllable, followed by a stop consonant. All sentences with English translations can be found in Appendix B. The participants were instructed to read each sentence out loud at their own preferred pace in a natural voice. The recordings were performed in an acoustically treated studio using a Røde Podcasting microphone connected to PCs running Audacity (2.3.2), and a small conference room using Zoom H2n microphone with a memory card. Due to technical difficulties, the long member of the sentence pair Y1YY1 was lost for three of the Finnish native controls and this pair could therefore not be included in the analysis for these participants.

2.4 Analysis

For the identification task, the goal was to investigate the perception of duration (short vs. long) in three vowel contexts (/y/, /æ/ or /ø/) at pretest and posttest. We used the following variables in the statistical analysis: all vowels, long vowels, short vowels, /y/, /yː/, /æ/, /æː/, /ø/and /øː/, both for pretest and posttest. These variables reflect the percentage of incorrect

identifications. Our aim was to investigate possible differences between the perception of long and short vowels, but also to examine whether some vowels would be more difficult to perceive as short or long.

Four-way ANOVAs (Group(2) x Time(2) x Vowel type(3) x Vowel length(2)) were carried out to investigate the possible differences in identification accuracy between pre- and posttest, the participant groups, the vowel contexts, and the duration contrasts. Further analyses, when justified by the four-way ANOVAs, were carried out as repeated measures ANOVAs (Group(2) x Time(2)) to look at possible differences between identification accuracy between pre and posttest, and between Q and NQ groups. In addition, possible differences between the NQ, Q and NF groups were investigated by one-way ANOVAs (Group(3)) and Bonferroni corrected post hoc tests, separately for the pretest and posttest, as the NF group only performed the identification task once. IBM SPSS Statistics 25 was used for all statistical analyses.

For the production task, 12 sentences were selected for analysis, four for each vowel (Table 4). Half of the sentences contained short examples of the vowels, and the other half long ones. The sentences were matched into two short-long pairs for each vowel, based on phonetic contexts. The context was matched as closely as possible between the long and short examples; however, full minimal pairs could mostly not be achieved as they do not exist in the Finnish lexicon.

Table 4

Sentences selected for analysis of the production task.

	Short sentences		**Long sentences**
Y1	Mummo **kykkii** pellolla	YY1	Te pelaatte **kyykkää**
Y2	**Typpi** on alkuaine	YY2	**Tyyppi** tulee huomenna
Ä1	Nainen heittää **läppää**	ÄÄ1	Mies **lääppii** koiraa
Ä2	Edessä on **mäki**	ÄÄ2	Lammas **määkii** pellolla
Ö1	Sinä olet **pölö**	ÖÖ1	**Pöö**, mörkö tulee!
Ö2	Jäätelö on **tötterössä**	ÖÖ2	Taksi **tööttää** kovaa

Note: Sentences on the same row form a sentence pair and are compared against each other. The words in which the vowel was analyzed in each sentence are in bold.

The duration of each target vowel was measured using Praat (version 6.0.49), and these durations were then used to calculate long/short ratios for each sentence pair by dividing the duration of the long vowel by the duration of the short one. This was done in order to normalize differences caused by different speaking rates between the participants that would have made direct comparisons of absolute durations unreliable. A ratio of 1 indicates equal duration for the short and long vowels. Values under 1 mean that the long one was produced shorter than the short one, and over 1 means the long one was produced longer. Finally, vowel type variables (/y/, /æ/ and /ø/) were formed by averaging the ratios for the two sentence pairs for each vowel.

The ratios calculated separately for each vowel type /y/, /æ/ and /ø/ for pretest and posttest were used for statistical analysis. Three-way ANOVAs (Group(2) x Time(2) x Vowel type(3)) were used to investigate the possible differences between production in pre- and posttest, the Q and NQ groups, and the 3 vowel types. In order to get a more detailed picture of the data, we also

performed repeated measures ANOVAs (Group(2) x (Time(2)) for the 3 vowel types. In addition, the participants' ratios for /y/, /æ/ or /ø/ were compared to a group of native Finnish speakers' ratios of the same vowels in a one-way ANOVA (Group(3)) and Bonferroni corrected post hoc tests, separately for pretest and posttest.

3 Results

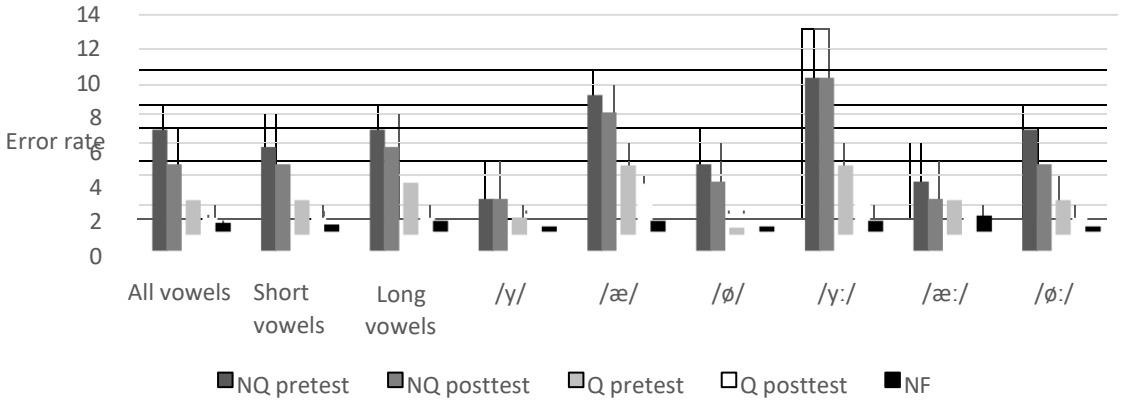
3.1 Identification task

The mean identification error rates are presented in Figure 1. Examination of the rates shows that overall, the participants performed quite well at the pretest stage, with no variable exhibiting error rates higher than 0.1, i.e. 10%. The lowest rates of errors are consistently seen for the NF group, and the highest for the NQ group in the perception of /yː/.

To investigate possible main effects and interactions, a four-way ANOVA was calculated (Group(2) x Time (2) x Vowel type(3) x Vowel length(2)), where Time represented pre- and posttest measurements, Group consisted of the two non-native groups, Vowel type of the three vowels /y/, /æ/ or /ø/, and Vowel length of short and long durations. The following main effects were statistically significant: Time ($F(1, 395) = 14.8$, $p < 0.001$, $\eta_{p^2} = 0.04$), and Group ($F(1, 395) = 16.4$, $p < 0.001$, $\eta_{p^2} = 0.04$), suggesting that there was a difference in overall performance over time, and that the groups' performances differed overall. Only one interaction was statistically significant, namely Vowel type * Vowel length ($F(2, 395) = 7.1$, $p = 0.001$, $\eta_{p^2} = 0.04$), suggesting that different vowels had different accuracies for short and long durations. To gain a more detailed picture of the data, repeated measures ANOVAs were then performed for each of the variables (Figure 1). All of the analyses followed the general structure of Group(2) X Time(2). All findings from these analyses are presented in Table 5.

Figure 1.

Mean identification error rate, percentages



Note. The error bars indicate standard error.

Table 5

Group comparisons of the identification error rates: Repeated measures ANOVAs (Time(2) x Group(2))

	Sentence	Main effect & interaction	Details
Quantity groups	All vowels	Time	$F(1,65) = 9.1$, ***p*** = **0.004**, $\eta_p^2 = 0.1$
		Group	$F(1,65) = 7.48$, ***p*** = **0.008**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,65) = 0.08$, $p = 0.79$, $\eta_p^2 = 0.001$
	Short vowels	Time	$F(1,65) = 2.0$, $p = 0.16$, $\eta_p^2 = 0.03$
		Group	$F(1,65) = 8.2$, ***p*** = **0.006**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,65) = 0.02$, $p = 0.90$, $\eta_p^2 < 0.001$
	Long vowels	Time	$F(1,66) = 6.7$, ***p*** = **0.01**, $\eta_p^2 = 0.1$
		Group	$F(1,66) = 14.0$, ***p*** = **0.049**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,66) = 0.06$, $p = 0.80$, $\eta_p^2 = 0.001$
	/y/	Time	$F(1,66) = 1.4$, $p = 0.24$, $\eta_p^2 = 0.02$
		Group	$F(1,66) = 4.3$, ***p*** = **0.042**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,66) = 0.44$, $p = 0.51$, $\eta_p^2 = 0.01$
	/æ/	Time	$F(1,66) = 1.8$, $p = 0.19$, $\eta_p^2 = 0.001$
		Group	$F(1,66) = 0.04$, $p = 0.09$, $\eta_p^2 = 0.001$
		Group * Time	$F(1,66) = 0.4$, $p = 0.85$, $\eta_p^2 = 0.001$
	/ø/	Time	$F(1,65) = 1.3$, $p = 0.26$, $\eta_p^2 = 0.02$
		Group	$F(1,65) = 9.6$, ***p*** = **0.003**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,65) = 1.3$, $p = 0.26$, $\eta_p^2 = 0.02$
	/yː/	Time	$F(1,66) = 4.1$, ***p*** = **0.048**, $\eta_p^2 = 0.1$
		Group	$F(1,66) = 5.2$, ***p*** = **0.025**, $\eta_p^2 = 0.1$
		Group * Time	$F(1,66) = 0.1$, $p = 0.77$, $\eta_p^2 = 0.001$
	/æː/	Time	$F(1,66) = 4.1$, ***p*** = **0.047**, $\eta_p^2 = 0.1$
		Group	$F(1,66) = 1.8$, $p = 0.18$, $\eta_p^2 = 0.03$
		Group * Time	$F(1,66) = 0.2$, $p = 0.70$, $\eta_p^2 = 0.002$
	/øː/	Time	$F(1,66) = 4.1$, ***p*** = **0.046**, $\eta_p^2 = 0.1$
		Group	$F(1,66) = 3.2$, $p = 0.08$, $\eta_p^2 = 0.1$
		Group * Time	$F(1,66) = 0.4$, $p = 0.54$, $\eta_p^2 = 0.01$

Note. p-values in bold indicate statistically significant findings.

The statistical analyses revealed statistically significant improvement in identification accuracy over time, as shown by the significant main effect of Time for All vowels. Further analysis showed that this effect was only true for long vowels, and analysis of individual vowels found that all long vowel stimuli showed improvement over time. This suggests that the non-native participants were able to improve their identification skills of long vowel duration during the intensive language course. Significant group level differences, on the other hand, emerged for both short and long vowels, and analysis of individual vowels showed that they occurred in /y/,

/ø/ and /yː/. In each of these cases the Q group outperformed the NQ group (Figure 1). This suggests that for perception of vowel duration, native speakers of a quantity language have an advantage over those that do not speak such languages.

In order to examine how native-like the participant groups' performances were, one-way ANOVAs were performed separately for the pre- and posttest identification scores between the non-native groups and the native Finnish control. All ANOVAs are presented in Table 6, and post hocs for significant findings in Table 7.

Table 6

Statistical analysis of identification error rates between all groups (one-way ANOVA)

	Variable	Main effect	Details
NQ, Q and NF	All vowels	Pretest: Group	$F(2,76) = 5.4$, $\mathbf{p = 0.006}$, $\eta_p^2 = 0.2$
		Posttest: Group	$F(2,77) = 6.0$, $\mathbf{p = 0.004}$, $\eta_p^2 = 0.14$
	Short vowels	Pretest: Group	$F(2,76) = 6.4$, $\mathbf{p = 0.003}$, $\eta_p^2 = 0.2$
		Posttest: Group	$F(2,77) = 5.5$, $\mathbf{p = 0.006}$, $\eta_p^2 = 0.13$
	Long vowels	Pretest: Group	$F(2,77) = 2.8$, $p = 0.068$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 3.4$, $\mathbf{p = 0.039}$, $\eta_p^2 = 0.1$
	/y/	Pretest: Group	$F(2,77) = 2.5$, $p = 0.09$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 3.0$, $p = 0.055$, $\eta_p^2 = 0.1$
	/æ/	Pretest: Group	$F(2,77) = 5.5$, $\mathbf{p = 0.006}$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 5.2$, $\mathbf{p = 0.008}$, $\eta_p^2 = 0.1$
	/ø/	Pretest: Group	$F(2,76) = 6.4$, $\mathbf{p = 0.003}$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 5.5$, $\mathbf{p = 0.006}$, $\eta_p^2 = 0.13$
	/yː/	Pretest: Group	$F(2,77) = 3.6$, $\mathbf{p = 0.03}$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 4.2$, $\mathbf{p = 0.019}$, $\eta_p^2 = 0.1$
	/æː/	Pretest: Group	$F(2,77) = 1.0$, $p = 0.38$, $\eta_p^2 = 0.03$
		Posttest: Group	$F(2,77) = 1.7$, $p = 0.18$, $\eta_p^2 = 0.04$
	/øː/	Pretest: Group	$F(2,77) = 3.2$, $p = 0.08$, $\eta_p^2 = 0.1$
		Posttest: Group	$F(2,77) = 2.2$, $p = 0.11$, $\eta_p^2 = 0.1$

Note. NQ = non-quantity group. Q = quantity group. NF = native Finnish control group. *p*-values in bold indicate statistically significant findings.

Table 7

Post hoc analyses of identification error rates, pretest and posttest

	Variable	Post hoc tests
NQ, Q and NF	**/æ/**	Pretest: NQ > Q p = 0.056. NQ > NF p = **0.016**. Q > NF p = 0.737
		Posttest: NQ > Q p = **0.033**. NQ > NF p = **0.037**. Q > NF p = 1.00
	/ø/	Pretest: NQ > Q ***p* = 0.005**. NQ > NF p = 0.066. Q > NF p = 1.00
		Posttest: NQ > Q p = **0.010**. NQ > NF p = 0.102. Q > NF p = 1.00
	/yː/	Posttest: NQ > Q ***p* = 0.037**. NQ > NF p = 0.132. Q > NF p = 1.00

Note. Post hoc analyses were conducted in the instances where a significant main effect of Group was found in the one-way ANOVA for individual vowel analysis (Table 6). Furthermore, entirely non-significant post hoc findings have been omitted. Statistically significant findings are in bold. NQ = non-quantity group. Q = quantity group. NF = native Finnish control group. The < and > symbols indicate which group had the larger error rate in each comparison.

These analyses revealed that the Q group did not differ significantly from the NF group in any of the tested variables, either at pretest or at posttest. Conversely, at pretest the NQ group's performance was less accurate than the NF group's in the identification of /æ/. This difference remained at the end of the course, suggesting that the NQ group did not improve during the course. In addition, new significant differences also emerged between the Q and NQ groups at posttest in the identification of /æ/ and /yː/, likely caused by the improved performance of the Q group, as in both of these cases the Q group was more accurate than the NQ group.

Overall, the identification tasks revealed that perception of non-native vowel duration contrasts improved during the intensive language courses. Improvement in identification accuracy was found for all three vowels after completing the course, and differences emerged between the groups, with the Q group outperforming the NQ group in the identification of /y/, /ø/ and /yː/. Comparison with the NF group is even more telling of the advantage of speaking a quantity language, as there were no statistically significant differences in performance between the Q group and native Finnish speakers.

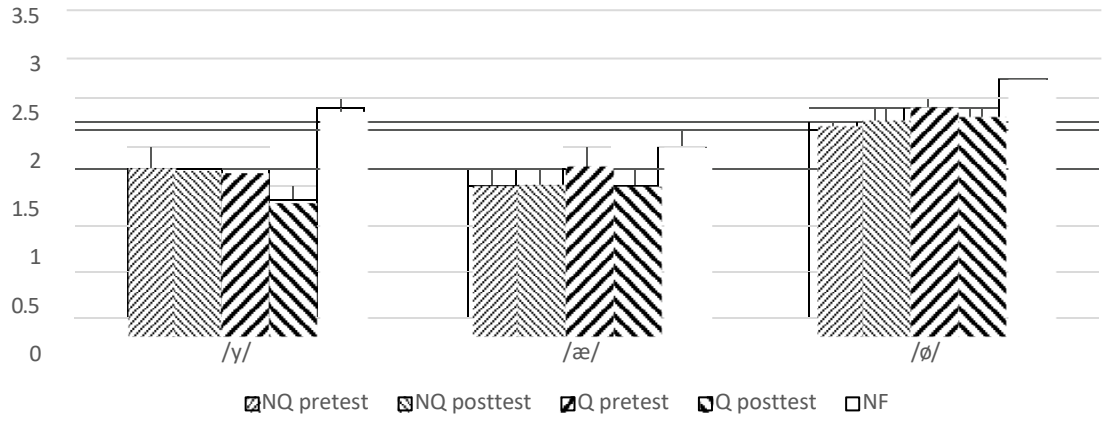
3.2 Production task

The long/short ratios for the different production task sentences are presented in Figure 2. First, a three-way ANOVA (Group(2) x Time(2) x Vowel Type(3)) was performed. Statistically significant main effects were found for Vowel Type ($F(2, 348) = 5.5$, $p = 0.004$, $\eta_{p^2} = 0.03$). Significant interactions were found for Time * Vowel Type ($F(2, 348) = 42.5$, $p < 0.001$, $\eta_{p^2} = 0.2$) and Group * Time * Vowel Type ($F(2, 348) = 3.6$, $p = 0.003$, $\eta_{p^2} = 0.02$). In the post hoc tests, a statistically significant difference was found between /y/ and /æ/ (p = 0.006) and between /y/ and /ø/ (p = 0.003).

More detailed analysis investigating each vowel type individually was done by carrying out repeated measures ANOVAs, comparing the non-native groups with a Group(2) X Time(2) structure. Time represented pre- and posttest measurements, and Group always consisted of the two non-native groups. Each vowel type was analyzed separately. Findings from the analyses are presented in Table 8.

Figure 2.

Mean production ratios



Note. The error bars indicate standard error.

Table 8

Repeated measures ANOVAs of the production ratios with the quantity groups

	Vowel type	Main effect & interaction	Details
Quantity groups	/y/	Time	$F(1, 63) = 5.5$, **$p = 0.02$**, $\eta_p^2 = 0.08$
		Group	$F(1, 63) = 8.9$, $p = 0.11$, $\eta_p^2 = 0.04$
		Group * Time	$F(1, 63) = 3.7$, $p = 0.06$, $\eta_p^2 = 0.06$
	/æ/	Time	$F(1, 63) = 2.1$, $p = 0.15$, $\eta_p^2 = 0.03$
		Group	$F(1, 63) = 0.64$, $p = 0{,}43$, $\eta_p^2 = 0.01$
		Group * Time	$F(1, 63) = 2.9$, $p = 0.09$, $\eta_p^2 = 0.04$
	/ø/	Time	$F(1, 63) = 0.06$, $p = 0.81$, $\eta_p^2 = 0.001$
		Group	$F(1, 63) = 0.64$, $p = 0{,}43$, $\eta_p^2 = 0.01$
		Group * Time	$F(1, 63) = 1.8$, $p = 0.19$, $\eta_p^2 = 0.03$

Note. p-values in bold indicate statistically significant findings.

The statistical analyses revealed that in the production task, only the production of the vowel /y/ underwent any statistically significant changes during the intensive language course, suggesting that the course was mostly not able to affect the participants' vowel duration production skills. More interestingly, the changes actually saw the groups move away from

native-like production of Y2YY2, as their ratios decreased, meaning that they produced short and long vowels with more similar durations at posttest than at pretest.

The vowel /y/ seemed to stand out from the results, as it was the only one of the vowels where statistically significant effects were observed in the production task. Further analyses were performed in order to find out whether participants with /y/ in their native languages differed from those who did not have it. This was done by dividing the participants into those who have a /y/ phoneme in their native language (n = 26, speakers of German, Hungarian, French, Chinese and Estonian) and to those who do not (n = 42, Russian, English, Italian, Polish, Czech, Spanish, Ukrainian, Japanese, Romanian, Udmurt, Georgian, Korean, Latvian and Tatar). A one-way ANOVA (Group(3)) with Bonferroni corrected post hoc tests was performed separately for the pretest and posttest production ratios for the /y/ vowel. It compared participants with /y/ in their native language to those without and the native Finnish speakers. At pretest, a significant main effect of Group ($F(2,71) = 4.8$, $p < 0.001$) was found, and the post hoc tests revealed a significant difference to the native speakers for both the /y/ group ($p = 0.02$) and the non-/y/ group ($p = 0.01$). For the posttest, a significant main effect of Group ($F(2,71) = 6.1$, $p = 0.004$) was found, with the post hoc tests again a significant difference to the native speakers for both the /y/ group ($p = 0.005$) and the non-/y/ group ($p = 0.003$). These findings suggest that the phoneme status of /y/ in the participants' native language did not affect their performance in the production task, as the /y/ and non-/y/ groups did not differ from each other either at pre- or posttest, only differing from the native speakers. The production values for the /y/ and non-/y/ groups that were used in these analyses can be found in Appendix C.

Overall, the production results for the non-native groups were quite different when compared to the identification task. No significant differences emerged between those that speak

quantity languages and those that do not, suggesting that being a native speaker of a quantity language was not advantageous in the production of vowel duration differences in Finnish contexts. The only change over time was observed in the vowel type /y/, with both non-native groups actually moving away from the native control group.

Table 9

Statistical analysis of production ratios between all groups (one-way ANOVA)

	Vowel type	Main effect	Details
NQ, Q and NF	/y/	Pretest: Group	$F(2,71) = 4.8$, $\mathbf{p = 0.009}$, $\eta_p^2 = 0.12$
		Posttest: Group	$F(2,71) = 9.3$, $\mathbf{p = 0.001}$, $\eta_p^2 = 0.21$
	/æ/	Pretest: Group	$F(2,74) = 2.6$, $p = 0.08$, $\eta_p^2 = 0.07$
		Posttest: Group	$F(2,74) = 3.03$, $p = 0.06$, $\eta_p^2 = 0.08$
	/ø/	Pretest: Group	$F(2,74) = 2.7$, $p = 0.08$, $\eta_p^2 = 0.07$
		Posttest: Group	$F(2,74) = 1.9$, $p = 0.16$, $\eta_p^2 = 0.05$

Note. NQ = non-quantity group. Q = quantity group. NF = native Finnish control group.

Table 10

Post hoc analyses of duration ratios in production, pretest and posttest

	Vowel type	Post hoc tests
NQ, Q and NF	/y/	Pretest: NQ < Q $p = 1.0$. NQ < NF $\mathbf{p = 0.015}$. Q < NF $\mathbf{p = 0.011}$.
		Posttest: NQ > Q $p = 0.07$. NQ < NF $\mathbf{p = 0.014}$. Q <. NF $\mathbf{p < 0.001}$.

Note. Post hoc analyses are provided only in the instances where a significant main effect of Group was found in the one-way ANOVA (Table 9). Statistically significant findings are in bold. NQ = non-quantity group. Q = quantity group. NF = native Finnish control group. The < and > symbols indicate which group had the larger ratios in each comparison.

The results from comparisons between the non-native groups and the native speakers (One-way ANOVAs, Table 9, and post hocs, Table 10) show that the participants did in fact produce native-like long/short ratios for the vowel types /æ/ and /ø/ at pretest and at posttest. Differences from native speakers were found for the vowel type /y/. Both non-native groups significantly differed from the native speakers by producing a smaller difference between the short and long vowels. All significant differences between the non-native groups and the

natives were due to the native speakers producing larger ratios between the short and long vowels than the non-native speakers did. Mostly, however, the learners did not differ from native speakers.

4 Discussion

In this study, we compared the effects of taking part in an intensive language course on the perception and production of non-native vowel duration contraststo earlier results achieved through laboratory training. Taking into account the varied linguistic background of the participants of the course, the following research questions were posed:

1. Does an intensive language course with a communicative focus affect perception and production of non-native vowel contrasts?
2. Does being a speaker of a quantity language offer an advantage?
3. Do perception and production develop in different ways?

For question 1, the answer was somewhat mixed. Overall, the identification results clearly show improvement in the perception of the Finnish vowel duration contrasts. Identification performance improved for all long vowels between pre- and posttest, suggesting that taking part in the language course may have helped the participants to improve their differentiation between long and short vowels, even though they were largely at a native-like level to begin with. For production, however, changes over time were only observed for one of the vowel types, /y/, suggesting that taking part in the course did little to change the way the participants produced the vowel duration contrasts being tested. Furthermore, the change in this vowel actually saw the participants move away from native-like productions. This could indicate that

the learning process is still ongoing and that the process of forming new representations for the non-native vowels is incomplete.

As for question 2, speaking a quantity language seemed to have a positive effect in the perception task, but not production. In the identification task, identification error rates for the Q group did not differ from the NF control group in any of the tested variables, whereas the NQ group differed from them in the identification of /æ/. Furthermore, the clear advantage for speakers of quantity languages was corroborated by the Q group outperforming the NQ group in the identification of short and long vowels. Some of these differences existed already at the pretest stage, but the Q group was able to improve their identification of /yː/, resulting in a significant difference where one did not exist at pretest. In the production task, however, the non-native groups did not significantly differ for any of the vowel types. Both non-native groups differed from the native speakers in the production of /y/, butthis was at least partially due to atypically exaggerated production ratios by the NF group. These results suggest that overall, speakers of quantity languages may have a perceptual advantage over those with no quantity contrasts in their native language, at least in the relatively short time frame of the course they took part in. In production, the difference is virtually nonexistent.

Finally, for question 3. perception and production did not, in fact, develop in a similar way. While overall improvements during the course were observed for all the vowels in the identification task, for production, any effects of taking part in the course seemed to be limited to the vowel type /y/. Furthermore, while in the identification task all significant differences between pretest and posttest suggested improved performance, i.e., lower error rates in the identification of the duration of the vowels, an opposite result was observed in the production task, with the only significant change over time actually resulting in more similar durations between short and long vowels, rather than less. As stated, where the productions of the

native speakers differed from the participant groups, it was due to the native controls producing larger ratios between short and long vowels.

Of the three vowels that were the subject of this study, most differences and difficulties for the participants seem to be caused by /y/, as the participants were not able to produce the /y/ duration contrasts in a native-like manner at either pre- or posttest. Furthermore, the participants actually performed worse overall in the production of /y/, with a significant change towards less native-like duration ratios. These findings were not explained by the native language background of the participants.

Another explanation for the differences in /y/ could be found in the fact that the native speakers in this experiment produced one of the /y/ contrasts, Y1YY1 (Table 4), with a higher ratio than could be expected based on the literature. It is typically 2.2–2.4 for first syllable vowel duration contrasts, whereas the native control group produced Y1YY1 with a ratio of 3.1. The words “kykkii” and “kyykkää” form a near minimal pair, with both vowels occurring in the first syllable between the same consonants. There is no clear reason for why the native speakers should have produced this contrast in such an exaggerated way and it may simply be an effect of these particular participants. While this may explain the significant difference between the native speakers and the participants, it does not offer an explanation for why the participants’ production performance became worse as the course progressed.

The asymmetrical development of perception and production was a noteworthy feature of the results. All vowels saw statistically significant changes in identification scores towards better performance, while in production the only statistically significant results saw participants actually perform worse after participating in the course. Given that all of the participants were at a relatively early stage in their language studies, it is likely that this is an indication of an incomplete learning process and reflects an U-shaped curve in their language learning (see

further, Rogers, Rakison & McClelland, 2009), namely learning the correct behavior, and then abandoning the correct behavior, only to finally return to the correct behavior. It has been suggested that the U-shaped curve in language learning reflects a step from an "associative" system producing constructions, i.e. phrases learned by heart, to a "rule-based" system applying language rules productively (see Rogers et al., 2009). Based on the results of the current study, it may be that the participants have passed the initial stage of language learning, and are entering a more advanced stage, resulting in less accurate language production.

As stated in the Introduction, it is no longer strictly thought that the proper perceptual representations, i.e. categories, for the non-native contrasts must first be formed, before they can be produced correctly. In this study, the results where the participants moved away from native-like values in the production task, while simultaneously improving their performance in the identification task, could, however, still imply that the representations were not fully formed for all of the Finnish vowels. This interpretation seems particularly likely given that the quantity group displayed a uniform, though non-significant, decrease in their production ratios. While they performed better than the non-quantity group in detecting quantity differences perceptually, it may be that their representations for Finnish vowels are not developed enough to enable them to correctly judge their own productions. As the significant findings are limited to only one of the vowel types, however, no definite conclusions can be drawn, and further research is required.

Overall, the identification performance of the participants was very good, with the quantity group exhibiting completely native-like performance at pretest and posttest, and the non-quantity group also performing at a native-like level in some of the tested variables. This is essentially ceiling-level performance, yet the participants were still able to improve upon it. This may have two possible explanations. First, the Desensitization Hypothesis by Bohn (1995)

suggests that non-native vowel duration contrasts may be salient to learners even if they do not have phonological duration contrasts in their own language. Bohn suggests that the use of duration to differentiate non-native contrasts may be a native language independent strategy that is used when spectral cues are insufficient. This is indeed the case with the stimuli in our experiment, as Finnish vowel qualities change very little, although measurably, with duration changes, with longer vowels exhibiting slightly more peripheral formant values than short vowels (Wiik, 1965, p. 60). Another explanation may lie in the identification stimuli themselves: as they were produced naturally by a Finnish speaker, with no acoustic modification to make them more easy or difficult for the participants, they exhibit Finnish acoustic cues for duration. Suomi et al. (2008) note that while the duration difference between the long and short vowel is the main cue in Finnish, the relative duration between the long vowel and other, short vowels within the word also changes, with short vowels typically becoming even shorter than in words containing only short vowels. In addition to the slight spectral differences between short and long vowels, it is possible that some participants were able to detect some of these other cues and achieve good performance that way.

In other words, being exposed to naturally produced Finnish duration contrasts, L2 learners of Finnish were able to push their already good identification performance further. This may suggest that particularly the non-quantity group initially relied on the native language independent strategy, described by Bohn (1995). Then, as they were exposed to the reliable cues that mark duration contrasts in Finnish, they became more aware of them and were able to utilize them to improve their performance. This, however, cannot be confirmed in the context of this study.

Comparing our findings to laboratory training studies, it does seem that at least perception of non-native duration contrasts can be improved both by specifically designed laboratory training

and more general classroom education. A laboratory training task would have been designed to place a much greater focus on specific contrasts in a given amount of time than is realistic in normal classroom education. With that in mind, the communicative instruction used in this study worked quite well, especially since it was not particularly aimed at improving perception, but rather overall spoken language skills. The fact that the perceptual skills of the participants improved, despite their initial performance already being near native level, shows that the teaching has at least partially shaped their internal representations of Finnish sounds in the relatively short time of four weeks. Our results echo the findings of Hardison and Motohashi-Saigo (2010), who were also able to achieve improved identification of non-native duration contrasts with "communicative activities", with no specific focus on long vowels or geminates, although on more modest scale, owing to the shorter duration of education our participants received.

5 Conclusion

To conclude, our findings suggest that taking part in the intensive language course positively affected the participants' perception of non-native duration contrasts to some extent, despite their already good performance. It is likely that a longer overall course time, with more time spent in the classroom, could have induced learning effects in production as well, as the internal representations of second language phonemes would have had more time to develop and mature, and the participants would have entered the rule-based understanding on the U-shaped language learning curve. It is possible that the participants would be able to learn correct productions better at this stage than they did during this study period, as their internal representations have now developed and could support production learning more, as per the Speech Learning Model (Flege, 1995).

Regarding the practical value of the results, it is unclear whether or not the improvement in perception would have translated to better understanding of the L2 being learned, but the fact that the course was able to effect change is in itself valuable information to L2 educators. Indeed, the purpose of our study was not to assess language skills, but rather the development of the perception and production of particular L2 features. In this regard, the fact that the participants of the study had such a varied native language background is not ideal, as a group consisting of native speakers of a single language would have been easier to interpret. On the other hand, one of the purposes of the study was to observe learning effects in an ecologically valid situation, i.e. the language classroom. The pedagogical implications of the study are the following: a 4-week intensive course using communicative methods, but not focusing specifically on perception, within a native speaking community seems to support L2 learners' perception of vowel duration which is important for understanding the meanings of words in quantity languages. For pedagogical implications for production of duration contrasts, more research is needed, as our findings do not offer clear evidence for the improvement of production skills.

Further studies with similar settings should employ more phonologically and acoustically controlled perceptual stimuli to reduce the ceiling effect and gain a more accurate understanding of the participants' initial perceptual skills. It could also be of interest to add vowel quality measurements to the production metric used in the current study in order to see, for example, whether the proper vowel quality needs to be learned before learning of duration can occur. In future studies, the development of pronunciation skills should also be evaluated in order to see whether changes in comprehensibility, accentedness (see, e.g. Munro & Derwing, 1995) or acceptability (Tulaja, 2019), for example, are linked to any of the features measured in the current study. Finally, a longer time frame, beginning right at the start of the

participants' language studies, could reveal whether the quantity language advantage that was now observed only in the identification test would become more apparent in production skills as well.

6 References

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, and Psychophysics*, 81(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Bohn, O.-S. (1995). Cross-language speech perception in adults: first language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 279–304). Baltimore: York Press.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–276). Baltimore: York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.) *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, *31*(1), 81-99. https://doi.org/10.1017/S0142716409990178

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, *116*(4), 2384–2394. https://doi.org/10.1121/1.1783351

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *The Journal of*

the Acoustical Society of America, *121*(6), 3837–3845. https://doi.org/10.1121/1.2734401

Isei-Jaakkola, T. (2004). *Lexical quantity in Japanese and Finnish*. University of Helsinki.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology*, *67*(2), 131–143. https://doi.org/10.1016/j.ijpsycho.2007.10.012

Kuukka, K. & Metsämuuronen, J. (2016). Perusopetuksen päättövaiheen suomi toisena kielenä (S2) -oppimäärän oppimistulosten arviointi 2015. Julkaisut 13:2016. [The Assessment of Finnish as a Second Language (S2) Curriculum for Basic Education 2015. Publications 13:2016.] Helsinki: Karvi. https://karvi.fi/app/uploads/2016/05/KARVI_1316.pdf

Lehtonen, J. (1970). Aspects of quantity in standard Finnish. *Studia Philologica Jyväskyläensia VI.* University of Jyväskylä.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258. https://doi.org/10.1006/jpho.2002.0174

Meister, E., Nemoto, R. & Meister, L. (2015). Production of Estonian quantity contrasts by Japanese speakers. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 6, pp. 330–334). International Speech and Communication Association. https://doi.org/10.12697/jeful.2015.6.3.03

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Okuno, T. (2014). *Acquisition of L2 Vowel Duration in Japanese by Native English Speakers*. Dissertation Abstracts International. Michigan State University.

Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration:

Training with technology. *Language Learning and Technology*, *20*(2), 61–80.

Rogers, T. T., Rakison, D. H., & McClelland, J. L. (2009). U-shaped curves in development: a PDP approach. *Journal of Cognition and Development*, 5(1), 137–145. https://doi.org/10.1207/s15327647jcd0501_14

Saloranta, A., Alku, P., & Peltola, M. S. (2017). Learning and generalization of vowel duration with production training: behavioral results. *Linguistica Lettica*, 25, 67–87.

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, 147(November 2019), 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243–261. https://doi.org/10.1017/S0142716400001417

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, phonology, phonotactics and prosody*. Oulu University Press.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, *123*(1), 397–413. https://doi.org/10.1121/1.2804942

Tsukada, K. (2012). Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese. *Applied Psycholinguistics*, 33(3), 501–516. https://doi.org/10.1017/S0142716411000452

Tulaja, L. (2019). *Dänische L2-Aussprache von Lernern mit Deutsch als Ausgangssprache : Fehler und Fehlerschwere* [German L2 Danish Learners' Pronunciation : Typical Errors and Error Gravity]. Christian-Albrechts-Universität zu Kiel. https://doi.org/10.21941/74F3-7P42

Wiik, K. (1965). *Finnish and English vowels: A comparison with special reference to the learning*

problems met by native speakers of Finnish learning English. University of Turku.

Ylinen, S., Shestakova, A., Alku, P., & Huotilainen, M. (2005). The perception of phonological quantity based on durational cues by native Speakers, second-language users and nonspeakers of Finnish. *Language and Speech*, *48*(3), 313–338. https://doi.org/10.1177/00238309050480030401

Appendix A

Identification token pairs

/lyp:y/, /ly:p:y/	/læp:æ/, /læ:p:æ/	/løp:ø/, /lø:p:ø/
/lyt:y/, /ly:t:y/	/læt:æ/, /læ:t:æ/	/løt:ø/, /lø:t:ø/
/myk:y/, /my:k:y/	/mæk:æ/, /mæ:k:æ/	/møk:ø/, /mø:k:ø/
/myp:y/, /my:p:y/	/mæp:æ/, /mæ:p:æ/	/møt:ø/, /mø:t:ø/
/myt:y/, /my:t:y/	/mæt:æ/, /mæ:t:æ/	/møp:ø/, /mø:p:ø/
/nyk:y/, /ny:k:y/	/næk:æ/, /næ:k:æ/	/nøk:ø/, /nø:k:ø/
/nyp:y/, /ny:p:y/	/næp:æ/, /næ:p:æ/	/nøp:ø/, /nø:p:ø/
/nyt:y/, /ny:t:y/	/næt:æ/, /næ:t:æ/	/nøt:ø/, /nø:t:ø/
/ryk:y/, /ry:k:y/	/ræk:æ/, /ræ:k:æ/	/røp:ø/, /rø:p:ø/
/ryt:y/, /ry:t:y/	/ræp:æ/, /ræ:p:æ/	/røt:ø/, /rø:t:ø/
/syk:y/, /sy:k:y/	/ræt:æ/, /ræ:t:æ/	/røk:ø/, /rø:k:ø/
/syp:y/, /sy:p:y/	/sæk:æ/, /sæ:k:æ/	/søk:ø/, /sø:k:ø/
/syt:y/, /sy:t:y/	/sæp:æ/, /sæ:p:æ/	/søp:ø/, /sø:p:ø/
/tyk:y/, /ty:k:y/	/sæt:æ/, /sæ:t:æ/	/søt:ø/, /sø:t:ø/
/typ:y/, /ty:p:y/	/tæk:æ/, /tæ:k:æ/	/tøk:ø/, /tø:k:ø/
/tyt:y/, /ty:t:y/	/tæp:æ/, /tæ:p:æ/	/tøp:ø/, /tø:p:ø/
	/tæt:æ/, /tæ:t:æ/	/tøt:ø/, /tø:t:ø/

Note: IPA transcriptions are used in this article, participants saw the orthographic forms of the words.

Appendix B

All Finnish production sentences and their English translations,

Me teemme kyykkyjä	We are doing squats
Kalliolla on kyy	There is an adder on the hill
Mummo kykkii pellolla	**Grandma is squatting in the field**
Sinulla on tyyliä	You've got style
On helppo tyriä	It is easy to mess up
Tuli kytee nuotiossa	The fire is smouldering
Sydämeni tykyttää kovasti	My heart is beating fast
Te pelaatte kyykkää	**You are playing skittles**
Tylli on kangas	Tulle is a fabric
Mies oli tyly	The man was rude
Typpi on alkuaine	**Nitrogen is an element**
Kylä on pieni	The village is small
Tyyppi tulee huomenna	**The guy is coming tomorrow**
Meri on tyyni	The sea is calm
Tyyny on pehmeä	The pillow is soft
Joukossamme on kyylä	There is a creep among us
Sinulla on kyky	You have the ability

Liisa kykenee tehtävään	Liisa is capable of the task
Minä tyydyn tähän	I am settling for this
Auto on tyyris	The car is expensive
Jussi pelaa lätkää	Jussi plays hockey
Tule tänne äkkiä!	Come here quickly!
Pasta laitetaan lävikköön	Pasta is put in a sieve
Niityllä on lääte	There is a common saw-wort in the field
Mies lääppii koiraa	**The man is petting the dog**
Lammas määkii pellolla	**The sheep is bleating in the field**
Sinä olet pätkä	You are a shorty
Matkustan huomenna Vääksyyn	I'm traveling to Vääksy tomorrow
Suomessa on ääkköset	Finnish has å, ä and ö ["ääkköset" refers to these three letters]
Kääpiö kaivaa maata	The dwarf is digging dirt
Huoneesi on läävä	Your room is a sty
Nainen heittää läppää	**The woman is telling jokes**
Kääty tehtiin kullasta	The necklace was made of gold
Kissalla on käpälä	The cat has a paw
Edessä on mäki	**There is a hill ahead**
Karhulla on iso kämmen	The bear has a large paw

Sängyn pääty hajosi	The headboard of the bed broke
Vain käteinen kelpaa	Only cash is accepted
Minulla menee käämit	I am getting angry
Torilla on väkeä	There are people in the marketplace

Ilmassa on pölyä	There is dust in the air
Ojassa on löpö	There is a scaber stalk mushroom in the ditch
Äiti höösää aina	Mother is always fussing
Anna minulle köö	Give me the [pool] cue
Lössi seikkailee kaupungilla	The gang is out and about in the city
Pöö, mörkö tulee!	**Boo, the bogeyman is coming!**
Sinä olet pölö	**You are a dummy**
Räikkösellä on ökytalo	Räikkönen has an opulent house
Kivellä on ötökkä	There is a bug on the rock
Ostan uuden mööpelin	I am buying a new piece of furniture
Aina joskus töppäilee	One makes mistakes sometimes
Taksi tööttää kovaa	**The taxi is honking loudly**
Jäätelö on tötterössä	**The ice cream is in a cone**
Matti puhuu lööperiä	Matti is talking nonsense
Maija on tökerö	Maija is crass

Imurista kuuluu mölyä	The vacuum cleaner is making noise
Pöllö on viisas	The owl is wise
Auto töötöttää tiellä	The car is honking on the road
Lehdessä on lööppi	There is a headline in the newspaper
Kööri laulaa laulun	The group sings a song

Note: Bolded sentences indicate those used in the production task of the study.

Appendix C

Mean production ratios based on the existence of a /y/ phoneme in the native language

		/ʏ/	/æ/	/ø/
NY	**Pretest**	1.69 (0.72)	1.58 (0.72)	2.12 (0.87)
	Posttest	1.56 (0.88)	1.51 (0.63)	2.18 (0.9)
Y	**Pretest**	1.72 (0.79)	1.69 (0.64)	2.42 (0.8)
	Posttest	1.56 (0.84)	1.62 (0.62)	2.33 (0.69)

Note. NY = speakers of languages with no /y/ phoneme. Y = speakers of languages with a /y/ phoneme.

Values in brackets denote standard deviations.

Author details:
Antti Saloranta
Koskenniemenkatu 4
20100 Turku
FINLAND
antti.saloranta@utu.fi

Leena Maria Heikkola
Tehtaankatu 2
20500 Turku
FINLAND
leena.maria.heikkola@abo.fi