

# Orthography does not hinder non-native production learning in children

Second Language Research

1–13

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/02676583221076645

[journals.sagepub.com/home/slr](https://journals.sagepub.com/home/slr)**Katja Immonen** **Kimmo U Peltola** **Henna Tamminen** 

University of Turku, Finland

**Paavo Alku** 

Aalto University, Finland

**Maija S Peltola** 

University of Turku, Finland

## Abstract

Children are known to be fast learners due to their neural plasticity. Learning a non-native language (L2) requires the mastering of new production patterns. In classroom settings, learners are not only exposed to the acoustic input, but also to the unfamiliar grapheme–phoneme correspondences of the L2 orthography. We tested how 9–10-year-old children, with Finnish as a native language (L1), respond to a two-day listen-and-repeat training paradigm, where they simultaneously hear acoustic stimuli and see orthographic cues. In the procedure, non-words containing the L2 vowel /ʌ/ were presented simultaneously with an orthographic cue showing <u>, guiding pronunciation towards the L1 vowel /u/ according to Finnish grapheme–phoneme correspondences. Earlier studies showed that Finnish adults rely on the orthographic cue over the acoustic one, leading them to produce /u/ instead of /ʌ/ when presented with the incongruent L1–L2 grapheme–phoneme correspondence (<u> – L1: /u/, L2: /ʌ/). Also, an earlier result from age-matched children receiving only acoustic input showed relatively fast pronunciation changes towards the target vowel. Our present results indicate clear and fast production learning of the non-native sound, and the misleading orthographic cue did not draw attention away from the target acoustic form. With orthographic cues, the participants learned to produce novel sounds faster than without them.

## Corresponding author:

Katja Immonen, Phonetics and Learning, Age and Bilingualism laboratory (LAB-lab), University of Turku, Koskenniemenkatu 4, 20014, Turku, Finland.

Email: [katja.m.immonen@utu.fi](mailto:katja.m.immonen@utu.fi)

## Keywords

audio-visual cues, children, orthography, production learning, training

## I Introduction

Models of non-native speech sound learning have proposed that the role of the native language is of crucial importance in how the non-native categories are perceived and thus produced. Several studies have proposed that perception precedes production in second language (L2) speech learning, meaning that accurate production of L2 sounds requires accurate L2 perception (Flege, 1995, 1999; Flege et al., 1999b), whereas more recent research suggests that L2 perception and production co-evolve (Flege and Bohn, 2021). In the framework of the Speech Learning Model (SLM; Flege, 1987) and its revised version – the SLM-r (Flege and Bohn, 2021) – the most difficult items are termed *similar*. This refers to a situation where the non-native sound bears a close resemblance to a native sound category, but is still distinct from it. In addition, the SLM-r (Flege and Bohn, 2021) predicts that the formation of a new phonetic category for a non-native sound is affected by the quantity and quality of phonetic input as well as the precision of existing first language (L1) categories. In contrast, the L2 Linguistic Perception (L2LP) model states that the most problematic items are the *new* items (Escudero, 2005). However, the term *new* in the L2LP model actually refers to the same situation as the term *similar* in the SLM: the L2 sound contains both non-native and native characteristics, thus making the target sound resemble a native sound. In terms of the Perceptual Assimilation Model (PAM-L2; Best and Tyler, 2007), the problems in non-native perception and production are most severe when two L2 categories are assimilated equally to one L1 category (*single-category assimilation*). In all of these models, the common factor is that a close resemblance of an L2 sound with an L1 sound category constitutes the most fundamentally difficult learning setting in L2 perception and production.

Earlier research has shown that L2 speech acquisition may depend on factors such as learning environment or age. Some of these studies have focused on the learning of perception (e.g. Flege and MacKay, 2004), whereas some others have focused on production (e.g. Immonen and Peltola MS, 2018). Immigration (Winkler et al., 1999) and language immersion education (Immonen and Peltola MS, 2018; Peltola MS et al., 2005) seem to be L2 learning environments that result in the formation of native-like perception and production patterns for the target language sounds. Non-native vowel contrasts elicit identical memory trace activation in immigrants and native speakers when pre-attentive perception is measured with mismatch negativity (MMN; Winkler et al., 1999). In contrast, classroom learning might not result in similar learning effects, since even proficient students show no neural memory traces for target language phoneme contrasts (Peltola MS et al., 2003). The imitation of a naturalistic learning setting seems to be conducive to learning L2 sounds, since it has been found that in early immersion settings children are able to discriminate non-native vowels through the formation of new memory traces (Cheour et al., 2002; Peltola MS et al., 2005). In contrast, another study by Peltola MS and Aaltonen (2005) showed results suggesting that the language context during testing may be the key for the activation of neural representations of non-native vowels in young adults. However, a study by Immonen and Peltola MS (2018) showed

that school-aged children in a language immersion class learn to produce non-native vowels more accurately than their peers in a non-immersion class.

Several studies have indicated that learning proceeds faster at an early age and that child learners may achieve more accurate performance than adults (pronunciation, sound discrimination and identification) in L2 phonetic learning (Flege et al., 1999a; Giannakopoulou et al., 2013; Taimi et al., 2014). Age effects may be seen in a study by Giannakopoulou et al. (2013) showing that perceptual training effects are more pronounced in children than adults. The study showed that child learners, when compared to adults, seem to benefit more from high-variability training in L2 sound identification and discrimination. In a more recent study, however, Giannakopoulou et al. (2017) showed contrary results, i.e. a detriment of HVPT (high-variability phonetic training). Older learners (aged 62–73 years) also showed rapid L2 production learning (Jähi et al., 2015). The participants completed a simple two-day listen-and-repeat training protocol and gained significant changes in L2 vowel articulation. In addition, factors such as continued use of the L1 (Flege et al., 1997; Flege and MacKay, 2004) have been shown to result in less accurate L2 perception and production patterns.

Training type also seems to affect learning. For instance, Iverson et al. (2012) used the high-variability phonetic vowel training method and showed that both experienced and inexperienced L2 learners benefit from this type of training. This was shown in vowel identification and discrimination, and to a lesser degree in vowel production tests. Along similar lines, Tamminen and Peltola MS (2015) indicated that even highly proficient students of an L2 can enhance their voice onset time memory traces with a simple listen-and-repeat training paradigm. The same type of training was found to be effective in vowel production with adult learners, and surprisingly, even mere passive auditory training resulted in production changes (Peltola KU et al., 2017). Altogether, it seems that training results may depend either on the age of the participants, and/or the type of training provided.

L2 training studies have found that the role of orthography may be significant in phonetic learning. For example, Escudero et al. (2008) investigated the phonetic and lexical mappings of auditorily confusable L2 non-words. The participants learned the non-words by matching the auditory stimuli to either the orthographic form or a picture. The results indicated that the orthographic forms may help learners to form separate representations of L2 non-words that contain L2 sounds that are similar to each other. When studying literate learners, the role of orthography in non-native language learning has been found to be important. For example, Hayes-Harb et al. (2010) have shown that the phonological analysis of new words can be affected by orthography in adult learners. They tested three groups of learners who heard the same auditory stimuli paired either with no orthographic input or with written forms. The written forms were either congruent (<kamad>, /kɑməd/) or incongruent (<kamand>, /kɑməd/) with their L1 spelling conventions. The group who saw the incongruent written forms of the auditory stimuli showed interference from the orthographic forms in an auditory word-picture matching test. Furthermore, Rastle et al. (2011) showed that language processing involves rapid and automatic interaction between orthographic and phonological representations (when tested with picture naming, shadowing, forced-choice spelling, picture spelling and auditory lexical decision). They introduced L2 learners to spelling-sound consistent and spelling-sound

inconsistent orthographic forms of novel spoken words. Underlining the importance of orthography in L2 learning, the results showed significant orthographic effects on speech perception (not on shadowing but on auditory lexical decision) and production (immediate and persistent effects on picture naming after the introduction of word spellings). In addition, when studying the learning of length production in experienced Italian L2 English learners, Bassetti (2017) found that the L1 grapheme–phoneme correspondences influenced L2 production. This effect was present even when the participants did not see the orthographic form. The results suggested that the interaction between L2 orthographic forms and L1 grapheme–phoneme correspondences can lead learners to produce L1 contrasts in L2 speech and that the role of orthographic conventions may extend far.

On the other hand, some training studies have found orthography to have no effect on L2 sound acquisition (Simon et al., 2010). In addition, the role of orthography in L2 sound learning may depend upon the manner in which the grapheme–phoneme correspondence functions in a particular language (Escudero and Wanrooij, 2010). A study by Peltola KU et al. (2015) demonstrated how L1 orthography affects L2 production learning when the spelling system is transparent and phonological. In that study, literate Finnish adult learners were presented with auditory cues (/ty:ti/, /tʌ:ti/) containing a non-native sound and visual cues (L1 orthography: <tyyti>, <tuuti>) simultaneously. Since the Finnish spelling system is highly phonemic, the visual and auditory cues provided incongruent information about the pronunciation. As a result, the participants learned to produce what they read, and not what they heard, suggesting that the visual written cue was more powerful. In fact, the results showed that the learners started to produce /u/ according to the visual cue <u> and ignored the auditory form /ʌ/. Taken together, it seems evident that the ability to read has an effect on how we perceive and thus also produce speech.

Based on earlier research, it is clear that learning to perceive and produce speech in a non-native language is a complex task. Learning may depend on several factors, such as age, training type and learning environment. The aim of this study is to see whether literate children, similarly to adults (Peltola KU et al., 2015), rely on the orthographic form over the acoustic one when making decisions about the phonological forms of words. Another aim is to see whether the audio-visual training results in the current study are different from those obtained with mere auditory training in an earlier study, where age-matched children learned to produce the exact same non-native vowel contrast after three training sessions (Taimi et al., 2014). The hypothesis is that after L1 grapheme–phoneme correspondences have been formed, they affect the learning and production of L2 sounds in literate children in a similar manner as in adults. Thus, children receiving audio-visual training, where the target L2 sound is paired with a grapheme that is incongruent with the L1 grapheme–phoneme correspondences and corresponds to a novel L2 phoneme, will not benefit from the training as much as age-matched children learning L2 sounds through mere auditory training.

## II Methods

### *I Participants*

Twelve 9–10-year-old monolingual Finnish children participated in the study (mean age 9.5 years, range 9–10 years, 6 females). The children were in the third grade in a Finnish

elementary school. They had started to study only English basics during the academic year and they only had a minor exposure to spoken English, none to written English and more importantly, none of them had any previous experience with Swedish and none had studied any other languages. All had resided in Finland for their whole childhood. All children had normal hearing and none had been diagnosed with language deficits of any kind. According to the teacher's report, none of the participants deviated from the normal reading skills in relation to the age group. All participants reported to be in normal health on the day of testing and none showed any signs of fatigue. The group thus represented typical Finnish monolingual children who have learned to read their native language. All participants gave informed consent prior to participating, and a written approval was also obtained from their parents. The test procedures were approved by the Ethics Committee of the University of Turku, Finland.

## 2 Stimuli

The auditory stimuli were semi-synthetic pseudo-words /ty:ti/ and /t̥ɛ:ti/ created using the Semisynthetic Speech Generation method, SSG (Alku et al., 1999). The stimuli were based on the natural production of a 24-year-old Finnish–Swedish bilingual male speaker. The glottal pulse was extracted from a natural speech signal produced by the male speaker and the formant patterns relevant for the distinction of the vowels in the first syllable were then synthesized over the natural glottis pulse. As a result, a word pair /ty:ti–t̥ɛ:ti/, containing the Swedish vowel contrast /y–ɛ/, was established. The non-target word contained the vowel /y/, which is phonemic both in Finnish and Swedish. The target word contained the Swedish (non-native) vowel /ɛ/. This vowel contrast is extremely difficult for Finns to perceive and to produce according to second language learning models. The target vowel /ɛ/ may be labelled as *similar* by the SLM and the SLM-r (Flege, 1987; Flege and Bohn, 2021) or *new* by the L2LP model (Escudero, 2005), and according to PAM-L2 (Best and Tyler, 2007) the Swedish vowels /y/ and /ɛ/ can be said to assimilate unevenly to Finnish /y/ with /y/ certainly being a better exemplar of the Finnish category /y/. The stimuli are referred to as target and non-target since the non-native vowel /ɛ/ is the target which the participants should learn to perceive and produce in the training setting of this study, whereas the vowel /y/ is already familiar and is therefore not a target for learning. The vowel /ɛ/ in the target stimulus had an F1 value of 338 Hz and F2 of 1,258 Hz, and the vowel /y/ in the non-target stimulus had an F1 and F2 of 269 Hz and 1,866 Hz, respectively. However, since the acoustic features in the voice of an adult male speaker are considerably different from those of a young child, the participants were not expected to reach the exact formant values of the stimuli.

The participants were also presented with the visual cues showing orthographic forms of the stimulus words during training. These visual cues derive from the Finnish writing system with the non-target vowel /y/ cued with <tyyti> and the target vowel /ɛ/ with <tuuti>. The visual cue followed Finnish grapheme–phoneme correspondences, which are transparent and according to which the grapheme <y> corresponds to the vowel /y/ and the grapheme <u> corresponds in all situations to the rounded back vowel /u/. Therefore, the target visual cue was misleading, since the symbol <u> corresponds to /u/ instead of /ɛ/ according to Finnish grapheme–phoneme correspondences.

In conclusion, the target vowel /ʉ/ is perceptually difficult to categorize according to second language learning models such as the SLM, the PAM-L2 and L2LP. More importantly for this study, the acoustic stimulus /tʉ:ti/ and the visual cue <tuuti> for the target word are incongruent as they may guide the participants' production towards two different speech sounds: the acoustic cue towards the non-native vowel /ʉ/ and the visual cue towards the native vowel /u/.

### 3 Procedure

The experiment was conducted on two consecutive days in a quiet room in an elementary school in Southern Finland using a portable laboratory consisting of an HP laptop with a Beyerdynamic MMX300 headset and an Asus Xonar U3 sound card. During recording sessions, Sanako Study Recorder software (version 8.22.0.0) was used to present the auditory stimuli and record the participants' productions. During training sessions, the auditory stimuli and the orthographic cues were presented as a PowerPoint slideshow where the acoustic and visual cues were played simultaneously. The interstimulus interval (ISI) was 3 seconds both in training and recording sessions and the stimuli were presented automatically in an alternating order so that every other word was the target /tʉ:ti/. The alternating pattern was used in both the training and recording to emphasize the acoustic characteristics of the non-native contrast /y-ʉ/.

The first day of the experiment started with a short familiarization block, where the participants heard both stimuli three times but did not see the orthographic presentations. The purpose of the familiarization phase was to give the children the opportunity to adjust the volume to a comfortable level and get accustomed to the pace of the experiment. The participants were then instructed to listen and repeat after the model in the following recording and training sessions. After familiarization and instruction, the first day continued with the first recording (baseline) followed by the first training, then a second recording and a second training.

The second day began with a third training session followed by a third recording, a fourth training and then concluded with the fourth and final recording. For a more detailed description of the procedure see Figure 1. Altogether, the children participated in four training (4 minutes each) and four recording (2 minutes each) sessions and the experiment lasted about 15 minutes per day. The recording sessions included 10 repetitions of each stimulus. In other words, 40 repetitions of each word were recorded from each participant during the experiment. The training sessions included 30 repetitions of each stimulus presented in auditory and orthographic forms simultaneously. In short, the participants repeated each word 120 times in the four training sessions and they were not given any feedback during or after training or recording. All in all, the participants repeated each of the two words 160 times during the entire experiment. We did not want the visual cues to have an effect on the actual pronunciation testing. Therefore, the participants were not presented with any orthographic cues during testing. That is to say, we wanted to investigate how the orthographic information presented during the training phase affects the development of the mental representation of the non-native sound and its production.

	Day 1		Day 2
<i>Auditory</i>	1st Recording session 10 × /tæ:ti/ 10 × /ty:ti/ → Recorded	<i>Audio-visual</i>	3rd Training session 30 × /tæ:ti/ – <tuuti> 30 × /ty:ti/ – <tyyti> → Not recorded
<i>Audio-visual</i>	1st Training session 30 × /tæ:ti/ – <tuuti> 30 × /ty:ti/ – <tyyti> → Not recorded	<i>Auditory</i>	3rd Recording session 10 × /tæ:ti/ 10 × /ty:ti/ → Recorded
<i>Auditory</i>	2nd Recording session 10 × /tæ:ti/ 10 × /ty:ti/ → Recorded	<i>Audio-visual</i>	4th Training session 30 × /tæ:ti/ – <tuuti> 30 × /ty:ti/ – <tyyti> → Not recorded
<i>Audio-visual</i>	2nd Training session 30 × /tæ:ti/ – <tuuti> 30 × /ty:ti/ – <tyyti> → Not recorded	<i>Auditory</i>	4th Recording session 10 × /tæ:ti/ 10 × /ty:ti/ → Recorded

**Figure 1.** The experiment procedure.

#### 4 Analysis

The acoustic recording data were analysed by a phonetician using Praat, version 5.3.01 (Boersma, 2001). A total of 80 tokens (10 productions of each word per recording session) were analysed from each speaker. We extracted the two lowest formant frequency values F1 and F2 (as well as F0) from the steady-state locations of the first syllable vowels using the Linear Predictive Coding (LPC) Burg algorithm. The values for F1 and F2 were then subjected to a repeated measures analysis of variance, and more importantly, they were analysed separately in more detail (ANOVA, IBM SPSS, version 22). In addition, we calculated each speaker's average standard deviation values of the formant values for both words from all ten productions in each of the four sessions. These data were statistically analysed using the same ANOVA models as in the initial analysis of the F1 and F2 values. The more detailed analysis was performed so that we extracted one variable at a time from the data set and analysed the remaining data with the same ANOVA model.

### III Results

The average Hertz values and the average standard deviations for both the target and the non-target vowels are shown in Tables 1 and 2.

**Table 1.** The average Hertz values for F1 and F2 in the four recording sessions.

		Session 1	Session 2	Session 3	Session 4
/tʌ:ti/	F1	437	447	451	453
	F2	2,148	1,849	1,792	1,792
/ty:ti/	F1	441	443	445	446
	F2	2,262	2,191	2,150	2,166

**Table 2.** The average of individual standard deviation values from all participants for F1 and F2 in the four recording sessions.

		Session 1	Session 2	Session 3	Session 4
/tʌ:ti/	F1	32	31	27	29
	F2	197	264	245	226
/ty:ti/	F1	26	25	28	30
	F2	115	137	112	140

We began the analysis by subjecting the average F1 and F2 values (Table 1) separately to omnibus ANOVAs with the factors of Session (first, second, third, fourth) and Stimulus (/ty:ti/, /tʌ:ti/). The analysis of the F1 values revealed no significant effects or interactions and thus we did not analyse the F1 values further. The analysis of the F2 values revealed a main effect of Stimulus ( $F(1,11) = 23.038, p = 0.001, \eta_p^2 = 0.677$ ) showing that the productions of the target and the non-target F2 values were systematically different. More importantly, we discovered a main effect of Session ( $F(3,9) = 4.775, p = 0.029, \eta_p^2 = 0.614$ ), suggesting an overall change as a function of training. In addition the Session  $\times$  Stimulus interaction ( $F(3,9) = 4.219, p = 0.040, \eta_p^2 = 0.584$ ) indicated that the words did not react similarly to training.

In order to clarify the role of the Stimulus in the analysis, we performed separate ANOVAs for the non-target and target vowel F2 values (Session: first, second, third, fourth). No significant effects were found in the F2 values of the non-target vowel and thus we continued the analysis only on the target vowel. A similar analysis on the target word F2 values revealed a main effect of Session ( $F(3,9) = 4.630, p = 0.032, \eta_p^2 = 0.607$ ), showing a significant change over time. These findings showed that while the F1 and F2 values for the non-target vowel remained unchanged, the F2 values of the non-native target vowel changed from /y/-like values towards /ʌ/ as a function of training. The effect of Session was investigated further for the target vowel /ʌ/ F2 values with a series of comparisons between sessions via ANOVA. These comparisons revealed a significant difference between sessions 1 and 2 ( $F(1,11) = 9.559, p = 0.010, \eta_p^2 = 0.465$ ), sessions 1 and 3 ( $F(1,11) = 16.759, p = 0.002, \eta_p^2 = 0.604$ ) and sessions 1 and 4 ( $F(1,11) = 14.865, p = 0.003, \eta_p^2 = 0.575$ ). This showed that the training effects were relatively quick (evident starting in session 2) and that they remained constant throughout the rest of the experimental procedure. We analysed all session pairs, but no significant differences between sessions 2 and 3, 2 and 4, and 3 and 4 were observed.



In order to study whether the larger standard deviations (SD) of the target vowel F1 and F2 were significant, we performed separate omnibus ANOVAs for the speakers' average F1 and F2 SD data with the factors of Session (first, second, third, fourth) and Stimulus (/ty:ti/, /tæ:ti/). There were no significant effects or interactions in the F1 SD values. The analysis of the F2 SD values revealed a main effect of Stimulus ( $F(1,11) = 16.436$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.599$ ), indicating that the larger SDs (Table 2) for the target vowel were significantly larger than in the non-target vowel. There was no main effect of Session, indicating that the formant SDs did not change between sessions. No other significant findings emerged.

Altogether, the statistical analyses showed that, despite the misleading orthographic input, the production of the target vowel /æ/ changed quickly towards /æ/. Moreover, it became evident that the changes were concentrated in the F2 values relevant for distinguishing between the categories /y/ and /æ/. In addition, the formant values in Table 1 show that the F2 values produced for the target vowel /æ/ remained considerably higher than the average F2 values for the Finnish vowel /u/ (adult norms F1 = 332 Hz, F2 = 690 Hz; Iivonen, 2012), which is linked with a more back articulation for Finnish /u/. Also, we were able to show that the native vowel did not undergo change as a result of the training and that despite the changes in the formant values, the standard deviations for the target vowel F2 remained larger than for the native vowel.

## IV Discussion

The acquisition of speech sounds is a significant part of the development of a language. In addition, reading skills play a role in how new languages are learned, since L1 grapheme–phoneme correspondences have an effect on how we perceive phoneme categories. Both adults and children encounter problems with the learning of a non-native language, but compared with adults, children benefit from neural plasticity. However, since neural commitment to L1 sounds happens already in early childhood, children are not immune to difficulties in the learning of new languages. For example, a study by Kuhl et al. (2008) found that infants' L1 and L2 phonetic perception abilities at 7.5 months of age predict their later language skills. This not only reveals the significance of phonological processing in L1 learning, but it also has important implications for L2 learning: the ability to distinguish L2 sounds is impaired in conjunction with the increasing dominance of the L1 sound categories. In fact, several studies have shown that the L1 dominance results in L2 perception and production difficulties irrespective of the age of acquisition or the learning setting (Peltola MS et al., 2003, 2007; Winkler et al., 1999).

Earlier studies have important implications for our main finding that children learn to produce new vowels quickly with only a short training paradigm containing both auditory and orthographic exposure. Firstly, it becomes evident that children are fast learners in comparison with adult (Peltola KU et al., 2015) and senior (Jähi et al., 2015) learners. The finding that child participants changed their production of /æ/ according to the acoustic input after only one training session suggests not only high plasticity, but also accurate analysis of the acoustic dimensions to be altered. The rather small difference in the acoustic values of the stimuli was clearly substantial enough to be perceived by child learners. The formant values for /æ/ produced by the children remained higher than the

exact values of the stimuli, but this was to be expected, as the stimuli were based on the voice of an adult male. Moreover, the acoustic data shows that instead of producing the Finnish vowel /u/ (F1 = 332 Hz, F2 = 690 Hz; Iivonen, 2012) as shown in the misleading visual cue, the children produced the vowel /ʉ/ according to the acoustic model. The /u/ F1 and F2 values described in Iivonen (2012) are from adult Finnish speakers. However, even considering the higher average formant values typical for child speakers, the F2 values produced by the participants of this study (Table 1) did not become /u/-like. Although the participants were exposed to the acoustic forms during familiarization before seeing the orthographic forms, the minimal auditory exposure (3 repetitions) is unlikely to affect the results. Secondly, the expected finding that the non-target vowel did not react to training implies that the stability of the native system was not affected by the non-native elements. This is contrary to an earlier finding, where child learners have shown changes also in the native categories as a function of exposure to a new language, but in that case the exposure took place in immersion education and the input was thus much more versatile (Peltola MS et al., 2007). Thirdly, despite clear production learning towards the acoustic stimulus, the fact that the standard deviations remained large for the target vowel indicates that variability in production persisted throughout the training. This probably suggests that the learning process is by no means finished by the end of this short exposure and/or that individual differences may play a role. To explore this effect in more detail, further research is needed.

The fourth conclusion is linked with the overall surprise that, contrary to our hypothesis, the misleading orthographic cues did not hinder production learning. On the basis of earlier adult data (Peltola KU et al., 2015), we assumed that the transparent Finnish orthography would result in problems in L2 sound contrast production in the audio-visual training paradigm, since the participants were already rather highly literate. However, this was clearly not the case, which suggests at least three alternative explanations: First, it may be that the grapheme–phoneme correspondence is not yet as strongly established at this stage of literacy development. Second, it could be that children are so attuned to auditory input that the changes take place irrespective of the visual cues. Most probably, however, the visual orthographic cues actually helped in directing attention to the non-native item. The third explanation seems plausible when combined with the observation that the participants in the present study changed their production faster than children in an earlier study where the same training paradigm with exactly the same amount of acoustic exposure was conducted with auditory stimuli only (Taimi et al., 2014). In Taimi et al. (2014) a listen-and-repeat protocol with no orthographic cues and the exact same acoustic ones changed the age-matched children's production by the third recording. In contrast, in the present study the changes took place already by the second recording. Altogether, it may be that the L1 grapheme–phoneme correspondences are not hard-wired at this literacy level and that the visual cue helps the child learners to focus on the differences in the auditory cues during training. It may also be that, in addition to the natural age-related aptitude, child learners may be more driven towards the acoustic cues when a task is not demanding as in the case of this study; adults, on the other hand, may be able to tolerate more difficult learning tasks to a better extent. But the interrelation between these factors cannot be teased apart in this experiment and further tests are

needed. In addition, it should be noted that the training paradigm in this study was not cognitively demanding and no acoustic variation inherent to normal speech was introduced, nor did we measure any potential long-term effects. Thus, new sets of data with longer exposures, more demanding trials and a longitudinal approach would address the complex task of learning in a more comprehensive manner.

## V Conclusions

In conclusion, our results show that children can learn to produce an extremely difficult non-native vowel contrast very rapidly with a training method that utilizes an audio-orthographic training paradigm. Surprisingly, the ability to read in the native language (and thus the potential to be misled by the native orthography) does not appear to hinder L2 production learning in an undemanding training task even though the orthographic cues provide misleading information according to L1 grapheme–phoneme correspondences. In this respect the child learners seem to be different from adults, who were extremely sensitive to the visual orthographic cues in this relatively easy task. In addition, it seems that the orthographically misleading visual cue does not hinder the learning process of a non-native sound. In comparison with age-matched children training with only acoustic stimuli, the participants of the current study actually changed their pronunciation with less training. The non-native target sound /ʉ/ is acoustically situated around the border of the native sound categories /y/ and /u/, which makes it difficult to learn according to the L2 learning models mentioned in this article. The findings of the present study suggest that orthographic cues may help to attract attention to non-native sounds. This enabled the participants to better focus on the demands of the learning task.

## Acknowledgements

We wish to thank MA Elina Lehtilä for her help in this study and Sanako Corp. for sponsoring the software used in data collection.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Katja Immonen  <https://orcid.org/0000-0002-9757-0178>

Kimmo U Peltola  <https://orcid.org/0000-0003-1214-6536>

Henna Tamminen  <https://orcid.org/0000-0003-1845-0854>

Paavo Alku  <https://orcid.org/0000-0002-8173-9418>

Maija S Peltola  <https://orcid.org/0000-0003-1639-2363>

## References

- Alku P, Tiitinen H, and 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: 1329–33.
- Bassetti B (2017) Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory and Cognition* 43: 1835–42.
- Best CT and Tyler M (2007) Nonnative and second-language speech perception. In: Bohn O and Munro MJ (Eds.) *Language experience in second language speech learning: In honour of James Emil Flege*. Amsterdam: John Benjamins, pp. 13–34.
- Boersma P (2001) Praat, a system for doing phonetics by computer. *Glott International* 5: 341–45.
- Cheour M, Shestakova A, Alku P et al. (2002) Mismatch negativity shows that 3–6 year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters* 325: 187–90.
- Escudero P (2005) *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: Netherlands Graduate School of Linguistics.
- Escudero P and Wanrooij K (2010) The effect of L1 orthography on non-native vowel perception. *Language and Speech* 53: 343–65.
- Escudero P, Hayes-Harb R, and Mitterer H (2008) Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60.
- Flege JE (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 JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (Ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press, pp. 229–73.
- Flege JE (1999) The relation between L2 production and perception. In: Ohala JJ, Hasegawa Y, Ohala M, et al. (eds) *Proceedings of the XIVth International Congress of Phonetic Sciences*. Berkeley, CA: Department of Linguistics, University of California, Berkeley, pp. 1273–76.
- Flege JE and Bohn O (2021) The revised Speech Learning Model (SLM-r). In: Wayland R (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 3–83.
- Flege JE and MacKay IR (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.
- Flege JE, Frieda EM, and Nozawa T (1997) Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–86.
- Flege JE, Yeni-Komshian GH, and Liu S (1999a) Age constraints on second-language acquisition. *Journal of Memory and Language* 41: 78–104.
- Flege JE, MacKay IR, and Meador D (1999b) Native Italian speakers’ perception and production of English vowels. *The Journal of the Acoustical Society of America* 106: 2973–87.
- Giannakopoulou A, Uther M, and 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: 201–18.
- Giannakopoulou A, Brown H, Clayards M, and Wonnacott E (2017) High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ* 5: e3209.
- Hayes-Harb R, Nicol J, and Barker J (2010) Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech* 53: 367–81.

- Iivonen A (2012) Kielten vokaalit kuuloanalisisessa vokaalikartassa [Vowels of languages in and auditory analog vocal map]. *Puhe ja kieli* 32: 17–43.
- Immonen K and Peltola MS (2018) Finnish children producing English vowels: Studying in an English immersion class affects vowel production. *Journal of Language Teaching and Research* 9: 27–33.
- Iverson P, Pinet M, and Evans BG (2012) Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics* 33: 145–60.
- Jähi K, Alku P, and Peltola MS (2015) Does interest in language learning affect the non-native phoneme production in elderly learners? In: The Scottish Consortium for ICPhS 2015 (ed.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.
- Kuhl PK, Conboy BT, Coffey-Corina S et al. (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 979–1000.
- Peltola KU, Alku P, and Peltola MS (2017) Non-native speech sound production changes even with passive listening training. *Linguistica Lettica* 25: 158–72.
- Peltola KU, Tamminen H, Alku P et al. (2015) Non-native production training with an acoustic model and orthographic or transcription cues. In: The Scottish Consortium for ICPhS 2015 (ed.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.
- Peltola MS and Aaltonen O (2005) Long-term memory trace activation for vowels depends on the mother tongue and the linguistic context. *Journal of Psychophysiology* 19: 159–64.
- Peltola MS, Kujala T, Tuomainen J et al. (2003) Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters* 352: 25–28.
- Peltola MS, Kuntola M, Tamminen H et al. (2005) Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters* 388: 121–25.
- Peltola MS, Tuomainen O, Koskinen M et al. (2007) The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. *Journal of Psycholinguistic Research* 36: 15–23.
- Rastle K, McCormick SF, Bayliss L, and Davis CJ (2011) Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory and Cognition* 37: 1588–94.
- Simon E, Chambless D, and Alves UK (2010) Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences* 32: 380–94.
- Taimi L, Jähi K, Alku P et al. (2014) Children learning a non-native vowel—The effect of a two-day production training. *Journal of Language Teaching and Research* 5: 1229–35.
- Tamminen H and Peltola MS (2015) Non-native memory traces can be further strengthened by short term phonetic training. In: The Scottish Consortium for ICPhS 2015 (ed.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.
- Winkler I, Kujala T, Tiitinen H et al. (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36: 638–42.