| | |
|---|---|
| AUTHOR | Katja Immonen, Paavo Alku, Maija S. Peltola |
| TITLE | Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel contrast production after one training session |
| YEAR | 2022 |
| DOI | https://doi.org/10.1075/jslp.21005.imm |
| VERSION | Author's accepted manuscript |
| CITATION | Katja Immonen, Paavo Alku and Maija S. Peltola 2022, Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel contrast production after one training session. Journal of Second Language Pronunciation. https://doi.org/10.1075/jslp.21005.imm |

**Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel contrast production after one training session**

Katja Immonen[1], Paavo Alku[2], Maija S. Peltola[1]

[1]University of Turku, [2]Aalto University

**Abstract**

The present study investigated children's ability to learn to produce a non-native vowel contrast through a listen-and-repeat training method that is traditionally used in foreign language classrooms. Sixteen Finnish preschoolers (aged 6–7 years) were tested. The stimuli were two semi-synthetic pseudo words with the familiar vowel /y/ and the novel vowel /ʉ/ embedded in the first syllable. The procedure included four training and four recording sessions on two consecutive days. The vowels produced by the children were acoustically analyzed to obtain the average values of the first and second formant. The results showed that the participants changed their production of /ʉ/ towards the acoustic model after the first training and the change remained throughout the experiment. Our findings suggest 6–7-year-old children learn to produce a non-native vowel contrast even with limited L2 sound exposure in a listen-and-repeat training setting.

**Keywords**

**1. Introduction**

The purpose of this study is to investigate the effects of listen-and-repeat training on Finnish preschoolers' (aged 6–7 years) production of a non-native vowel contrast, and to determine the extent to which early L2 exposure in a classroom-like setting affects L2 speech learning. More specifically, the study seeks to answer the following

research questions. First, can 6–7-year-old children learn to produce a difficult non-native sound contrast through listen-and-repeat training in four short training sessions? Second, how fast and in what direction does the change occur, if the children's production of the L2 sound contrast changes as a function of training? The results of the current experiment are compared to earlier findings from studies that tested 7–10-year-old children (Taimi et al., 2014) and adults (Peltola, K. U. et al., 2017, 2020) with different amounts of listen-and-repeat training using the same stimuli. This allows us to obtain information on how age of learning (AOL) affects L2 production learning outside a naturalistic L2 environment with a listen-and-repeat method that is traditionally used in classroom teaching.

Earlier research has shown that children are often more successful in second language (L2) speech learning than adults at least in naturalistic learning contexts (e.g. Tsukada et al., 2005; Oh et al., 2011) and in phonetic training settings (e.g., . Giannakopoulou, Uther, & Ylinen, 2013). These research findings have informed the educational reform of L2 teaching in Finnish elementary schools; since 2020 all Finnish children start to learn their first non-native language (usually English, Swedish, French, German or Russian) already in the first grade, at the age of seven, because an early onset of L2 learning is thought to result in better language skills later in life (Pyykkö, 2017). After starting their first L2 studies in the first grade, children start to learn at least one other non-native language in the fourth or fifth grades. Because Swedish is Finland's second official language (L1 for 5.2% of the population; Official Statistics of Finland, 2018), one of the languages introduced in first, fourth of fifth grades has to be Swedish, but the order in which a child starts to study each language is optional. Before the implementation of the new national core curriculum, Finnish children began their

first compulsory L2 (English) lessons at the age of nine in the third grade. Compulsory Swedish and other optional foreign languages (e.g., German, French or Russian) were studied after the fourth grade. There is an ongoing discussion on the possibility of introducing L2 teaching and L2 immersion education methods to preschools. This means that in the future, Finnish children might begin to learn their first L2 already at the age of six, before they start elementary school. Since most of the findings on the effects of AOL on L2 learning are from immigrant or bilingual populations, there is not enough research on how early L2 teaching and exposure in instructed L2 environments affect children's L2 perception and production.

Major theories into L2 speech learning predict that native language (L1) phonological categories strongly influence L2 category perception and production (e.g., Kuhl, 1991; Kuhl, Williams, & et al, 1992; Best & Strange, 1992; Best, 1994, 1995; Flege, 1987, 1995; Flege, Munro, & MacKay, 1995). For example, the Speech Learning Model (Flege, 1987, 1995; Flege & Bohn, 2021) and the Perceptual Assimilation Model (Best, 1994, 1995; Best & Tyler, 2007) are established models of L2 learning that take a comparative approach to the phonetic and phonological differences between languages. The SLM and PAM provide the theoretical basis for the stimulus selection of the current study to ensure that the L2 sound contrast used in the training paradigm would represent a theoretically challenging L2 contrast for Finnish speaking children.

The SLM (Flege, 1987, 1995) predicts that difficulties in L2 sound perception and production are greatest when an L2 sound is *similar* to an L1 sound and that L2 input affects L2 speech learning (Flege, 2009). A new revised version of the Speech Learning Model (SLM-r) expands the perspective on L2 category learning. The SLM-r proposes that, in addition to the perceived dissimilarity between L2 and L1 categories, the quality

and quantity of L2 input as well as the precision of L1 categories affect the formation of new L2 phonetic categories (Flege & Bohn, 2021). The SLM-r states that the quality and quantity of L2 phonetic input can differ greatly even when speakers have resided in an L2 speaking environment for the same amount of time, and therefore length of residence (LOR) is not a reliable measure for L2 input (Flege & Bohn, 2021). Another core proposition of the original SLM is that L2 perception precedes L2 production (Flege, 1995, 1999; Flege, MacKay, & Meador, 1999). In other words, accurate perception of L2 sounds does not require accurate production, but accurate production does require accurate perception. The SLM-r, however, states that rather than perception preceding production, L2 sound perception and production coevolve (Flege & Bohn, 2021).

The PAM (Best, 1994, 1995) and its updated version the Perceptual Assimilation model of L2 speech learning (PAM-L2; Best & Tyler, 2007) suggest that a situation where two L2 sounds assimilate equally to one L1 sound category (*single-category assimilation*) causes the most difficulties in L2 perception and production for language learners. Perception and production are also predicted to be potentially challenging in situations where two L2 sounds assimilate to one L1 category differently, so that one of them is perceived as an acceptable and the other as a deviant exemplar of the L1 category (*category-goodness difference*). The formation of a new category for the deviant L2 sound depends on the degree of perceived similarity between the categories (Best, 1994, 1995; Best & Tyler, 2007).

The predictions made by the SLM and the PAM-L2 are predominantly based on L2 learning in naturalistic L2 environments, but for the purposes of the present study they should also be discussed in the context of L2 learning in classroom environments.

According to Tyler (2019), the predictions of PAM-L2 apply differently in classroom settings, because instructed L2 learning often includes L2-accented spoken input and the use of written input in vocabulary and grammar teaching. These factors make *category-goodness assimilations* and *single-category assimilations* less likely to be learned in classroom environments, especially if the phonetic differences between categories are perceptually small or the learners rehearse words containing the L2 sounds from their orthographical forms (Tyler, 2019). The SLM-r, on the other hand, defines L2 phonetic input as sensory stimulation received in meaningful conversations in an L2. This poses important implications for L2 learning in instructed settings, where the amount of L2 input (in meaningful conversations) is limited and the quality of phonetic input is likely to be considerably different from the input received in a naturalistic L2 environment.

## 2. AOL and second language sound learning

*2.1. Phonetic studies on child and adult learners in naturalistic L2 environments and instructed learning settings*

The majority of earlier research on children's L2 speech learning has focused naturalistic learning environments. For example, studies by Oh et al. (2011) and Tsukada et al. (2005) examined the effects of AOL on acquisition of English vowels in Japanese and Korean adults and children who had immigrated to the United States or Canada. The results of Oh et al. (2011) showed that Japanese children (mean age 9.9 years) reached higher accuracy in English vowel production than Japanese adults in a year's time, but they did not reach the same production accuracy as the age-matched native English children. The results of Tsukada et al. (2005) showed that Korean children (aged 9–17 years) were able to discriminate English vowels better than Korean adults (aged 23–41 years). In addition, children with a longer LOR obtained better

discrimination scores than children with a shorter LOR. Furthermore, the child

participants' production of English vowels did not differ from native English children's

productions, whereas the Korean adults did not reach nativelike production accuracy

(Tsukada et al., 2005). In addition, a more recent study by Baigorri et al. (2019)

compared early and late Spanish-English bilinguals' perception of American English

(AmE) vowels and vowel contrasts. The participants were native Spanish speakers who

had learned English as an L2 after moving to the US. The early learners had moved to

the US before the age of 11 and the late learners had moved there after the age of 13.

All the participants were adults (18–48 years) at the time of testing. The results

indicated that early L2 learning is associated with better L2 sound perception than late

L2 learning.

Studies by Darcy and Krüger (2012), and Immonen and Peltola (2018) offer

insight into children's L2 sound learning in different immersion settings. The study by

Darcy and Krüger (2012) investigated the effects of early L2 acquisition on the

interaction of children's L1 and L2 by comparing early sequential bilingual and

monolingual children. The sequential bilingual participants were L1 Turkish speaking

children who started to learn L2 German in a dual-language daycare between the ages of

two and four. All the participants were 9–12 years old at the time of testing. The results

showed that the sequential bilinguals' discrimination of L2 vowel contrasts differed

from the native speakers' discrimination, but their productions of the same L2 vowel

contrasts were mostly target-like. The authors conclude that L2 perception is influenced

by L1 categories even when L2 exposure begins in early childhood. The study by

Immonen and Peltola (2018), on the other hand, examined how studying in an English

immersion education program in elementary school affects Finnish children's (aged 11–

13 years) production of British English vowels. The early learner group included 17 children from an English immersion education program and the control group had 15 children from a regular Finnish speaking class from the same school. The experiment consisted of a simple listen-and-repeat task where the children heard English words produced by a native British English speaker and repeated them on tape. The stimuli were 23 monosyllabic words containing 11 standard British English vowels in voiced and voiceless consonant contexts. Acoustic analysis of the groups' productions revealed that the groups produced the vowels differently in the voiced context words. The early learner group produced the English vowels closer to the native model (Immonen & Peltola, 2018). This result suggests that L2 immersion education in elementary school can result in better production accuracy of L2 sounds.

Studies on children's L2 sound learning in instructed settings are particularly relevant for the purposes of the present study. For example, a study by Kopečková, Dimroth and Gut (2019) investigated the perception and production of L2 phonemes by children and adults in the first hours of exposure to the L2. The participants were 10 German children (aged 9–11 years) and 19 German adults. Their perception and production of Polish sibilants was tested with discrimination and sentence imitation tasks during a two-week Polish course (14 hours of teaching in total). Both L2 sibilant perception and production were tested at two time points during the course (discrimination after 4.5 and 11.5. hours of teaching and sentence imitation after 9 and 13.5 hours of teaching). The results showed that the adult learners perceived L2 sibilants more accurately than the children. No differences between the groups were found in the production tests. The authors conclude that children are not necessarily

more successful learners of L2 sound perception and production than adults in instructed learning settings (Kopečková et al., 2019).

A study by Morales Reyes, Arechabaleta-Regulez and Montrul (2017) examined how 4–8-year-old monolingual and bilingual children learn foreign language phonology in a classroom. They tested 19 monolingual English and 6 bilingual Korean-English children who were learning Spanish as a foreign language in an instructed classroom setting. The children's production of Spanish rhotics was tested with a picture-naming task and their productions were compared to L1 Spanish speaking children's productions. The results showed that the children learned nativelike production of Spanish rhotics quickly, but the bilingual children performed better than the monolingual children (Morales Reyes et al., 2017).

Taken together, these findings indicate that children are often successful in learning L2 sound perception and production when they learn the language in a naturalistic L2 environment (Baigorri et al., 2019; Oh et al., 2011; Tsukada et al., 2005). In addition, studies on immersive classroom environments show that early L2 exposure in daycare or elementary school can result in more nativelike L2 production (Darcy & Krüger, 2012; Immonen & Peltola, 2018). On the other hand, some studies on L2 learning in classroom settings indicate that children might not be more successful than adults in instructed learning of L2 sound perception and production (Kopečková et al., 2019) and that bilingual children may have an advantage over monolingual children in instructed L2 production learning (Morales Reyes et al., 2017). Overall, much of the research on children's phonetic L2 learning has focused on naturalistic L2 environments and considerably more research is required in order to discover how early L2 exposure in classroom settings affects children's L2 perception and production learning.

*2.2. Child and adult learners in phonetic L2 training studies*

      Several studies have investigated how child and adult learners respond to different types of phonetic L2 training. For example, a perceptual training study by Giannakopoulou et al. (2013) examined whether high-variability phonetic training (HVPT) would improve the identification and discrimination accuracy of English phonemes by Greek adults (20 participants) and children (20 participants, aged 7–8 years). The results showed that both Greek adults and children improved their L2 identification and discrimination accuracy after the HVPT. However, the training effects were more pronounced for the child learners. Giannakopoulou et al. conclude that this finding suggests enhanced plasticity for spoken language in the child participants' developmental stage (2013). Another training study focusing on the perceptual development of L2 phonological contrasts (Heeren & Schouten, 2010) found contrasting results to those of Giannakopoulou et al. (2013). Heeren and Schouten (2010) investigated how perceptual sensitivity to Finnish consonant quantity develops in Dutch children (22 participants, aged 12 years). They compared the results to an earlier study that used the same experiment design on Dutch adults (Heeren & Schouten, 2008). Results showed that the children's identification scores improved slightly as a function of training; in other words their category boundary tended to move towards native Finnish listeners' category boundary (Heeren & Schouten, 2010). The same tendency was found in adults (Heeren & Schouten, 2008). No significant changes in the children's perceptual sensitivity were found in the discrimination tests (Heeren & Schouten, 2010). In addition, the adult subjects obtained higher overall discrimination scores than the children (Heeren & Schouten, 2010, 2008).

Although the present study focuses on L2 sound production instead of perception, the findings of Giannakopoulou et al. (2013) and Heeren and Schouten (2010) offer valuable insight into how child learners respond to perceptual training of L2 phonological contrasts. Overall, these earlier results from perceptual training studies indicate that HVPT training of L2 sounds benefits children more than adults (Giannakopoulou et al., 2013) but that the perceptual training of phonological quantity contrasts does not lead to significant improvements in discrimination accuracy (Heeren & Schouten, 2010).

Phonetic production training studies on L2 learners are particularly interesting from the point of view of the current study. The most important of such studies was conducted by Taimi et al. (2014) who examined the effects of a two-day listen-and-repeat training on 7–10-year-old children's production of the Swedish vowel contrast /y/ - /ʉ/. Thirteen monolingual Finnish children were tested. The experiment procedure and stimuli were the same as in the present study. The experiment paradigm consisted of four training sessions (with 30 repetitions of each stimulus) and four recording sessions (with 10 repetitions of each stimulus) on two consecutive days. The L2 vowel contrast /y/ - /ʉ/ embedded in two semi-synthetic pseudowords /ty:ti/ and /tʉ:ti/ were used as stimuli. The children's productions from the recording sessions were acoustically analyzed to evaluate the effects of the training. The results showed that the children changed their production of the novel vowel /ʉ/ on the second day of the experiment, after three training sessions (Taimi et al., 2014). The same stimulus words /ty:ti/ and /tʉ:ti/ and a similar training paradigm have also been used to study L1 Finnish (10 participants) and L1 English (9 participants) speaking adults (Peltola, K. U. et al., 2017). The study by Peltola, K. U. et al. (2017) included only one listen-and-repeat

training session (with 30 repetitions of each stimulus) with baseline and endpoint recordings. The results showed that the adult speakers did not learn to produce the L2 vowel contrast with one training session adults (Peltola, K. U. et al., 2017). A later study by Peltola K. U. et al. (2020) tested two groups of monolingual Finnish adults with an active listening (11 participants) and a listen-and-repeat (11 participants) training protocols using the same stimulus words /ty:ti/ and /tʉ:ti/. The first day of the two-day procedure consisted of baseline measurements (identification and production) and two training sessions (with 30 repetitions of each stimulus per session). The second day started with two training sessions and concluded with endpoint measurements (identification and production). The stimuli were presented in a pseudo-randomized order during training. The results showed that while the adults in the active listening group did not benefit from auditory training of the L2 contrast /y/ - /ʉ/,  the listen-and-repeat training protocol did improve the other group's identification and production of the novel L2 vowel /ʉ/ after four training session (i.e., between baseline and endpoint measurements). The findings of Taimi et al. (2014) and Peltola K. U. et al. (2017, 2020) suggest that listen-and-repeat training can alter L2 vowel production in both children and adults.

**3. Material and methods**

*3.1. Participants*

Eighteen monolingual Finnish children from a preschool in Southern Finland participated in the study. One participant did not complete the experiment and another participant's data had to be excluded from analysis due to technical difficulties during recordings. Therefore, a total of sixteen children (aged 6–7;4, mean age 6;8, median 6;10, age reported as years;months, 13 females) were tested. Before participating in the

experiment, all children and their parents gave written informed consent and completed a language background questionnaire. None of the children had any experience with Swedish or any other foreign languages. Two participants had a history of minor difficulties with the production of /r/ and /s/ in early childhood. Both had overcome all difficulties in articulation at the time of the experiment. These participants were not excluded from the data, as no words containing the sounds /r/ or /s/ were used in the experiment procedure.

*3.2. Stimuli*

The stimuli were two semi-synthetic pseudo words /ty:ti/ and /tʉ:ti/ with the close rounded vowels /y/ and /ʉ/ embedded in the first syllable. The Swedish vowel /ʉ/ is not phonological in Finnish and is situated on the border of Finnish vowel categories /y/ and /u/. This has been shown in studies investigating the perception of /ʉ/ by monolingual Finnish and bilingual Finnish-Swedish speakers (e.g. Peltola et al., 2010; Peltola, Tamminen, Toivonen, Kujala, & Näätänen, 2012). The Swedish vowel /ʉ/ is likely to be perceived as *similar* to the Finnish vowel /y/ by L1 Finnish speakers because the categories exist in the same phonetic space (Peltola et al., 2012) and it can therefore be hypothesized to cause difficulties in perception and production according to the SLM (Flege, 1995). In addition, according to the PAM, the Swedish vowel /ʉ/ represents a challenging *category-goodness difference assimilation* situation for Finnish speakers, since /ʉ/ and /y/ are both likely to assimilate to the Finnish /y/ category as an acceptable and slightly deviant exemplar of the L1 category (Best, 1994, 1995). Therefore, the Swedish contrast /y/ – /ʉ/ represents a theoretically difficult L2 sound contrast for L1 Finnish speakers. The aim of the present study is not to investigate the production of any particular language, but L2 sound production learning in general, and the Swedish

contrast /y/ – /ʉ/ was selected because it allows us to obtain exact and measurable data on the effects of listen-and-repeat training on L2 vowel production.

The stimuli were created with the Semisynthetic Speech Generation method (Alku, Tiitinen, & Näätänen, 1999). The natural productions of a 24-year-old Finnish-Swedish bilingual male speaker were used as the basis of the stimuli. The glottal pulse waveform (i.e. the air flow excitation signal of voiced speech generated by the vocal folds) was extracted from the natural speech signal produced by the speaker. The formant structures of the vowels were then synthesized over the natural glottal pulse waveform to create the pseudo word pair /ty:ti - tʉ:ti/. The L1 relevant (familiar) vowel /y/ embedded in the stimulus word /ty:ti/ had the first formant (F1) value of 269 Hz and the second formant (F2) value of 1866 Hz. The F1 and F2 values of the L1 irrelevant (novel) vowel /ʉ/ in the word /tʉ:ti/ were 338 Hz and 1258 Hz respectively. Because the vocal tract size of child speakers is considerably smaller than that of an adult male, the participants in this study were not expected to reach the exact formant values of the stimulus vowels. Instead, the focus was on the direction of the possible changes in the children's production.

*3.3. Procedure*

The experiment procedure was a short phonetic listen-and-repeat training paradigm consisting of four alternating recording and training sessions on two consecutive days. The experiment was conducted in a quiet room in a preschool during school hours. The data was collected using a portable laboratory consisting of a laptop computer with a Beyerdynamic MMX300 headset and an Asus Xonar U3 sound card. During recording and training sessions, the stimuli were presented automatically with Sanako Study Recorder software (version 8.22.0.0) in an alternating order, so that every

other word was /tʉ:ti/ with the novel vowel /ʉ/, and every other word was /ty:ti/ with the familiar vowel /y/. The stimulus word with the novel sound /ʉ/ was always presented first. The fixed stimulus order was selected to emphasize the acoustic characteristics of the non-native vowel contrast. The interstimulus interval (ISI) was 3 seconds. The same Sanako Study Recorder software was also used to record the participants' productions during recording sessions. Each stimulus was repeated thirty times during training sessions and ten times during recording sessions. The children were instructed to focus on what they heard and repeat the stimuli after the acoustic model during training and recording sessions. They were informed before each session whether the upcoming task was a training or a recording. They received no articulatory instructions or feedback on their productions during the experiment.

Before starting the experiment, the children completed a short familiarization task, where they heard both stimuli three times. The familiarization phase allowed the children to get accustomed to the pace and nature of the task and to adjust the volume to a comfortable level. The first day of the experiment then proceeded with a baseline recording followed by the first training, then a second recording and a second training. The second day continued in reverse order, starting with a third training, then a third recording, a fourth training and concluding with the final recording. The children were offered the chance to take a one minute break after each session before continuing on to the next session. The experiment lasted approximately 15 minutes per day including the short breaks between sessions. The participants showed no visible signs of fatigue during testing. In total, the participants heard and repeated both words 120 times during training sessions and 40 times during recordings. The experiment procedure is summarized in Table 1.

*3.4. Analysis*

The recorded speech signals were acoustically analyzed using Praat software version 6.0.43 (Boersma & Weenink, 2020). The maximum frequency was set at 6000 Hz. One F1 and F2 value per utterance was extracted from the steady state phase of the first syllable vowels using the Linear Predictive Coding (LPC) Burg algorithm. The participants produced each word ten times during the four recording sessions (40 repetitions per word). Individual average formant values for /y/ and /ʉ/ from the ten repetitions within each recording session were calculated from each speaker's productions. The F1 and F2 values of the vowels /y/ and /ʉ/ in the four recording sessions were then subjected to statistical analysis using IBM SPSS Statistics (version 25.0.0.1) software.

## 4. Results

A repeated measures Analysis of Variance (ANOVA) was performed for the average formant values with the 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 participants' productions changed significantly in any way across recording sessions. The main effects of FORMANT are not reported, since the F1 and F2 values are expected to differ automatically from each other. The analysis revealed the main effect of WORD ($F(1,15)=6.865$, $p=0.019$), indicating that the two words were produced differently. In addition, a WORD $\times$ SESSION interaction ($F(3,13)=5.725$, $p=0.010$) was found, which means that the two words changed differently across sessions. The analysis also revealed a WORD $\times$ FORMANT ($F(1,15)=7.297$, $p=0.016$) interaction, meaning that the F1 and F2 values were produced differently in the vowels /y/ and /ʉ/. Furthermore, the initial analysis revealed a WORD

× SESSION × FORMANT (F(3,13)=5.831, p=0.009) interaction, which indicates that the two words and the formant values of their first syllable vowels developed differently across sessions.

To understand the findings of the initial ANOVA, the sessions were then examined in pairs to see how the two words developed in the second, third and fourth sessions compared to the first session (baseline). A SESSION (2) × WORD (2) × FORMANT (2) repeated measures ANOVA was performed for each session pair. The analysis of the first and second sessions revealed the main effects of SESSION (F(1,15)=6.232, p=0.025) and a SESSION × FORMANT interaction (F(1,15)=7.850, p=0.013). These findings indicate that there is a significant change in the children's production after the first training session and that the F1 and F2 values changed differently between the sessions. The main effect of WORD (F(1,15)=5.188, p=0.038) was discovered when comparing the first and third sessions with the same ANOVA. In addition, WORD × SESSION (F(1,15)=5.519, p=0.033), WORD × FORMANT (F(1,15)=6.420, p=0.023) and WORD × SESSION × FORMANT (F(1,15)=6.718, p=0.020) interactions were found. Finally, the same ANOVA for the first and fourth sessions revealed the main effect of WORD (F(1,15)=11.115, p=0.005) as well as WORD × SESSION (F(1,15)=13.880, p=0.002), WORD × FORMANT (F(1,15)=12.140, p=0.003) and WORD × SESSION × FORMANT (F(1,15)=13.146, p=0.002) interactions. To summarize, the same main effects and interactions were found when comparing the first session to the third and fourth sessions. These findings suggest that the familiar vowel /y/ and the novel vowel /ʉ/ were produced as separate sounds and that the two vowels changed differently between the two session pairs, which can also be seen in Figure 1. The average formant values shown in Figure 1 suggest that the

significant difference found between the vowels across sessions is situated in the F2 and not in the F1. The F1 values in both vowels seem to remain unchanged across all sessions. This was to be expected, since the F2 value is the primary acoustic difference between /y/ and /ʉ/ and it was the only spectral difference between the stimuli.

Next, in order to investigate the interactions between WORD and SESSION further, the two words were analyzed separately within each session pair with a SESSION (2) × FORMANT (2) repeated measures ANOVA. The word /tʉ:ti/ was subjected to analysis first by comparing the first and second sessions, which revealed the main effect of SESSION ($F(1,15)=5.403$, $p=0.035$) and a SESSION × FORMANT interaction ($F(1,15)=6.231$, $p=0.025$). This indicates that there is a change in the children's production of /tʉ:ti/ between the first and second sessions. The same main effect of SESSION ($F(1,15)=6.914$, $p=0.019$) and a SESSION × FORMANT ($F(1,15)=6.966$, $p=0.019$) interaction were found when comparing the first and third sessions. In addition, a main effect of SESSION ($F(1,15)=7.806$, $p=0.014$) and a SESSION × FORMANT ($F(1,15)=7.573$, $p=0.015$) interaction were also discovered between the first and fourth session. To summarize, the same main effect of SESSION, as well as the same SESSION × FORMANT interactions were found in all three session pairs. This indicates that there is a change in the F1 or F2 values of the children's production of /tʉ:ti/ already after the first training session and that the change remains throughout the experiment (see Figures 1 and 2). The same ANOVA revealed no significant changes in the word /ty:ti/. The word /ty:ti/ was not subjected to further analysis, since no significant findings concerning session were found.

The main effects of SESSION and the SESSION × FORMANT interactions found for the word /tʉ:ti/ in all session pairs suggest that the F1 or F2 values differed between

the sessions. Paired samples t-tests for the F1 and F2 values in the novel vowel /ʉ/ in all three session pairs were performed to see how the formants developed in the second, third and fourth sessions compared to the baseline. A significant difference in the F2 values of the novel vowel /ʉ/ was found between the first and second sessions (t(15)=2.419, p=0.029), first and third sessions (t(15)=2.675, p=0.017) as well as the first and fourth sessions (t(15)=2.816, p=0.013). Figure 3 shows the development of all speakers' F2 values in both vowels across sessions. No significant changes were found in the F1 values of the novel vowel /ʉ/ in the word /tʉ:ti/.

## 5. Discussion

Earlier research on children's L2 perception and production learning has mainly focused on naturalistic learning settings and perceptual training paradigms. Our results, based on a study with young instructed learners, seem in line with those earlier studies, which have shown that early AOL often benefits L2 sound production (Darcy & Krüger, 2012; Oh et al., 2011; Tsukada et al., 2005). The results show that 6–7-year-old children benefit from phonetic listen-and-repeat training of an L2 sound contrast and change their production of a novel L2 vowel rapidly as a function of training. While previous studies have mainly focused on naturalistic L2 environments, our results suggest that 6–7-year-olds benefit even from very limited and controlled L2 sound exposure in an instructed training setting. The statistical analysis revealed that the training effects were immediate, as the children changed their production of the novel vowel /ʉ/ significantly in the second recording session, after the first training. The change in production was reflected in the lowering of the F2 value of /ʉ/, which can be seen in Figures 2 and 3 showing the F2 values of /ʉ/ produced by the children across all four recording sessions.

The primary spectral difference between the vowels /y/ and /ʉ/ is the F2, which is lower in the central vowel /ʉ/ than in the front /y/. The fact that there were no significant changes in the production of the familiar vowel /y/ shows that the children were able to distinguish the two sounds from each other and produce them as separate categories, even though the vowel contrast is not phonologically relevant in their L1. The finding that the children started to produce /ʉ/ with lower F2 values after the first training session suggests that they either perceived the spectral difference between the vowels in the stimuli and then applied it to their own speech, or that their perception and production co-evolved during the training, as proposed by the SLM-r (Flege & Bohn, 2021). The latter explanation seems plausible when considering the nature of the listen-and-repeat training paradigm, where the children are exposed to their own speech in addition to the acoustic stimuli, meaning that they receive repetitive auditory feedback from their own productions during training. However, any hypotheses on perception cannot be verified without additional perceptual measurements.

The difference between the F2 values in /y/ and /ʉ/ grew to 124–199 Hz after the first recording, where the difference was merely 22 Hz (Figure 2). This indicates that after the baseline recording, the two sounds were produced as two spectrally and perceptually distinct vowels, since the frequency difference between the vowels' F2 values in sessions 2–4 clearly exceeds all just noticeable difference (JND) thresholds observed in previous studies for vowel quality in this frequency range (e.g., Kewley-Port & Watson, 1994). In addition, previous studies have shown that monolingual Finnish and bilingual Finnish-Swedish speakers are able to categorize vowel stimuli from the Swedish /y/-/ʉ/-/u/ continuum with relatively small differences in the F2 values (Tamminen et al., 2013). This supports our statistical findings that the two words

were produced differently across sessions and that the F2 in the novel vowel /ʉ/ lowered

significantly in the second, third and fourth sessions compared to the baseline (Figure

3). The fact that the children did not reach the exact F2 values of /ʉ/ in the acoustic

stimulus /tʉ:ti/ (F1 = 338 Hz, F2 = 1258 Hz) was expected, because the stimuli were

based on the voice of an adult male speaker with a considerably larger vocal tract than

children.

   When compared to earlier results by Taimi et al. (2014) and Peltola K. U. et al.

(2017, 2020), the results of the current study show that 6–7-year-old preschoolers

change their pronunciation of an L2 vowel contrast faster through listen-and-repeat

training than 7–10-year-old children or adults. Taimi et al. (2014) and Peltola K. U. et

al. (2017, 2020) used different amounts of listen-and-repeat training with the same

stimulus words /ty:ti/ and /tʉ:ti/ that were used in the current experiment. Their results

showed that the thirteen 7–10-year-old children changed their production of the novel

vowel /ʉ/ on the second day of the experiment, after three training sessions (Taimi et al.,

2014). The ten monolingual Finnish adults tested by Peltola K. U. et al. did not change

their production of /ʉ/ after one listen-and-repeat training session in a one-day training

procedure (2017). A later study showed slight improvements in Finnish adults'

production of the novel vowel /ʉ/ after four sessions of listen-and-repeat training on two

consecutive days (Peltola, K. U. et al., 2020). Contrary to the present study, the training

protocol in Peltola K. U. et al. (2020) used a pseudo-randomized stimulus order, but the

amount of repetitions per training session (30 repetitions of each word) was identical in

all these studies. The fact that the children in the current study changed their

pronunciation already after one training on the first day indicates that they were able to

change their articulatory patterns to fit the acoustic model more rapidly than the older

age groups studied by Taimi et al. (2014) and Peltola K. U. et al. (2017, 2020) with the same listen-and-repeat training method.

It could be that the children investigated in the current experiment benefitted from an enhanced ability to imitate the acoustic model, resulting in faster changes in production than in their older counterparts (Taimi et al., 2014; Peltola K. U. et al., 2017, 2020). This could partly explain the differences between the 6–7-year-olds and adults, but seems an unlikely explanation for the results from 6–7 and 7–10-year-old children, considering the relatively small age difference between the groups. In addition, as discussed above, the listen-and-repeat protocol includes repetitive auditory feedback from the speaker's own productions in addition to the trained stimuli. Therefore, a more probable explanation could be that, in line with the SLM-r (Flege & Bohn, 2021), the perception and production processes co-evolved faster during training in the 6–7-year-old children than in 7–10-year-olds or adults. However, since neither the current study nor the study by Taimi et al. (2014) included perceptual measurements, no definite conclusions on the vowel perception accuracy of the children can be drawn without further studies.

Another perspective that needs to be considered is the participants' previous experience with foreign languages. The children tested in the present study had no experience with any other language than their L1 (Finnish). Some of the 7–10-year-olds tested by Taimi et al. (2014) had minimal exposure to some foreign languages, but none of them had started to study any L2 in school. The Finnish adults tested by Peltola K. U. et al. (2017, 2020), on the other hand, had studied English and Swedish and they had some experience with other foreign languages. Therefore, the participants tested in these studies had different levels of experience and knowledge on foreign languages prior to

the training, which could affect their perception and production of the trained vowel contrast. However, our hypothesis is that the adult participants (Peltola K. U. et al., 2017, 2020) would more likely benefit from their previous experience with English and Swedish, rather than be hindered by it, in a training paradigm where the trained vowel contrast is phonemic in Swedish. Therefore, the differences in groups' L2 experience does not seem to conclusively explain why the 6–7-year-olds tested in the current study responded to the listen-and-repeat training faster than the older children (Taimi et al., 2014) or adults (Peltola K. U. et al., 2017, 2020).

Finally, we suggest that the result that the 6–7-year-old children changed their production of the novel L2 vowel already after the first training may indicate that their L1 phonological categories are not yet fully established at this developmental stage (Flege & Bohn, 2021). The L1 sound categories are likely to be more established for the older children and even better established for the adults. The developmental stage of L1 category formation could be a possible explanation for our findings, since the neural commitment to L1 sound categories has been found to reduce the ability to perceive L1 irrelevant sound contrasts (e.g., Kuhl et al., 2008) and L1 category precision could affect the formation of L2 categories (Flege & Bohn, 2021). Our findings could indicate that the younger children's L1 categories were relatively precise (i.e., had less F1-F2 variability within L1 categories), and even more precise than the L1 categories of the older children or adults. We suggest that the still evolving and/or precise L1 phonological system allowed the 6–7-year-olds to change their production of the novel vowel /ʉ/ rapidly towards the acoustic model and to produce the L1 irrelevant /y/ – /ʉ/ contrast.

To summarize, the results of the current study show that children are successful and effective L2 production learners in an instructed listen-and-repeat training setting. Most importantly, comparing the results to previous findings from older age groups suggests that 6–7-year-old children learn to produce a novel L2 vowel faster through listen-and-repeat training than 7–10-year-old children or adults, which could indicate less established and/or more precise L1 phonetic categories in 6–7-year-old learners. However, the effects of AOL and L1 category precision on children's L2 perception and production learning and development in instructed settings need to be investigated further. The current study could be extended by adding discrimination and identification measurements to gain more information on how listen-and-repeat training affects children's L2 sound perception in addition to L2 production.

## 6. Conclusions

The results show that 6–7-year-old children change their production of a novel L2 vowel rapidly towards the acoustic model /tʉ:ti/ as a function of listen-and-repeat training. The changes in production were immediate and persisted throughout testing. The results suggest 6–7-year-old children learn to produce a non-native vowel contrast even with limited L2 sound exposure in a training setting that is traditionally used in foreign language classrooms.

**References**

Alku, P., Tiitinen, H., & Näätänen, R. (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.

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech; Lang Speech, 62*(4), 002383091880693-700. doi:10.1177/0023830918806933

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* (pp. 167-224). Cambridge, MA: MIT Press.

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* (pp. 171-206) York Press, Baltimore.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Best, C. T., & Tyler, M. (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* (pp. 13-34). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Boersma, P. & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program].Version 6.0.43.

Darcy, I., & Krüger, F. (2012). *Vowel perception and production in turkish children acquiring L2 german* doi:https://doi.org/10.1016/j.wocn.2012.05.001

Flege, J. E. (2009). *Give input a chance!* In T. Piske & M. Young-Scholten (Eds.), Input matters in SLA (pp.175–190). Clevedon: Multilingual Matters.

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*, 47-65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273) Timonium, MD; York Press.

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.

Flege, J. E., & Bohn, O. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge, UK: Cambridge University Press.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native italian speakers' perception and production of english vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973-2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5 Pt 1), 3125-3134. doi:10.1121/1.413041

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Giannakopoulou, A., Uther, M., & Ylinen, S. (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.

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the finnish /t-t:/ Distinction in dutch 12-year-old children: A training study. *Journal of Phonetics, 38*, 594-603.

Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *The Journal of the Acoustical Society of America, 124*(4), 2291-2302. doi:10.1121/1.2967472

Immonen, K., & Peltola, M. S. (2018). Finnish children producing english vowels - studying in an english immersion class affects vowel production. *Journal of Language Teaching and Research, 9*(1), 27-33. doi:10.17507/jltr.0901.04

Kewley-Port, D., & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *The Journal of the Acoustical Society of America, 95*(1), 485-496.

Kopečková, R., Dimroth, C., & Gut, U. (2019). Children's and adults' initial phonological acquisition of a foreign language. *Journal of Second Language Pronunciation, 5*(3), 374-401.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93-107.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1493), 979-1000.

Kuhl, P. K., Williams, K. A., & et al. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608.

Morales Reyes, A., Arechabaleta-Regulez, B., & Montrul, S. (2017). The acquisition of rhotics by child L2 and L3 learners. *Journal of Second Language Pronunciation, 3*(2), 242-266.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (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.

Official Statistics of Finland (OSF): Population structure [e-publication]. ISSN=1797-5395. 2018, Appendix table 1. Population according to language 1980–2018. Helsinki: Statistics Finland [referred: 22.4.2021]. Access method: http://www.stat.fi/til/vaerak/2018/vaerak_2018_2019-03-29_tau_001_en.html

Peltola, K. U., Rautaoja, T., Alku, P., & Peltola, M. S. (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.

Peltola, K. U., Tamminen, H., Alku, P., Kujala, T., & Peltola, M. S. (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.

Peltola, M. S., Tamminen, H., Salonen, L., Toivonen, H., Kujala, T., & Näätänen, R. (2010). Two languages–one brain. Paper presented at the *Third ISCA Workshop on Experimental Linguistics,* 145-148.

Peltola, M. S., Tamminen, H., Toivonen, H., Kujala, T., & Näätänen, R. (2012). Different kinds of bilinguals – different kinds of brains: The neural organisation of two languages in one brain. *Brain and Language, 121*(3), 261-266. doi://dx.doi.org/10.1016/j.bandl.2012.03.007

Pyykkö, R. (2017). Monikielisyys vahvuudeksi. Selvitys Suomen kielivarannon tilasta ja tasosta. *Publications of the Ministry of Education and Culture 2017:51.* Helsinki.

Taimi, L., Jähi, K., Alku, P., & Peltola, M. S. (2014). Children learning a non-native Vowel–The effect of a two-day production training. *Journal of Language Teaching and Research, 5*(6), 1229-1235.

Tamminen, H., Peltola, M. S., Toivonen, H., Kujala, T., & Näätänen, R. (2013). Phonological processing differences in bilinguals and monolinguals. *International Journal of Psychophysiology, 87(1),* 8-13.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of english vowel production and perception by native korean adults and children. *Journal of Phonetics, 33*(3), 263-290.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. Bothe Jespersen & M. Hjortshøj Sørensen (Eds.), *A sound approach to language Matters—In honor of Ocke-Schwen Bohn* (pp. 607-630). Aarhus: Dept, of English, School of Communication & Culture, Aarhus University.

| Day 1 | Day 2 |
|---|---|
| 1st Recording session | 3rd Training session |
| 10 x /tʉ:ti/ | 30 x /tʉ:ti/ |
| 10 x /ty:ti/ | 30 x /ty:ti/ |
| →Recorded | |
| *Optional 1 minute break* | *Optional 1 minute break* |
| 1st Training session | 3rd Recording session |
| 30 x /tʉ:ti/ | 10 x /tʉ:ti/ |
| 30 x /ty:ti/ | 10 x /ty:ti/ |
| →Not recorded | |
| *Optional 1 minute break* | *Optional 1 minute break* |
| 2nd Recording session | 4th Training session |
| 10 x /tʉ:ti/ | 30 x /tʉ:ti/ |
| 10 x /ty:ti/ | 30 x /ty:ti/ |
| *Optional 1 minute break* | *Optional 1 minute break* |
| 2nd Training session | 4th Recording session |
| 30 x /tʉ:ti/ | 10 x /tʉ:ti/ |
| 30 x /ty:ti/ | 10 x /ty:ti/ |

**Table 1.** The experiment procedure. The stimuli were presented in alternating order, i.e. every other word was /tʉ:ti/ and every other word was /ty:ti/. The order of the stimuli remained the same throughout the experiment. The participants listened to and repeated the words during training and recording sessions.
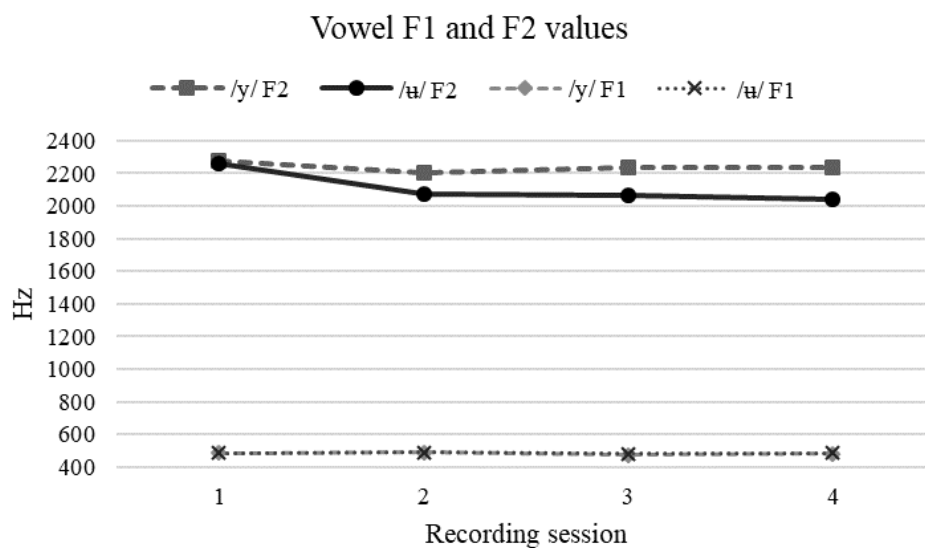
**Figure 1.** The average F1 and F2 values of the first syllable vowels produced by the participants across recording sessions.
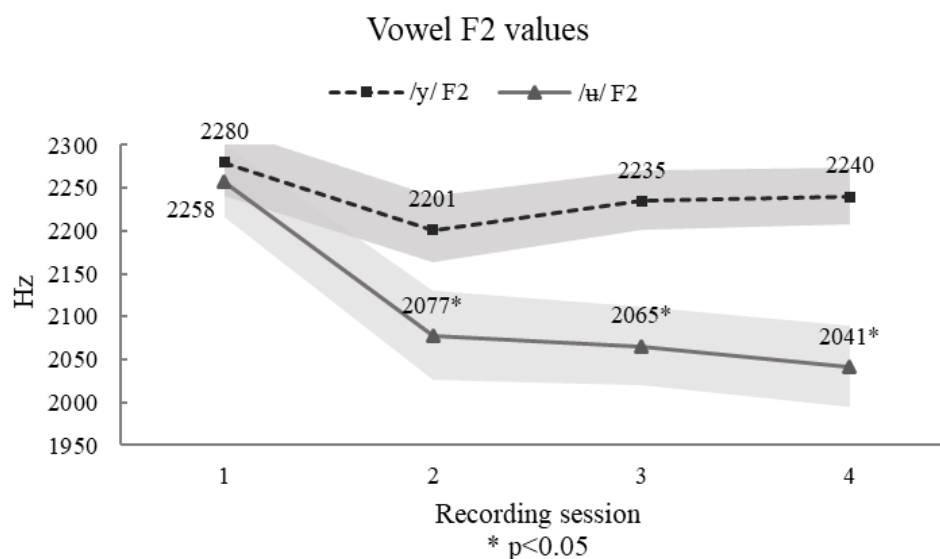


**Figure 2.** The average F2 values (with 95% confidence intervals) in /y/ and /ʉ/ produced by the children across recording sessions. The SESSION (2) × FORMANT (2) repeated measures ANOVA revealed the main effect of SESSION for the vowel /ʉ/ in all three session pairs when sessions 2, 3 and 4 were compared to the baseline. All significant between-session changes observed in the ANOVA are marked with an asterisk (* $p < 0.05$). No significant findings emerged for /y/. The difference between the vowels' F2 values was 22 Hz in the baseline and 199 Hz in the fourth recording.
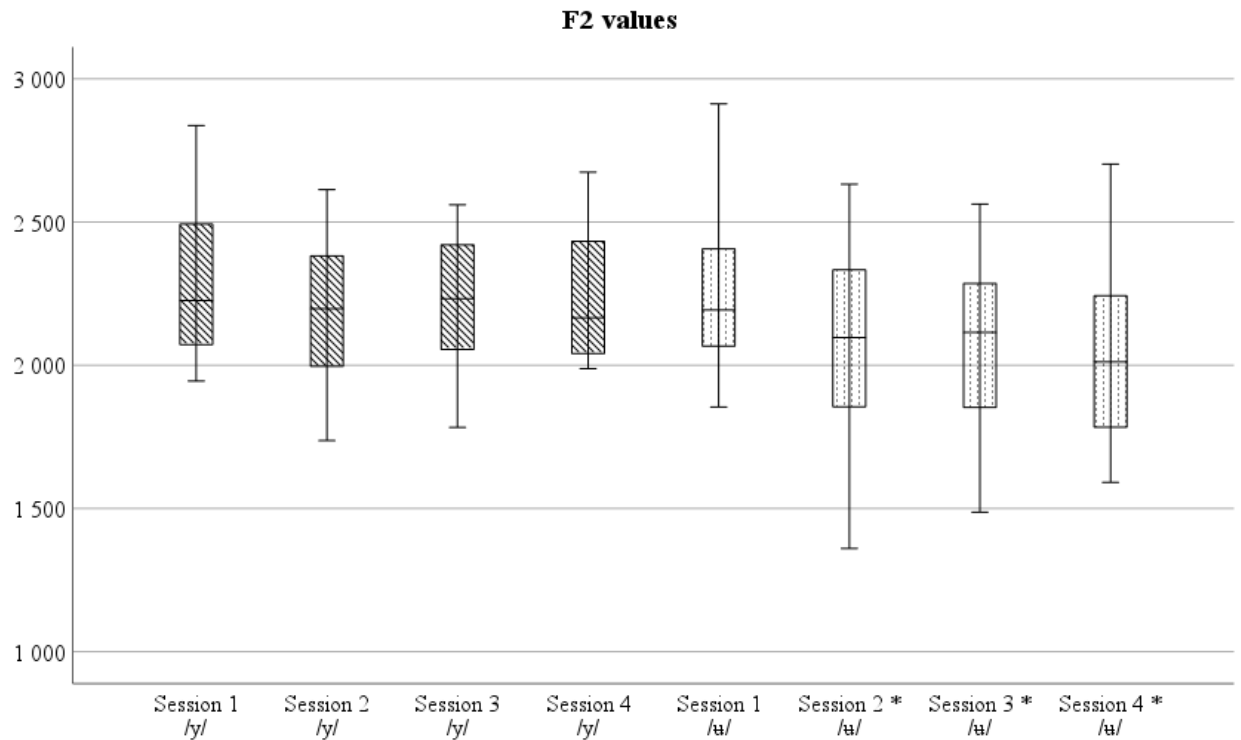
**Figure 3.** The F2 values in /y/ and /ʉ/ from all the speakers across recording sessions. No significant changes emerged for /y/ in a SESSION (2) × FORMANT (2) repeated measures ANOVA when sessions 2, 3 and 4 were compared to the baseline and no further t-tests were performed. After a SESSION (2) × FORMANT (2) repeated measures ANOVA, the /ʉ/ F2 values produced in sessions 2, 3 and 4 were compared to the baseline with paired samples t-tests. Significant between-session changes are marked with asterisks (* $p < 0.05$).