ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 12, No. 5, pp. 678-687, September 2021
DOI: http://dx.doi.org/10.17507/jltr.1205.06

Does Studying in a Music-oriented Education Program Affect Non-native Sound Learning? — Effects of Passive Auditory Training on Children's Vowel Production

Katja Immonen
Phonetics and Learning, Age & Bilingualism Laboratory (LAB-lab), Department of Computing, University of Turku, Finland

Jemina Kilpeläinen
Phonetics and Learning, Age & Bilingualism Laboratory (LAB-lab), Department of Computing, University of Turku, Finland

Paavo Alku
Department of Signal Processing and Acoustics, Aalto University, Finland

Maija S. Peltola
Phonetics and Learning, Age & Bilingualism Laboratory (LAB-lab), Department of Computing, University of Turku, Finland

***Abstract*—Earlier studies have shown that children are efficient second language learners. Research has also shown that musical background might affect second language learning. A two-day auditory training paradigm was used to investigate whether studying in a music-oriented education program affects children's sensitivity to acquire a non-native vowel contrast. Training effects were measured with listen-and-repeat production tests. Two groups of monolingual Finnish children (9–11 years, N=23) attending music-oriented and regular fourth grades were tested. The stimuli were two semisynthetic pseudo words /ty:ti/ and /tʉ:ti/ with the native vowel /y/ and the non-native vowel /ʉ/ embedded. Both groups changed their pronunciation after the first training. The change was reflected in the second formant values of /ʉ/, which lowered significantly after three trainings. The results show that 9–11-year-old children benefit from passive auditory training in second language production learning regardless of whether or not they attend a music-oriented education program.**

***Index Terms*—auditory training, children, music, pronunciation, second language learning, vowels**

I. Introduction

The primary aim of this study was to examine whether children attending a music-oriented education program in elementary school have a sensitivity to acoustic variation that is transferred to the trainability of second language (L2) sound contrasts. In a music-oriented education program, children participate in various musical activities that are not included in the regular elementary school curriculum. The program is meant for children who are interested in music and they take musicality tests before admission. We tested two groups of 9–11-year-old Finnish children from a music-oriented and a regular fourth grade with a two-day auditory training paradigm. Training effects were measured with listen-and-repeat production tests before, during and after the experiment. The passive auditory training paradigm included the Swedish close rounded vowel contrast /y/-/ʉ/. The hypothesis was that the children who attend a music-oriented fourth grade could be more sensitive to subtle acoustic differences than the children who attend a regular fourth grade. If this hypothesis is true, this sensitivity might enable them to learn to perceive and produce a non-native sound contrast more efficiently through auditory training. Since the aim of the experiment was to discover possible differences in the children's sensitivity to acoustic contrasts in L2 sounds, the training paradigm was designed to be a simple listening task without any production training or articulatory instructions. This was to ensure that any changes in articulation found in the production tests would be the effect of the auditory training.

II. Literature Review

A. Models of Second Language Learning

The phonetic learning of a second language (L2) poses the learner with the challenge of acquiring non-native

perception and production patterns that do not necessarily comply with the sound categories of their native language (L1). Studies in cross-language sound categorization have offered evidence that problems in non-native sound perception and production arise from the relative differences and similarities between the phonological systems of the speaker's L1 and L2 (e.g. Best & Strange, 1992; Best, 1994; Flege, 1987; Flege, Munro, & Mackay, 1995). These cross-language studies have led to the emergence of several theoretical frameworks of second language learning, such as the Speech Learning Model (SLM, Flege, 1987, 1995), the Perceptual Assimilation Model (PAM, Best, 1994, 1995) and the Second Language Linguistic Perception model (L2LP, Escudero, 2005). These theoretical models take a comparative approach to the phonetic and phonological differences between languages, aiming to explain and predict the possible problems that may arise when learning a new language.

According to the SLM (Flege, 1995), the degree of similarity or dissimilarity between the L1 and L2 sounds determines the degree of difficulty that the speaker is faced with when learning new speech sounds. In other words, in order to learn to perceive and produce L2 sounds, speakers need to learn to adapt the phonetic categories of their L1. Compared to the native sound system, non-native sounds can be perceptually either identical, similar or new. The SLM proposes that the probability of a new L2 category to be formed increases linearly as the perceived similarity between L1 and L2 sounds decreases. When an L2 sound is identical to an L1 category, the formation of a new L2 category is not necessary and therefore unlikely. On the other hand, when an L2 sound is completely new and does not resemble any L1 categories, the formation of a new sound category is more probable, but requires input and practice. Similar L2 sounds that resemble a native sound to some extent, but are still distinct from all L1 categories, are considered to be the most difficult to learn, since category formation is less easily achieved but necessary for accurate perception and production of the L2 sound. In addition, the SLM proposes that perception precedes and guides production in L2 speech learning. In other words, accurate L2 perception does not require accurate production, but accurate production does require accurate perception (Flege, 1995, 1999; Flege, MacKay, & Meador, 1999).

The PAM (Best, 1994, 1995; Best & Tyler, 2007) takes a slightly different approach to cross-language category perception, though it shares the comparative aspect of the SLM. Instead of concentrating on the perception of individual L2 sounds, the PAM approaches the issue of L2 category perception through examining non-native sound contrasts. The PAM aims to explain L2 sound perception through assimilation patterns, which describe the degree to which L2 sounds can be assimilated to one or several L1 sound categories. The PAM suggests that the equal assimilation of two L2 sounds to one L1 category (single-category assimilation) is most likely to cause persistent difficulties in L2 perception and production.

A more recent model for L2 sound perception is the L2LP established by Escudero (2005). The L2LP model proposes that L2 perception and production requires learners to either create new perceptual mappings for L2 sounds, or adjust the existing perceptual mappings of their L1 to fit the L2 phonological representations. The L2LP model predicts that a learning situation where two or more sounds are perceptually mapped to one L1 category causes challenges in L2 perception and production. This is because the learner is required to create new perceptual representations for these sounds and possibly adjust existing overlapping perceptual L1 categories. The L2LP refers to these L2 sounds as new but the learning task is comparable to the classification of similar sounds described by the SLM.

To summarize, the SLM, the PAM and the L2LP all propose that the close resemblance of an L2 sound with an L1 sound category potentially causes major difficulties in L2 perception and production. These models of L2 phonetic learning provide the theoretical basis for stimulus selection in the present study.

B. The Close Rounded Vowels of Finnish and Swedish in the Light of L2 Learning Models

To ensure that the stimuli used in the present study would represent a theoretically difficult learning situation for monolingual Finnish speakers, the vowel categories of Finnish and Swedish were viewed in the light of the L2 learning models (SLM, PAM, L2LP). The Finnish close rounded vowel continuum is divided into two categories /y/–/u/, whereas Swedish has three close rounded vowel categories /y/–/ʉ/–/u/ in the same vowel space. The Swedish vowel /ʉ/ is therefore situated on the border of Finnish /y/ and /u/ categories. This makes /ʉ/ difficult for Finnish speakers to perceive and produce, as it is perceptually relatively close to both of the native categories and can be categorized to either one or both according to the SLM (Flege, 1995), the PAM (Best, 1994, 1995; Best & Tyler, 2007) and the L2LP (Escudero, 2005). Based on these L2 learning theories, it can be assumed that the perceptual challenge of categorizing the non-native sound is also reflected in production. This means that the accurate pronunciation of the Swedish vowel /ʉ/ also requires the accurate perception of the L2 vowel contrast by Finnish speakers. Therefore, the SLM's proposition that L2 perception precedes L2 production is an essential part of the present study, as the experiment was designed to measure auditory sensitivity to L2 sound contrasts by examining production learning through perceptual training. To be more precise, the training paradigm of the present study is designed to measure whether perceptual sensitivity to acoustic differences is transferred to the trainability of an L2 vowel contrast, and whether auditory training affects the articulation of the trained vowels.

C. Second Language Perception and Production Studies on Children and Adults

The processes and mechanisms of L2 learning have been studied widely in different age groups, and especially the comparison of child and adult learners has received considerable attention. For example, a study investigating the perception and production of English vowels by native Korean adults and children found that child learners were able to

produce and discriminate L2 vowels more accurately than adult learners (Tsukada et al., 2005). These findings were later supported by another study that measured the pronunciation of English vowels by Japanese adults and children living in the United States (Oh et al., 2011). The participants' productions were recorded on two occasions: shortly after arrival in the USA and one year later. The results showed that children reached higher production accuracy than adults when their pronunciation was compared to age-matched native speakers of English. Furthermore, a phonetic training study by Giannakopoulou, Uther and Ylinen (2013) revealed that high-variability perceptual training (HVPT) improved L2 vowel identification and discrimination both in Greek adults and children (7–8 years), but the degree of improvement was more pronounced for the child learners. The authors suggest that these results indicate enhanced plasticity for language learning at this developmental stage (Giannakopoulou et al., 2013).

These findings have also been supported by other experimental training studies on children. For example, a production training study by Taimi, Jäähi, Alku and Peltola (2014) examined how a simple listen-and-repeat training affects 7–10-year-old children's production of an L2 vowel. The results showed that children changed their pronunciation of the L2 sound after three short training sessions. This indicates that children are able to adapt their existing L1 production patterns quickly towards an acoustic model through phonetic listen-and-repeat training. Another training study on adult learners showed that adults did not benefit from auditory training of an L2 vowel contrast, but their perception and production results did improve with a two-day listen-and-repeat training paradigm (Peltola, Tamminen, Alku, Kujala & Peltola 2020). However, results from an earlier study by Peltola, Rautaoja, Alku and Peltola (2017) showed that L1 Finnish and L1 English speaking adults did not learn to produce an L2 vowel contrast after a one-day listen-and-repeat training paradigm. These results suggest that even though adults can improve their perception and production of L2 sounds through motoric training, the amount of training affects training results.

Overall, these findings offer evidence that children benefit efficiently both from experimental L2 training paradigms as well as naturalistic L2 input. The findings on adult learners show that they are often less successful L2 production learners than children in naturalistic L2 learning settings. On the other hand, adults can benefit from some types of phonetic training, but changes in perception and production are not as rapid or as pronounced as in child learners.

D. Music and Second Language Learning

The relationship between music and language has received considerable attention from scientists in recent years, but there are many different ways of defining and measuring musicality. However, the present study does not focus on musicality or any specific area of musical ability. Instead, we examine musical experience from a more general perspective by testing children who study in a music-oriented education program. There is little or no previous research on how music-oriented programs or musical experience in general might affect children's L2 perception and production learning. Nonetheless, there are some relevant findings from earlier studies that have examined the relationship between music and L2 learning from slightly different perspectives.

Studies on the effects of musical background factors on L2 perception are of particular interest for the present study. For example, Marie, Delogu, Lampis, Belardinelli and Besson (2011) investigated the perception of tonal and segmental variations in Mandarin Chinese by French musicians and non-musicians with behavioral discrimination tests and electroencephalography (EEG) measurements. The musicians showed better discrimination accuracy for both tonal and segmental variations than non-musicians. Furthermore, the event-related potential (ERP) results revealed that the tonal variations elicited an earlier N2/N3 response in musicians than non-musicians and that the musicians also showed enhanced P3b components for the tonal and segmental variations. Another study on the perception of lexical tones of Mandarin Chinese (Delogu, Lampis, & Belardinelli, 2006) investigated the connection between melodic ability and tone discrimination in L1 Italian speakers. The overall results showed that the participants were better at identifying phonological variation than tonal variation, but speakers with high melodic ability performed better in tonal discrimination tasks. However, melodic ability did not influence phonological discrimination. Delogu et al. (2006) conclude that, in spite of the different role of pitch variations in music and language, it seems that a music-to-language transfer effect does occur. Ghaffarvand Mokari and Werner (2018) examined whether musical ability affects participants' response to high variability intensive phonetic (HVIP) training of L2 vowels. Vowel discrimination and production were measured in pre- and post-tests. The participants were L1 Azerbaijani speakers who were trained with British English vowels. The results showed no connection between overall musical ability and L2 perception and production, but there was a connection between discrimination accuracy and tonal memory.

Contrary to the results obtained by Ghaffarvand Mokari and Werner (2018), Bhatara, Yeung and Nazzi (2015) did not find a connection between L2 experience and melody perception when they tested the correlations between L2 learning and different areas of music perception in 147 French speakers. However, their results showed positive correlations between rhythm perception and L2 experience as well as rhythm perception and music training. According to Bhatara et al., both music training and L2 learning are related to native French speakers' perception of rhythm but not melody, and the results suggest a common perceptual basis for rhythm in language and music (2015). Furthermore, another study by Boll-Avetisyan, Bhatara, Unger, Nazzi and Höhle, (2016) found that adult native French listeners' sensitivity to rhythm can be enhanced through music and L2 experience. The experiment tested forty French late learners of German who participated in a rhythmic grouping task where they listened to sequences of co-articulated syllables that varied in intensity or duration. The results showed that musical experience as well as L2 input quality and quantity influenced grouping preferences.

The interplay between music and L2 learning has also been investigated in children, though from a very different perspective than in the present study. For example, a study by Milovanov, Huotilainen, Välimäki, Esquef, and Tervaniemi (2008) examined the relationship between musical aptitude and L2 production. The participants were Finnish children (aged 10–12 years) who were divided into two groups according to their English pronunciation skills. When tested with a Seashore musicality test (standardized musical aptitude listening test; Seashore, Lewis, & Saetveit, 2003), the children with more advanced L2 pronunciation skills obtained higher musical aptitude scores than children with less advanced L2 pronunciation skills for pitch discrimination, timbre, sense of rhythm and sense of tonality. In addition, the pre-attentive processing of chords was measured with EEG and the results showed that children with advanced L2 pronunciation skills had more pronounced sound-change evoked activation with music stimuli than children with less accurate L2 linguistic skills. Milovanov et al. (2008) conclude that it appears that musical and linguistic skills might partly be based on shared neural mechanisms.

Taken together, the results from earlier studies on the relationship between music and language provide evidence that there is some interplay and overlap between some musical factors and phonetic L2 processing. The results by Marie et al. (2011), Delogu et al. (2006), and Ghaffarvand Mokari and Werner (2018) indicate that musical expertise as well as melodic and tonal abilities might be connected to the perception of tonal and segmental variations in speech sounds. On the other hand, the findings of Bhatara et al. (2015) and Boll-Avetisyan et al. (2016) indicate connections between rhythm perception, musical expertise and L2 experience. Most importantly, the results of Milovanov et al. (2008) suggest that children's L2 production skills and musical aptitude might be connected. However, the effects of musical experience on L2 perception and production learning remain somewhat unclear. As the models of L2 phonetic learning propose, difficulties in L2 sound production are often caused by difficulties in L2 sound perception. Therefore, if there is a difference in auditory sensitivity to sounds between children from a music-oriented and a regular education program, it should be reflected in the production tests.

III. Methods

A. Participants

Altogether 25 monolingual Finnish speaking children from two elementary schools in Southwest Finland participated in the study. However, two of the children did not complete the experiment and they had to be excluded from the data. Therefore, a total of 23 children (aged 9;10–11;2 years, mean age 10;5, 20 females) were tested. The participants were divided into two groups: a Music group and a Non-music group. The two groups were compared to see whether the amount of musical activities in school would affect children's training results.

The Music group included 11 children (aged 9;10–10;9 years, mean age 10;4, ten females) from a music-oriented fourth grade. Children in a music-oriented education program participate in various musical activities, such as solo and choir singing, playing different instruments, listening to music across genres as well as preparing and performing musical shows and productions. To be admitted, children have to take a test that measures their musical abilities. The music-oriented education program starts at the beginning of the third grade and continues throughout elementary school, until the sixth grade. At the time of the experiment, the children were in their second year of the program.

The Non-music group included 12 children (aged 10;1–11;2, mean age 10;7, ten females) from regular fourth grades from two elementary schools. The group represented typical monolingual Finnish children who have one compulsory music lesson per week, as dictated by the national core curriculum.

Before participating in the experiment, the children and their parents gave informed written consent and answered to a language background questionnaire. The questionnaire was used to ensure that none of the children knew any Nordic languages and that they had not lived in any Nordic countries outside Finland. All participants had studied English basics in school for a little over a year. All children reported to have normal hearing. None of the children reported having speech defects, except for one participant who reported having had minor difficulties in the production of /r/ in early childhood. This participant was not excluded from the experiment as no words containing the sound /r/ were used in the experiment.

B. Stimuli

The stimuli were two semisynthetic pseudo words /ty:ti/ and /tʉ:ti/ with the close rounded vowels /y/ and /ʉ/ embedded in the first syllable. The stimuli were created using the Semisynthetic Speech Generation method (SSG, Alku, Tiitinen, & Näätänen, 1999). The natural speech productions of a 24-year-old Finnish-Swedish bilingual male speaker were used as the basis of the stimuli. The glottal excitation waveform was first extracted from the natural speech signal. This waveform was then used to excite a digital vocal tract model with a desired formant structure to create the semisynthetic word pair /ty:ti - tʉ:ti/. The first formant (F1) value for the non-native vowel /ʉ/ in the stimulus word /tʉ:ti/ was 338 Hz and the second formant (F2) value was 1258 Hz. The F1 and F2 values for the native vowel /y/ in /ty:ti/ were 269 Hz and 1866 Hz respectively. Therefore, the primary acoustic difference between the vowels /y/ and /ʉ/ lies in the F2.

For the purpose of this study, any changes in the F2 towards the acoustic model /tʉ:ti/ will be considered as a sign of learning, since the non-native vowel /ʉ/ has considerably lower F2 values than /y/, due to more backed tongue position

during articulation. However, the pitch (F0) of the male voice used in the stimuli is considerably lower than that of a child, and the F1 and F2 values of the first syllable vowels in the stimuli were those typical for an adult male speaker. Therefore, the child participants in this study are not expected to reach the exact formant values of the stimuli in their own productions. The focus of the experiment is on the direction of the possible change in the participants' pronunciation.

C. Procedure

The procedure was a short training paradigm consisting of four alternating recording and training sessions on two consecutive days. The experiment was conducted during school hours. The participants were tested in a quiet room using a portable laboratory consisting of an HP laptop computer with a Beyerdynamic MMX300 headset and an Asus Xonar U3 sound card. The auditory stimuli were presented automatically in an alternating order during recording and training sessions with Sanako Study Recorder software (version 8.22.0.0) with an interstimulus interval (ISI) of 3 seconds. The same software was also used to record the participants' productions. During recording and training sessions, the stimuli were presented in turns in a fixed order, so that every other word was /tʉ:ti/ with the non-native vowel /ʉ/. During recording sessions, the participants listened and repeated each stimulus word ten times. During training sessions, they listened to each stimulus word 30 times without repeating them. Both days of the experiment consisted of two recording sessions and two training sessions. Overall, the participants heard each stimulus word 120 times during trainings and produced them 40 times during recordings.

The first day of the experiment started with a short familiarization, where the participants heard both stimuli three times. The purpose of the familiarization phase was to allow the children to adjust the volume to a comfortable level and get accustomed to the pace of the experiment. After familiarization, the first day continued with the first recording (baseline) followed by the first training, then a second recording and a second training. On the second day the experiment proceeded in reverse order, in other words the day began with a third training session followed by a third recording, a fourth training and ended with the final recording. The experiment lasted around 15 minutes per day. The children were instructed to listen carefully to the auditory stimuli without repeating them during training sessions, whereas during recordings they were instructed to listen and repeat what they heard. The procedure was designed not to include any production during training, because the aim was to test the children's auditory perception skills and their ability to adapt their own production through merely listening to an L2 sound contrast. For a summary of the experiment procedure, see Table I.

TABLE I
THE EXPERIMENT PROCEDURE

	Day 1		Day 2
Listen and repeat	1st Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →Recorded	*Listen*	3rd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →Not recorded
Listen	1st Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →Not recorded	*Listen and repeat*	3rd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →Recorded
Listen and repeat	2nd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →Recorded	*Listen*	4th Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →Not recorded
Listen	2nd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →Not recorded	*Listen and repeat*	4th Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →Recorded

The stimuli were presented in a fixed order during trainings and recordings, so that every other word was /tʉ:ti/ and every other word was /ty:ti/.

D. Analysis

The production data was acoustically analyzed using Praat software version 6.0.43 (Boersma & Weenink, 2020). The first and second formants were measured from the steady state phase of the first syllable vowels using the Linear Predictive Coding (LPC) Burg algorithm. The participants produced both words ten times during each of the four recording sessions and all these productions were analyzed. Altogether 920 productions of /ty:ti/ and 920 productions of /tʉ:ti/ were analyzed (a total of 1840 tokens), of which 440 repetitions per word were produced by the Non-music group and 480 by the Music group. After acoustic analysis, the speakers' individual average formant values for /y/ and /ʉ/ from the ten repetitions within each recording session were calculated. The F1 and F2 values of the two vowels in all four sessions were then subjected to statistical analysis using IBM SPSS Statistics (version 25.0.0.1) software. During the analysis, special attention was paid to any changes in the F2 values of the children's productions.

IV. RESULTS

A repeated measures Analysis of Variance (ANOVA) was performed for the average formant values with the between-subject factor defined as Group (Music, Non-music) and the within-subject factors defined as Session (first, second, third, fourth), Word (/ty:ti/, /tʉ:ti/) and Formant (F1, F2). The initial ANOVA was performed in order to see whether the groups differed in any way in their productions across recording sessions. The main effects of Formant are not reported, because the F1 and F2 values are expected to differ automatically from each other. The analysis revealed the main effect of Word ($F(1,21)=32,419$, $p<0.001$). The main effect of Word suggests that the participants had produced the target and non-target vowels differently, in other words that the vowels had not been assimilated to a single sound category. The initial analysis also revealed a Word × Formant interaction ($F(1,21)=35,461$, $p<0.001$), which means that the formants of the target and non-target vowels were produced differently in the two words. Furthermore, a Session × Formant interaction ($F(3,19)=4,314$, $p=0.018$) was discovered, which suggests that the formants F1 and F2 developed differently between sessions. More importantly, the interaction between Session and Formant indicates that there is a significant change in the F1 or F2 values of the vowels between sessions, meaning that there is some change in the children's productions across time. The initial ANOVA did not reveal any significant differences between the two groups. In other words, even though there seems to be a slight difference in the F2 values of /ʉ/ produced by the two groups (Fig. 1), the difference did not reach significance.

The interaction between Session and Formant was investigated further with a Group (2) × Session (2) ×Word (2) × Formant (2) repeated measures ANOVA with the same factors as in the initial analysis, but only two sessions tested at a time. The first recording session (baseline) was compared to the second, third and fourth sessions separately. The comparison of the first and second sessions revealed the main effects of Session ($F(1,21)=4,681$, $p=0.042$) and Word ($F(1,21)=27,026$, $p<0.001$). In addition, Session × Formant ($F(1,21)=7,082$, $p=0.015$) and Word × Formant ($F(1,21)=30,043$, $p<0.001$) interactions were found. Next, the main effects of Session ($F(1,21)=8,678$, $p=0.008$) and Word ($F(1,21)=29,109$, $p<0.001$), as well as Session × Formant ($F(1,21)=9,569$, $p=0.006$) and Word × Formant ($F(1,21)=33,298$, $p<0.001$) interactions were discovered when comparing the first and third sessions with the same ANOVA. Comparison of the first and fourth sessions revealed the same main effects of Session ($F(1,21)=8,533$, $p=0.008$) and Word ($F(1,21)=30,394$, $p<0.001$), as well as Session × Formant ($F(1,21)=8,222$, $p=0.009$) and Word × Formant ($F(1,21)=31,503$, $p<0.001$) interactions. To summarize, the same main effects and interactions were found in all three session pairs. This finding shows that there is a change in the participants' productions already after the first training session and that the change remains throughout the experiment. The Session × Formant interactions suggest that the F1 or F2 values somehow differed between the sessions.

Paired samples t-tests for both words' F1 and F2 values in all three session pairs were performed to see how the formants developed in the second, third and fourth sessions compared to the baseline. Significant differences in the target vowel's F2 values (Fig. 2) were found between the first and third sessions ($t(22)=2,842$, $p=0.009$) as well as the first and fourth sessions ($t(22)=3,206$, $p=0.004$). There were no significant changes in the F1 values of the target word. In addition, no significant changes were found in the F1 or F2 values of the vowel /y/ in the non-target word /ty:ti/.

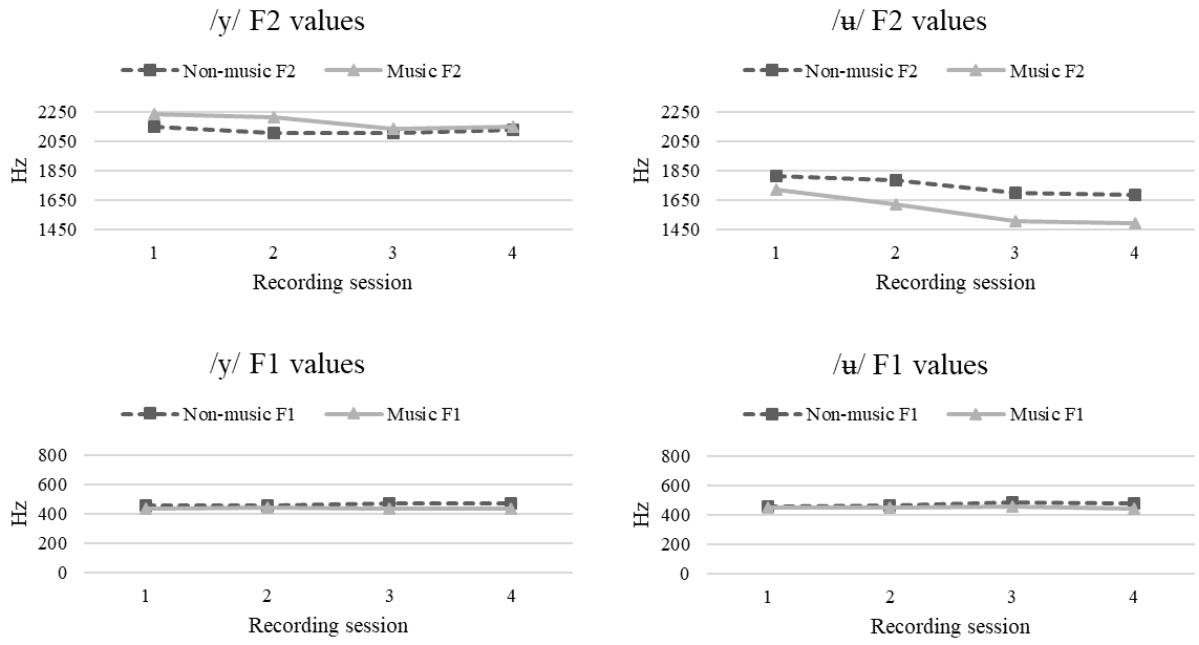

Figure 1. The Average F1 and F2 values for both vowels produced by the two groups.

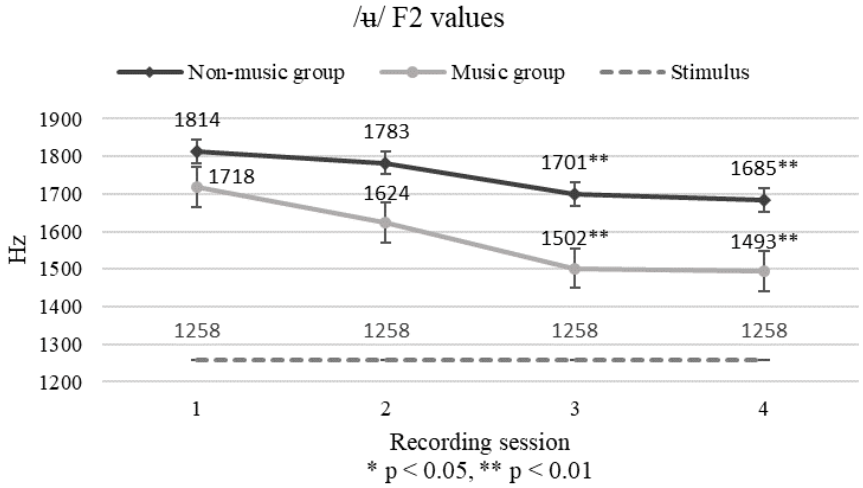

Figure 2. Both groups' average F2 values for the vowel /ʉ/ across recording sessions. The dashed line indicates the F2 of the acoustic stimulus. Paired samples t-tests revealed that the /ʉ/ F2 values lowered significantly between the first and third recording session. Significant between-session changes are marked with asterisks. No significant changes emerged for the vowel /y/ or the F1 values in /ʉ/.

V. Discussion

The results show that both groups benefitted from the auditory training paradigm and that their production of /ʉ/ changed as a function of training by the second recording session. The change in production was also reflected in the F2 values of /ʉ/, which lowered significantly by the third recording session. The fact that there were no changes in the native vowel /y/ shows that the children were able to perceive and produce the vowels as two separate sounds. Their production of /y/ remained stable while they adjusted their production of the difficult non-native sound /ʉ/ to fit the acoustic model. The results indicate that the 9–11-year-old children in this study were able to distinguish subtle acoustic differences in vowel qualities, even when the difference is not relevant in their L1. Moreover, the results show that they can rapidly modify their own articulation patterns towards the acoustic target. Contrary to the children tested in the present experiment, adult learners tested by Peltola et al. (2020) did not benefit from auditory training of an L2 vowel contrast. In addition, the findings of Taimi et al. (2014) showed that 7–10-year-old children changed their production of an L2 vowel already after three listen-and-repeat training sessions. Comparing our findings to these results further supports the proposition that children benefit rapidly and more efficiently than adults from auditory phonetic L2 training paradigms. In addition, our results suggest that children are sensitive to acoustic differences in non-native sounds and benefit from auditory training regardless of whether or not they attend a music-oriented education program in elementary school.

The statistical analysis revealed that the participants changed their pronunciation already in the second recording session and produced the non-native vowel /ʉ/ with significantly lower F2 values in the third recording session. The fact that the main effect of Session was found already between the first and second recording sessions indicates that there was a significant change in production already after the first training, even though there were no significant changes in the F2 values of /ʉ/ until the third recording session. The fact that the main effect of Session found between the first and second sessions was not reflected in the F2 analysis is most probably explained by the rather large overall standard deviations (Table II). This can be interpreted as a reflection of the learning process being in progress. This means that the children were already making changes in their pronunciation in the second session, but were not yet consistent in their articulation of the non-native vowel /ʉ/. This implies that the changes in the children's production were immediate, but they still needed to practice their articulation before producing the difference consistently in their own speech. As the F2 value in vowels is usually related to tongue backness or lip rounding during articulation, the change in the children's F2 values implies that they started to move their tongue backwards or rounded their lips more when articulating the vowel /ʉ/ during recordings. However, both /y/ and /ʉ/ are rounded vowels and the Finnish close rounded vowel space also has the backed rounded vowel /u/. Therefore, it is more likely that the children produced the difference between the vowels by articulating /ʉ/ with a more backed tongue position. In addition, the formant values in Table II show that the participants did not produce the non-native vowel /ʉ/ as the Finnish close back rounded vowel /u/, since their F2 values remained considerably higher than those of a typical Finnish /u/ (F1=332 Hz and F2=690 Hz; Iivonen, 2012).

TABLE II
THE AVERAGE VOWEL FORMANT VALUES (HZ) PRODUCED BY THE GROUPS IN EACH RECORING SESSION

			Session 1	Session 2	Session 3	Session 4
/ʉ/	Non-music	F1	454 (34)	461 (41)	481 (28)	474 (36)
		F2	1814 (540)	1783 (523)	1701 (509)	1685 (508)
	Music	F1	445 (31)	449 (28)	456 (58)	439 (35)
		F2	1718 (383)	1624 (362)	1502 (403)	1493 (416)
	Both groups	F1	450 (32)	455 (35)	469 (46)	457 (39)
		F2	1768 (463)	1706 (450)	1606 (462)	1593 (466)
/y/	Non-music	F1	460 (33)	458 (35)	474 (30)	470 (30)
		F2	2148 (176)	2104 (188)	2103 (162)	2127 (210)
	Music	F1	438 (32)	441 (29)	438 (45)	437 (51)
		F2	2232 (158)	2215 (144)	2135 (242)	2149 (230)
	Both groups	F1	449 (34)	450 (33)	457 (41)	454 (44)
		F2	2188 (169)	2157 (174)	2118 (200)	2137 (215)

Both words were repeated ten times in each session by all speakers. The standard deviations are reported in parentheses.

Our hypothesis was that the Music group would have an enhanced auditory sensitivity to vowel quality differences due to musical experience received in the music-oriented education program or their musical abilities in general. Our hypothesis was based on previous results from studies on the connection between musicality and L2 processing (e.g. Marie et al. 2011; Milovanov et al. 2008). However, the hypothesis was not confirmed. The results indicate that studying in a music-oriented education program did not affect L2 production learning facilitated by auditory training in the children tested in the current experiment. There are two possible explanations that need to be discussed in order to understand this finding.

Firstly, since the sample size of this study is comparable to other studies in the same field, it is unlikely that the individual differences in the data would affect the result to a great extent. However, it is possible that the tentative group difference in the F2 values of /ʉ/ seen in Fig. 2 could reach significance with a larger sample size. To test this possibility, additional data would need to be collected to both groups.

Most importantly, it is possible that the participants were at a linguistically sensitive age, making them naturally efficient learners of L2 sounds. The finding that the two groups responded similarly to training indicates that their perceptual abilities were naturally accurate enough to distinguish subtle acoustic differences in speech sounds. Therefore, it is plausible that at this developmental stage, the benefits of age are greater than the possible benefits of musical experience on L2 sound production learning through auditory training. This hypothesis is supported by earlier findings showing that children are more successful and efficient L2 learners than adults (e.g. Giannakopoulou et al., 2013; Oh et al., 2011; Tsukada et al., 2005) due to their young age and plasticity. Therefore, we propose that in the present experiment, the effects of age outweighed the possible effects that studying in a music-oriented program might have on L2 sound learning. Musical experience and musical aptitude might have a greater effect on the perception and production of L2 sounds in adult or adolescent learners, who are no longer at the same linguistically sensitive developmental stage as 9–11-year-old children. This question could be explored further by recreating the experiment with different age groups to see whether exposure to music affects auditory L2 training results for instance in adolescent or adult learners.

VI. CONCLUSIONS

The results show that the 9–11-year-old children examined in this study had a sensitivity to acoustic differences in L2 sound contrasts and can change their production of a difficult L2 vowel after just one session of passive auditory training. The results indicate that attending a music-oriented education program does not affect children's L2 production learning. This suggests that, at least at the developmental stage of the children tested in the current study, the benefits of linguistic plasticity and age may outweigh the possible benefits of musical experience on L2 sound learning. However, further research is needed to draw more definite conclusions on the effects of music on L2 perception and production learning in different age groups.

ACKNOWLEDGEMENTS

The authors would like to thank all the children and their parents for willingness to participate, as well as Sanako Corp. for sponsoring the software used for data collection.

REFERENCES

[1] Alku, P., H. Tiitinen & R. Näätänen. (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, 1329–1333.

[2] Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman, & H. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press, 167–224

[3] Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language speech research*. Baltimore: York Press, 171–206.
[4] Best, C. T. & W. Strange (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–330.
[5] Best, C. T. & M. Tyler (2007). Nonnative and second-language speech perception. In O. Bohn, & M. J. Munro (eds.), *Language experience in second language speech learning: In honour of James Emil Flege*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 13–34.
[6] Bhatara, A., H. H. Yeung & T. Nazzi. (2015). Foreign language learning in French speakers is associated with rhythm perception, but not with melody perception. *Journal of Experimental Psychology: Human Perception and Performance* 41.2, 277–282.
[7] Boersma, P. & D. Weenink. (2020). Praat: doing phonetics by computer [Computer program].Version 6.0.43. https://www.fon.hum.uva.nl/praat/ (no date).
[8] Boll-Avetisyan, N., A. Bhatara, A. Unger, T. Nazzi & B. Höhle. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition* 19.5, 971–986.
[9] Delogu, F., G. Lampis & M. O. Belardinelli. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing* 7.3, 203–207.
[10] Escudero, P. (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. Utrecht: Netherlands Graduate School of Linguistics.
[11] Flege, J. 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.
[12] 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*. Timonium, MD; York Press, 229–273.
[13] Flege, J. E. (1999). The relation between L2 production and perception. *Paper presented at the Proceedings of the XIVth International Congress of Phonetics Sciences*, 1273–1276.
[14] Flege, J. E., I. R. A. MacKay & D. Meador. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106.5, 2973–2987.
[15] Flege, J. E., M. J. Munro & I. R. A. MacKay. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16.1, 1–26.
[16] Ghaffarvand Mokari, P. & S. Werner (2018). Perceptual training of second-language vowels: Does musical ability play a role? *Journal of Psycholinguistic Research* 47.1, 95–112.
[17] Giannakopoulou, A., M. Uther & S. Ylinen (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy* 29.2, 201–218.
[18] Iivonen, A. (2012). Kielten vokaalit kuuloanalogisessa vokaalikartassa. *Puhe Ja Kieli* 1, 17–43.
[19] Marie, C., F. Delogu, G. Lampis, M. O. Belardinelli & M. Besson. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience* 23.10, 2701–2715.
[20] Milovanov, R., M. Huotilainen, V. Välimäki, P. A. Esquef & M. Tervaniemi. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research* 1194, 81–89.
[21] Oh, G. E., S. Guion-Anderson, K. Aoyama, J. E. Flege, R. Akahane-Yamada & T. Yamada. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics* 39.2, 156–167.
[22] Peltola, K. U., T. Rautaoja, P. Alku & M. S. Peltola. (2017). Adult learners and a one-day production training–Small changes but the native language sound system prevails. *Journal of Language Teaching and Research* 8.1, 1–7.
[23] Peltola, K. U., H. Tamminen, P. Alku, T. Kujala & M. S. Peltola. (2020). Motoric training alters speech sound perception and production—Active listening training does not lead into learning outcomes. *Journal of Language Teaching and Research* 11.1, 10–16.
[24] Seashore, C. E., D. Lewis & J. G. Saetveit. (2003). Seashore measures of musical talents CD, digitally remastered version by esquef, P. Finland: Helsinki University of Technology.
[25] Taimi, L., K. Jähi, P. Alku & M. S. Peltola. (2014). Children learning a non-native Vowel–The effect of a two-day production training. *Journal of Language Teaching and Research* 5.6, 1229–1235.
[26] Tsukada, K., D. Birdsong, E. Bialystok, M. Mack, H. Sung & J. E. Flege. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33.3, 263–290.

Katja Immonen, MA, is a PhD candidate in Phonetics at the Learning, Age & Bilingualism Laboratory (LAB-lab), University of Turku, Finland. Her main research is focused on the effect of different language learning backgrounds on children's second language learning. Her research interests also include children's second language perception and production.

Jemina Kilpeläinen, MA, is a PhD candidate in Phonetics at the Learning, Age & Bilingualism Laboratory (LAB-lab), University of Turku, Finland. Her research is focused on the effect of audiovisual learning cues on non-native language learning in different types of adult learners. Her research interests also include the relationship between language learning and music.

Paavo Alku, Dr. Tech, is an Academy Professor from Department of Signal Processing and Acoustics, Aalto University, Finland. His research interests include a wide range of topics related to the speech communication technology, e.g. analysis and

parameterization of voice production and HMM-based speech synthesis.

Maija S. Peltola, PhD, is a Professor and the head of Phonetics and Learning, Age & Bilingualism Laboratory at the University of Turku, Finland. Her research interests include a wide range of topics related to the perceptual and productional acquisition of non-native speech.