LEARNING AND GENERALIZATION OF VOWEL DURATION WITH PRODUCTION TRAINING: BEHAVIORAL RESULTS

Antti Saloranta[1,2], Paavo Alku[3], Maija S Peltola[1,2]

1) Department of Phonetics, University of Turku 2) Learning, Age and Bilingualism laboratory (LAB-lab), University of Turku 3) Department of Signal Processing and Acoustics, Aalto University

Abstract

The purpose of this study was to determine whether a short listen-and-repeat training paradigm can be used to train vowel duration discrimination and production, and whether any learning effects are transferred to an untrained vowel or a non-linguistic sound. Similar training has previously been used to train vowel quality contrasts to young adults and children, with results showing up both in behavioral and psychophysiological measurements. Unlike vowel quality, segment duration can be considered to be a suprasegmental feature that is not directly dependent on any other acoustic feature of the sounds being trained. It is therefore plausible that it can be learned a separate skill and generalize to untrained segments, and even non-linguistic sounds. Participants were 18-30-year-old healthy adults with normal hearing, who were not native Finnish speakers and had spent little time in Finland. The stimuli were semisynthetic Finnish pseudoword pairs /tite/-/ti:te/ and /tote/-/to:te/. A sinusoidal tone pair served as the non-linguistic stimulus. The behavioral measurements employed in the study were an oddball discrimination task for all three stimulus pairs, and a listen-and-repeat production task for both of the vowel pairs. No feedback was given. The experiment was conducted in three sessions over three days. The first two consecutive days consisted of baseline measurements for all the stimuli and four blocks of production training. The third day, taking place 1-2 weeks after the second, consisted of full progress measurements. The results show that the training did induce changes in discrimination sensitivity and production of the trained length contrasts, though not all effects remained at the end of the experiment. This suggests that while the processing mechanisms related to the processing of duration contrasts are somewhat separated from the processing of vowel quality, it seems that they can be accessed with this kind of training.

Keywords: *phonetic training, production training, second language acquisition, vowel length*

1. Introduction

Finnish is a quantity language, which means that the duration of phonetic segments has a phonologically distinctive role in it. The Finnish quantity system is fairly extensive, with all vowels and most consonants displaying length differences that can occur in all syllables for vowels (Suomi et al. 2008:48). For example, all lengthening combinations of the voiced segments in the Finnish word /tule/ are phonologically distinctive. Quantity systems of this extent are somewhat rare in most major languages in the world; segment duration is more typically used to denote stress or other similar, not necessarily phonological features. Many L2 learners of Finnish, therefore, are faced with having to learn this ubiquitous feature and it is often considered quite difficult. The purpose of this paper is to present results from a training scheme aimed at training the perception and production of vowel duration in speakers whose native language does not contain such features.

Most modern models of second language acquisition predict that perception and production of sounds that do not exist as such in the speaker's native language is difficult. Both the Speech Learning Model (SLM) by James E. Flege (e.g. Flege 1987) and the Perceptual Assimilation Model (PAM) by Catherine Best and Winifred Strange (e.g. Best and Strange 1992) present a similar view, in which a novel phoneme category from a foreign language is assessed based on the existing phonemic system in the native language. Both models differentiate between completely new categories, such as consonants that do not resemble any existing native language ones, and similar categories that partially match ones in the native language, but differ systematically. The most difficult situations arise in the case of similar categories, when the learner needs to be able to perceive a phoneme category difference where one does not exist in their native language. An example of this is the existence of a phonemic contrast between voiced and voiceless sibilants in English from the viewpoint of a Finnish speaker: Finnish only has one voiceless sibilant phoneme that may become slightly voiced in certain phonetic environments. Both the PAM and the SLM view this as a situation where both categories are considered to be members of a single category in the native language, and therefore both sibilants are thought to be exemplars of the native sibilant category.

The learning problems detailed in the previous paragraph stem from the early influence of the native language on the perceptual system. Very young children of only a few months of age are able to differentiate phonetic differences from any language very accurately, but already by six months the brain starts to become desensitized to phonetic variation that is non-distinctive in the native language (Kuhl et al. 1992). This has been demonstrated also by Iverson et al (2003), who showed that native Japanese speakers show highly reduced sensitivity to changes in the F3 formant in the English /r/-/l/ contrast, explaining the notorious difficulty Japanese speakers have with this contrast. Accordingly, Finnish speakers have been shown to have a higher degree of sensitivity to phonetic duration differences than Germans (Kirmse et al. 2008), and they also appear to have phonemic categories for lengths while Russian speakers do not (Ylinen, Shestakova, et al. 2005). Furthermore, in a study examining the acquisition of a Swedish quantity contrast, speakers of Estonian, a quantity language, performed better than speakers of English and Spanish in which segment duration has a much less prominent role (McAllister et al. 2002).

Relatively few training studies have been conducted on training the Finnish quantity system using L2 learners of Finnish as the test subjects. Some studies, however, have been conducted on L2 learning and training of the phonological quantity in Japanese, which is highly similar to the Finnish system. Both languages have a binary quantity system, meaning that they differentiate between long and short phonemes (Isei-Jaakkola 2004:1). Quantity can be contrastive in all vowels and most consonants (Meister and Meister 2013:79; Suomi et al. 2008:41) and both languages also allow vowel lengthening in any syllable (Meister and Meister 2013:80; Suomi et al. 2008:41). Due to these similarities, an L2 learner of Japanese can be said to face difficulties similar to a learner of Finnish, and it is therefore of interest to examine results from training studies conducted in a Japanese learning context.

Hirata, Whitehurst and Cullings (2007) trained young adult native English speakers to identify Japanese vowel length contrasts in carrier sentences at different speaking rates. The training consisted of a basic forced-choice identification task with feedback using naturally produced stimuli. Subjects trained with a total of 540 stimuli, divided in to four sessions over 11-17 days. The study found significant, although weak, improvement in the perception of the length contrasts for the group that received training with the lowest speaking rate. Generalization effects were not tested. Tajima et al (2008) used minimal-pair identification training with feedback in training 19–25-year-old native Canadian English speakers to distinguish between Japanese vowel and consonant length contrasts. All tasks consisted of forced-choice identification, where the subjects heard the stimulus and were instructed to choose between transcriptions. Subjects were given immediate feedback during training. The total amount of training was 15 sessions over 5 days, lasting on average 35-60 minutes/day with a total of 3600 training trials. Identification performance improved slightly on the tested contrasts, but no generalization effects to untrained stimuli or to new talkers were found. Okuno (2014) trained L1 English university students of various experience levels on identifying Japanese vowel length contrasts using audio only and audiovisual training. The training groups underwent eight 25-minute sessions of forced-choice identification training of the stimuli, with waveforms acting as the visual component for the audiovisual group. The perception task and the training used natural bisyllabic stimuli produced by native speakers of Japanese and feedback was given for all answers. Both training groups were able to improve their identification accuracy in comparison to the control group, with no statistically significant difference between the two training types. The effects generalized to both untrained tokens and previously unheard talkers. Notably, production accuracy also increased in both group *vs* the control group, even though no production training was given.

While the previous studies show that vowel and consonant length perception (and to some extent, production) can be improved with perceptual training, the results are somewhat mixed. All studies reported improvement in the perception of length contrasts, but the effects remained relatively minor, and generalization to new talkers or stimuli did not occur consistently, in spite of the large amount of training the subjects received. It seems, therefore, that perceptual training alone may not be sufficient for reliable acquisition of length contrasts. What is notable is the lack of studies using production training on length contrasts, even though production training has been shown to elicit good results in training of other L2 contrasts, with results visible both behaviorally and psychophysiologically. A recent study (Taimi et al. 2014) showed that young children can learn to produce a novel vowel quality contrast in just two days of listen-and-repeat training. In the study, 7-10-year-old Finnish children trained the production of the Swedish /ʉ/ – /y/ contrast that is not found in Finnish. Already after three of the four short sessions of training, meaning 90 of a total of 120 repetitions of the contrast over two days, the children were able to modify their productions and accurately produce the previously unfamiliar vowel. In another study (Saloranta et al. 2015) the same listen-and-repeat procedure, enhanced with instructions, was used to train the same contrast in 18-30-year-old adults. The participants were able to modify their productions after just one session of training, during which they had simply been made explicitly aware of the novel contrast in the stimuli. Listen-and-repeat training also proved effective with linguistically oriented senior subjects, who were able to modify their production of a foreign vowel contrast after two days of training (Jähi et al. 2015). It has also been shown that a similar, three-day production training scheme can create new memory traces for novel second-language contrasts (Tamminen et al. 2015) or further strengthen existing ones (Tamminen and Peltola 2015).

The success of the production training schemes described above may result from a combination of several aspects. Previous studies (e.g. Guion and Pederson 2007; Pederson and Guion-Anderson 2010) have shown that training focusing the subjects' attention specifically on aspects relevant to the trained feature may aid in their learning. This approach was highly successful in Saloranta et al (2015), in which the training was enhanced with instructions with the dual purpose of both focusing the subjects' attention on the acoustically relevant features of the contrast and helping them form the articulatory gestures necessary for the production of the non-native vowel. It seemed that redirecting of attention enable the subjects to focus on the relevant acoustic differences in the trained contrast. Another reason for the success of listen-and-repeat training may be the interconnectedness of speech perception and production. The Directions Into Velocities of Articulators (DIVA) model (e.g. Guenther and Hickok 2015) suggests that the development of correct motor patterns for speech production in childhood is linked to the development of perceptual categories. The model posits that children develop their native phonetic categories during the first months of life, and then use them as models for the development of production motor patterns. During its development, the system employs the brain's motor and acoustic feedback systems, but in later life the system is mainly maintained by the acoustic feedback system (Perkell 2012), i.e. the person hearing themselves and others speak and subconsciously altering and correcting their own speech. Experimentally, it has been shown that perceptual changes can be elicited in as little as 45 minutes of phonetic categorization or discrimination training with feedback due to warping of the perceptual space by new stimuli (Guenther et al. 1999). Phonetic training has been shown to improve production of non-native consonant contrasts (Bradlow et al. 1997; Tajima et al. 2008), and it has also been shown that relatively little articulatory production training can also improve perception of novel vowel and consonant contrasts (Catford and Pisoni 1970). Listen-and-repeat training may be efficient in combining the benefits of these methods by providing the subject with a new target that they can learn to both discriminate and produce simultaneously, with feedback coming from the subject's own productions.

The purpose of the current study is to examine the effectiveness of a listen-and-repeat production training scheme on the perception and production of vowel duration differences. There are three main research questions stemming from previous research. Firstly, can vowel duration be trained similarly to and as quickly as vowel quality? Speech segment duration and quality are processed separately in the brain, suggesting they are two different systems (Ylinen, Huotilainen, et al. 2005) meaning that they may not behave similarly under similar training. Secondly, are any possible learning effects transferred to other, untrained vowels? Previous duration training studies, such as the ones presented earlier, have not consistently shown generalization effects, but the training employed in them has focused mainly on discrimination with no production elements. Finally, does training with linguistic stimuli affect discrimination of duration in non-linguistic sounds? It has been suggested that temporal processing in both speech and non-speech is done using a general neural mechanism (Liégeois-Chauvel et al. 1999), and if duration discrimination can indeed be learned as separate feature from individual segments, it stands to reason any increases in accuracy could also be seen in non-speech sounds.

2. Materials and methods

2.1. Subjects

Subject recruitment was under special scrutiny in this project, as care had to be taken to ensure that the subjects' native languages contained no phonological length in any types of sounds. While a quantity system similar to the Finnish one in extensiveness is found in few languages and the quantity distinctions in the study would likely prove difficult for speakers of most major world languages, any languages containing phonological length contrasts were ruled out. This was done to ensure that all subjects would be able to discriminate the length distinction equally poorly at the beginning of the experiment. Subjects consisted of 7 (6 female) 19-29-year-old healthy, normally hearing and right-handed adults, who were recruited among the exchange students entering the University of Turku. All subjects volunteered to take part in the project and were not compensated for their participation in any way. Upon volunteering for the project, subjects were first asked to provide basic information about their eligibility, which included age, native language, other spoken languages, pre-existing neurological conditions or medications, handedness and length of stay in Finland. Should the subject be eligible, this information was clarified further on the first test day before the beginning of the actual experiment in order to make the final eligibility decision. Language skills were self-evaluated in three separate sections regarding overall language ability, frequency of language use and frequency of passive exposure to the languages in the media etc. The subjects were spoken to in English during the experiment. The subjects' hearing on the 100-4000 Hz range at 5-25 dB was tested using a Grason-Stadler GSI 18 audiometer. No subject showed any substantial hearing problems in this range. Written consent was obtained from all subjects for the use of the data in this and future projects.

2.2. Stimuli

Three stimulus pairs were synthesized for the experiment. Two of them functioned as linguistic stimuli, consisting of Finnish two-syllable pseudoword pairs /tite/ - /ti:te/ and /tote/ - /to:te/, differing in the length of first syllable. The former will from here on be called the trained linguistic pair, and the latter the untrained linguistic pair. The third, non-linguistic pair consisted of sinusoidal tones mimicking the main temporal and spectral structure of the pseudowords. The short members of the pairs were 392 ms long, and the long ones 428 ms; first syllable lengths were 154 ms and 194 ms, respectively. The linguistic stimuli were synthesized using the semisynthetic method SSG (Alku et al. 1999), which uses an extracted glottal excitation waveform from a real speaker, producing natural sounding stimuli with phonetic features that can be carefully controlled. The lowest sinusoidal tone of the non-linguistic stimuli were adjusted to be equal to the mean of the F1 and F2 frequencies in the corresponding vowel of the linguistic stimuli. In addition to the lowest tone, the non-linguistic stimuli consisted of one sinusoidal per every 1 kHz and the non-voiced sections were synthesized with a 6^{th} order linear prediction filter (Makhoul 1975) excited with white noise. All stimuli were synthesized at the Department of Signal Processing and Acoustics at Aalto University.

2.3. Discrimination task

The discrimination tasks were all performed using an oddball paradigm, with 130 short stimuli as the standards and 20 long ones as the deviants, with an interstimulus interval (ISI) of 1000 ms. The stimuli were presented binaurally with Sennheiser HD 25-1 II headphones connected to a desktop PC running Presentation version 16.3 by NeuroBehavioral Systems. In the task, subjects were told that they would hear two stimuli, either words or sounds, with one being repeated often and the other only occasionally, and that they should press the response button as quickly as they could when they heard the less frequently occurring stimulus. Subjects were not told any specific qualities of the stimuli, only whether they would be hearing words or sounds. No feedback was given. Reaction times to and detection rates of the deviants were measured in each task from the beginning of the stimulus, and the latter were used to calculate discrimination sensitivity (d’) in all instances of the task.

2.4. Production task and training

A listen-and-repeat paradigm was used as both the training task and as the measure of production skills in the experiment. In both cases, a stimulus pair was presented binaurally in an alternating long–short pattern with an ISI of three seconds using the Sanako SLH-07 headset and Sanako Lab 100 software. Similarly to the discrimination task, subjects were not given any specific information pertaining to the stimuli they heard; they were simply told to repeat each word they heard as accurately as they could, according to their own judgment. The main difference between the production task and the training phase was the length of the stimulus blocks: during training, the subjects repeated the stimulus pair 30 times per block, while the production task consisted of 10 repetitions. In total, the subjects therefore repeated the trained linguistic pair 150 times and the untrained 10 times. Before each training block, the subjects were told that if they felt they had any problems repeating the words correctly during the baseline recording, they could use the following blocks as a chance to practice. If they did not feel like they needed to practice, they were told to still repeat each word in order to ensure the same amount of repetitions between subjects. The production task was performed only with the trained linguistic pair on the first two days, and with both linguistic pairs on the third. The untrained linguistic stimuli were only tested on the third day in order to minimize production experience with them, and to have them remain truly untrained compared to the trained stimuli. Only the trained linguistic pair was used during training.

Whole word durations, initial syllable durations and initial syllable vowel durations were measured from all recorded words using the Praat software version 6.0.0.5 (Boersma and van Heuven 2001). It was decided that analysis would focus on the length of the vowels produced in the first syllable. This was done because several of the subjects produced the ends of the words with strongly breathy articulation, making the endpoints difficult to determine. Furthermore, the change in duration that was being trained occurred in its entirety in the vowel of the first syllable, and changes in production were also expected to happen there. The data was further normalized by calculating the ratios between the long and short syllables by dividing the length of the long vowels by the length of the short vowels in each session. This minimized the effects of speaking rate between individual subjects, as only the relative difference between long and short productions was being examined.

2.6. Experiment structure

A three-day structure was employed in the experiment, consisting of baseline/progress measurements on each day and listen-and-repeat training on the first and second day. The measurements were performed on all stimulus pairs on the first and third days, and only on the trained pair on the second day. This was done to gauge the effectiveness of the training on the trained stimuli immediately after training had stopped, while simultaneously giving any transfer of learning effects more time to manifest. The order of the measurement and training blocks can be seen in Figure 1. At the end of the third day, subjects were asked to briefly self-evaluate their performance in the experiment and to discuss any difficulties they faced during it.

In addition to the behavioral experiments discussed in this article, electroencephalography (EEG) recordings were also performed for all stimulus pairs in order to measure mismatch negativity (MMN) responses for each of the contrasts. In each block, the subjects heard each short member of the stimulus pairs 874 times, and each short member 140 times while attending to a silent movie. Responses for the trained linguistic stimuli were measured on all test days, and on the first and third days for the untrained linguistic and non-linguistic pairs. Results from these measurements will not be discussed in this article.

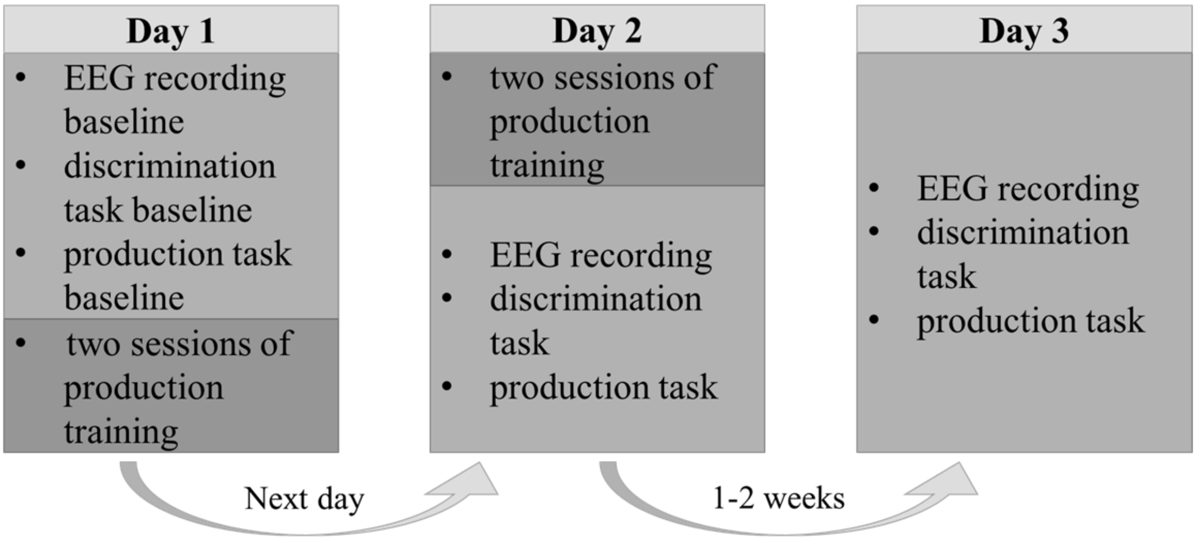

Figure 1. Structure of the experiment. On Day 1 and Day 3, all three stimulus pairs were used in the baseline/progress measurements, apart from the production task baseline in which the untrained linguistic or the non-linguistic pairs were not used. All four training sessions and the progress measurements on Day 2 were conducted with only the trained linguistic stimuli.

3 Results

In the self-evaluation, all subjects correctly identified segment length as the feature being studied, despite having not received any feedback or information regarding it. Most subjects felt that the experiment had become easier as it progressed, but some of them felt that they could still have done better. The non-linguistic stimuli were considered to be the most difficult to discriminate by nearly all subjects, followed by the untrained linguistic, but most subjects felt their performance was comparable with all stimulus types by the end of the experiment. No subjects self-reported any regression in their performance on the third day.

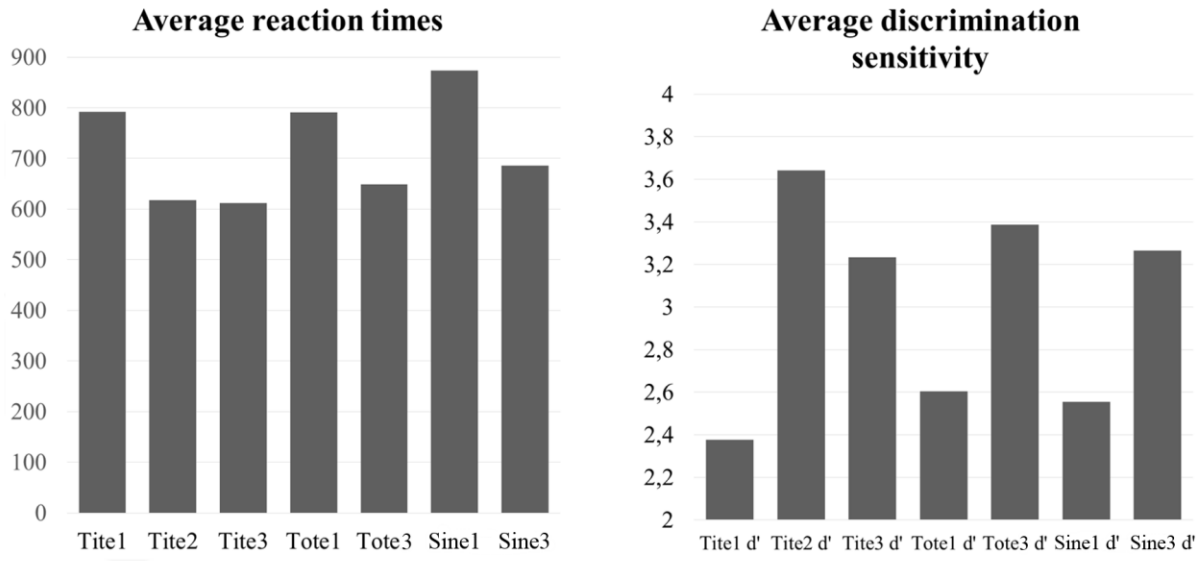

Figure 2. Average reaction times and average sensitivity in the discrimination task on separate days. The numbers 1-3 on the x-axis represent the experiment day on which each value was measured for the given stimulus: the trained linguistic stimulus /tite/ was measured on all three days, while the other two were measured on the first and last day.

Individual subject discrimination sensitivity scores							
	Tite1 d'	**Tite2 d'**	**Tite3 d'**	**Tote1 d'**	**Tote3 d'**	**Sine1 d'**	**Sine3 d'**
S01	0,23	2,41	3,05	0,36	3,94	0,51	3,28
S02	3,05	4,61	4,31	0,71	2,26	2,41	3,10
S03	0,71	2,30	0,96	2,41	2,31	0,71	1,02
S04	4,61	4,37	3,95	4,31	4,31	3,70	4,31
S05	3,19	4,31	3,70	3,47	4,28	3,51	3,34
S06	1,15	2,88	2,06	3,04	2,00	3,70	4,31
S07	3,70	4,61	4,61	3,95	4,61	3,34	3,51
AVG	**2,45**	**3,78**	**3,35**	**3,00**	**3,48**	**2,53**	**3,00**
STDEV	**1,53**	**0,82**	**1,18**	**1,5**	**1,1**	**1,2**	**1,04**

Table 1. Scores of individual subjects in the discrimination task and the group averages and standard deviations.

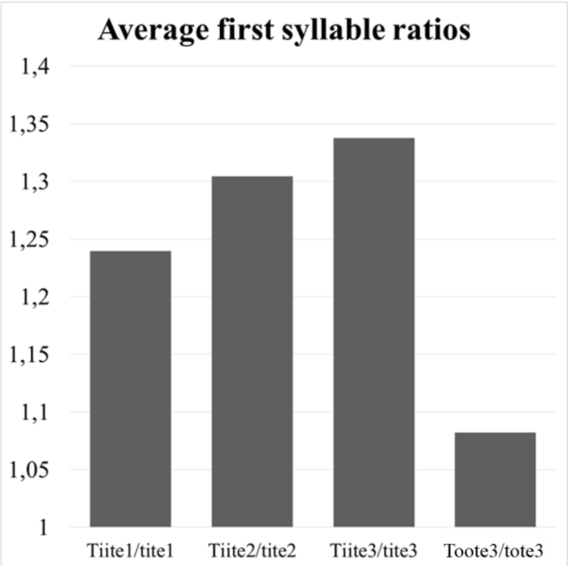

Figure 3. Average production ratios for the trained stimuli (recorded on all days) and the untrained stimuli (recorded on the last day).

Statistical analysis of the results started with the average discrimination sensitivity scores (Figure 2). A Repeated Measures Analysis of Variance (ANOVA) with Word(3) X Session(2) was run to compare scores between the baseline and endpoint in all words. No effects reached significance. In order to examine the effectiveness of the training on the trained stimulus alone, an ANOVA was performed with Session(3). This resulted in a main effect of Session ($F(2,5) = 9,907$; $p = 0,018$), indicating that discrimination sensitivity was different between sessions. Post hoc paired samples t-tests were performed in order to compare differences between individual sessions. This revealed a significant difference between Sessions 1 and 2 ($t(6) = -4,280$; $p = 0,005$) but not between Sessions 1 and 3, indicating that compared to the baseline, discrimination of the trained stimuli was significantly improved directly after the training, but that the effect was possibly not retained.

The statistical analysis of discrimination sensitivity indicates that it peaks in the second session, with the third showing values similar to the untrained linguistic and non-linguistic stimuli. In order to further examine this result and rule out the effects of major outliers, individual scores for discrimination sensitivity are presented in Table 1. This reveals a wide range of baseline values. The ceiling value for this score is 4,61, and one of the subjects, S04, was able to achieve it already in the first session, before any training had taken place. Most, however, display lower baseline scores. Examination of Session 2 shows an increased score for all but one subjects, S04, whose score decreased from the ceiling level they were able to achieve in the first session. In the third Session, however, 5/7 subjects show a decrease in their discrimination scores compared to the second session. Only subjects S01, who continued a steady increase from baseline, and S07, who achieved ceiling level in Session 2, were able to increase or maintain their scores. The score for S04, who achieved ceiling level in the first Session, decreased further. These individual data therefore corroborate the results of the statistical analysis: for most subjects, discrimination performance with the trained stimuli indeed improved after training, but declined between the second and third session.

Statistical analysis of the baseline and endpoint discrimination reaction times started with an ANOVA with Word (3) X Session (2). This resulted in a main effect of Session ($F(1,6)$ = 9,001; p = 0,024) and Word ($F(2,5)$ = 8,399; p = 0,025), indicating that the reaction times were different between the baseline and endpoint sessions and that they were different in different words. In order to examine these effects further, post hoc tests were conducted, starting with paired samples t-tests between the first and last sessions for each word. Of these, only the reaction times for the non-linguistic stimuli showed statistically significant decrease between sessions 1 and 3 ($t(6)$ = 3,144; p = 0,02). Next, paired samples t-tests comparing reaction times between words within sessions were conducted. The difference between the trained linguistic stimuli and the non-linguistic stimuli reached significance in Session 1 ($t(6)$ = -2,687 ; p = 0,036) and Session 3 ($t(6)$ = -3,144; p = 0,02). No other tests reached significance.

Analysis of the production ratios, seen in Figure 3, began with an ANOVA with Session(3) for the trained words. No significant effect was found. The ratios for the trained words in different sessions were then compared to the single ratio for the untrained word using paired samples t-tests. A significant difference was observed in Session 2 ($t(6)$ = 1,824; p = 0,024) and in Session 3 ($t(6)$ = 3,776; p = 0,009) but not in Session 1, indicating that the trained words were produced with long/short vowel ratios similar to the untrained ones before training, but not after it.

4 Discussion

In this article, a listen-and-repeat training scheme for the training of vowel duration perception and production was presented. The aim was to find out whether improved vowel duration perception and production could be achieved with a relatively short three-day training paradigm, whether any training effects would transfer to an untrained vowel, and whether improved perception could also be detected in non-linguistic duration differences of similar magnitude. Research was backed by previous successful training effects results from similar production training paradigms in improving both behavioral and psychophysiological perception and production of vowel quality contrasts. Furthermore, some results from earlier studies using perceptual training of vowel durations showed that improved perception of duration can be trained, although the effects may be minor, and that generalization of the training results is also possible, though not consistently achieved.

The results from the study are somewhat mixed, although promising. Signs suggesting potential efficacy for the method could be observed, as some training effects for the trained stimuli emerged in both the discrimination scores and the production task. Discrimination sensitivity showed a statistically significant increase on the second day, but the effect disappeared again on the third. The reasons for this apparent regression can be twofold. It could be due to the amount of training not being enough to actually produce lasting changes for the novel contrast, meaning that the internal model of sound perception and production, as per the DIVA model, had begun to be modified, but the changes were too slight to have lasting effects. Perhaps duration information is stored in the internal model differently to quality information; the results suggesting that phoneme duration is processed separately from quality (Ylinen, Huotilainen, et al. 2005) lend credence to this interpretation. While a similar amount of production training did produce learning effects for vowel quantity (e.g. Saloranta et al. 2015; Taimi et al. 2014; Tamminen et al. 2015), training may need to take into account the more complex nature and separate processing of speech duration in order to produce more lasting perceptual effects. On the other hand, another reason for the lack of statistical significance may simply be the low sample size of the experiment and the increased standard deviation, which somewhat limit the statistical power of the analyses. The decrease in the average discrimination score between the second and third day is quite small compared to the increase between the first and second day (0,43 *vs* 1,33, respectively) and it may be that more data could maintain the effects also on the third day.

As stated, learning effects were also observed in the production results. While the trained stimuli showed no significant between-session differences, comparative analysis between recordings of the trained and untrained stimuli showed that while productions between the two were statistically similar on the first day, the difference between them was statistically significant on the second and third days. This suggests that subjects were able to change their production of the trained stimuli, but not the untrained stimuli; had the latter been the true, results for both words would likely have stayed similar throughout the experiment. This effect is likely mainly caused by poor performance by the subjects in differentiating short and long vowels in their production of the untrained stimuli; differentiation was notably, though not significantly, better with the trained stimuli already before training and continued to improve throughout the experiment in relation to the untrained ones. Reasons for this are unclear, as discrimination scores were comparable in all stimuli, and no subjects reported major difficulties with producing either of the linguistic pairs compared to the other, although the untrained ones were thought to be slightly more difficult. More data and further analysis may shed light on this discrepancy.

As a somewhat surprising finding, the only statistically significant changes in discrimination reaction times were observed with the non-linguistic stimuli, whose times showed statistically significant decrease between sessions, but remained significantly higher than the ones for the trained stimuli throughout the experiment. The reaction times for the untrained linguistic stimuli were somewhere in between, showing no significant change in overall reaction times or in their relationship to the other stimuli. This seems to indicate that the non-linguistic sounds were slower for the subjects to process in spite of the discrimination scores showing no significant differences between the words. It may reflect the perceived difficulty of the stimuli that was often mentioned in the self-evaluation: while subjects were able to discriminate the stimuli as well as the others, the process was more demanding. The slow reaction times could be explained simply by the foreign nature of the non-linguistic sounds: pure sine tones are quite rare in everyday life and their processing is therefore likely to be more demanding and therefore slower. The significant improvement, on the other hand, seems like a generalization effect at first. However, considering no improvement in reaction times was observed with either of the linguistic stimuli, the improvement here is unlikely to show generalization. It is more likely to be a task familiarization effect: in the beginning subjects may have been more hesitant with the foreign sounding sine tones than with the linguistic stimuli, but in the end they felt more confident in their judgments and were able to make decisions more rapidly, though not at the same rate as with the trained linguistic stimuli. This is supported by the self-evaluations, where most subjects felt that the task had become easier as it progressed from day to day.

5 Conclusion

Overall, it seems that the methodology described in this study shows promise in producing learning results with suprasegmental contrasts, as the training was able to elicit clear, significant changes in discrimination sensitivity and production ratios. Although statistically significant improvement was observed in the reaction times to the non-linguistic stimuli, any generalization effects as a result of training could not be statistically confirmed. The study also hinted at a processing difference between linguistic and non-linguistic sounds, as the reaction times were consistently significantly slower for the non-linguistic stimuli than for the trained linguistic ones. The reasons for this are not entirely clear. Several studies have successfully used similar trainings methods with vowel quality contrasts, and while the processing mechanisms related to the processing of duration contrasts are somewhat separated from the processing of vowel quality, it seems that they, too, can be accessed with this kind of training. However, further research with more data is required before any confident conclusions can be drawn about the results acquired thus far.

6 References

Alku, Tiitinen, Näätänen 1999 – Alku, Paavo; Tiitinen, Hannu; Näätänen, Risto. 1999. A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 110, 8, 1329–33.

Best, Strange 1992 – Best, Catherine T.; Strange, Winifred. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–30.

Boersma, van Heuven 2001 – Boersma, Paul; van Heuven, Vincent. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 9–10, 341–47.

Bradlow, Pisoni, Akahane-Yamada, Tohkura 1997 – Bradlow, Ann R.; Pisoni, David B.; Akahane-Yamada, Reiko; Tohkura, Yoh'ichi. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101, 4, 2299–2310.

Catford, Pisoni 1970 – Catford, J. C.; Pisoni, David B. 1970. Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54, 7, 477–81.

Flege 1987 – Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language : evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 1, 47–65.

Guenther, Hickok 2015 – Guenther, Frank H.; Hickok, Gregory. Role of the Auditory System in Speech Production. In Celesia, G. and Hickok, G., editors. *Handbook of Clinical Neurology 129*. Elsevier B.V, 2015, 161-175.

Guenther, Husain Cohen, Shain-Cunningham 1999 – Guenther, Frank H.; Husain, Fatima T.; Cohen, Michael A.; Shinn-Cunningham, Barbara G. 1999. Effects of categorization and discrimination training on auditory perceptual space. *The Journal of the Acoustical Society of America* 106, 5, 2900–2912.

Guion, Pederson 2007 – Guion, Susan G.; Pederson, Eric. Investigating the Role of Attention in Phonetic Learning. In Bohn, O.-S. and M. J. Munro, editors. *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*,. John Benjamins Publishing Company, 2007, 57–77.

Hirata, Whithurst, Cullings 2007 – Hirata, Yukari; Whitehurst, Elizabeth; Cullings, Emily. 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, 3837–45.

Isei-Jaakkola 2004 – Isei-Jaakkola, Toshiko. 2004. “Lexical Quantity in Japanese and Finnish.” University of Helsinki.

Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, Siebert 2003 – Iverson, Paul; Kuhl, Patricia K.; Akahane-Yamada, Reiko; Diesch, Eugen; Tohkura, Yoh’ichi; Kettermann, Andreas; Siebert, Claudia. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, 1, 47–57.

Jähi, Peltola, Alku 2015 – Jähi, Katri; Peltola, Maija S.; Alku, Paavo. 2015. Does interest in language learning affect the non-native phoneme production in elderly learners? in *Proceedings of the 18th International Congress of Phonetic Sciences*. Available online: *https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0234.pdf*.

Kirmse, Ylinen, Tervaniemi, Vainio, Schröger, Jacobsen 2008 – Kirmse, Ursula; Ylinen, Sari; Tervaniemi, Mari; Vainio, Martti; Schröger, Erich; Jacobsen, Thomas. 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, 131–43.

Kuhl, Williams, Lacerda, Stevens, Lindblom 1992 – Kuhl, Patricia K.; Williams, Karen A.; Lacerda, Francisco; Stevens, Kenneth N.; Lindblom, Björn. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–8.

Liégeois-Chauvel, De Graaf, Laguitton, Chauvel 1999 – Liégeois-Chauvel, Catherine; De Graaf, Jozina B.; Laguitton, Virginie; Chauvel, Patrick. 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex* 9, 5, 484–96.

Makhoul 1975 – Makhoul, John. 1975. Linear Prediction: A Tutorial Review. *Proceedings of the IEEE* 63, 4, 561–80.

McAllister, Flege, Piske 2002 – McAllister, Robert; Flege, James E.; Piske, Thorsten. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 2, 229–58.

Meister, Meister 2013 – Meister, Einar; Meister, Lya. 2013. Production of Estonian quantity contrasts by Japanese speakers. Pp. 330–34 in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 6. International Speech and Communication Association.

Okuno 2014 – Okuno, Tomoko. 2014. *Acquisition of L2 Vowel Duration in Japanese by Native English Speakers*. Michigan State U.

Pederson, Guion-Anderson 2010 – Pederson, Eric; Guion-Anderson, Susan. 2010. Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America* 127, 2, 54-59.

Perkell 2012 – Perkell, Joseph S. 2012. Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics* 25, 5, 382–407.

Saloranta, Tamminen, Alku, Peltola 2015 – Saloranta, Antti; Tamminen, Henna; Alku, Paavo; Peltola, Maija S. 2015. Learning of a non-native vowel through instructed production training. *Proceedings of the 18th International Congress of Phonetic Sciences*. Available online: *http://www.icphs2015.info/pdfs/Papers/ICPHS0235.pdf*

Suomi, Toivanen, Ylitalo 2008 – Suomi, Kari; Toivanen, Juhani; Ylitalo, Riikka. *Finnish Sound Structure. Phonetics, Phonology, Phonotactics and Prosody*. Oulu: Oulu University Press, 2008.

Taimi, Jähi, Alku, Peltola 2014 – Taimi, Laura; Jähi, Katri; Alku, Paavo; Peltola, Maija S. 2014. Children learning a non-native vowel – the effect of a two-day production training. *Journal of Language Teaching and Research* 5, 6, 1229–35.

Tajima, Kato, Rothwell, Akahane-Yamada, Munhall 2008 – Tajima, Keiichi; Kato, Hiroaki; Rothwell, Amanda; Akahane-Yamada, Reiko; Munhall, Kevin G. 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America* 123, 1, 397–413.

Tamminen, Peltola 2015 – Tamminen, Henna; Peltola, Maija S. 2015. Non-native memory traces can be further strengthened by short term phonetic training. *Proceedings of the 18th International Congress of Phonetic Sciences*. Available online: *http://www.icphs2015.info/pdfs/Papers/ICPHS0285.pdf*.

Tamminen, Peltola, Kujala, Näätänen 2015 – Tamminen, Henna; Peltola, Maija S.; Kujala, Teija; Näätänen, Risto. 2015. Phonetic training and non-native speech perception — New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology* 97, 1, 23–29.

Ylinen, Huotilainen, Näätänen 2005 – Ylinen, Sari; Huotilainen, Minna; Näätänen, Risto. 2005. Phoneme quality and quantity are processed independently in the human brain. *Neuroreport* 16, 16, 1857–60.

Ylinen, Shestakova, Alku, Huotilainen 2005 – Ylinen, Sari; Shestakova, Anna; Alku, Paavo; Huotilainen, Minna. 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, 313–38.