frontiers in Education

ORIGINAL RESEARCH
published: 20 May 2021
doi: 10.3389/feduc.2021.612457

The Effects of a Digital Articulatory Game on the Ability to Perceive Speech-Sound Contrasts in Another Language

Sari Ylinen[1,2]*, *Anna-Riikka Smolander*[1], *Reima Karhila*[3], *Sofoklis Kakouros*[3,4], *Jari Lipsanen*[5], *Minna Huotilainen*[1,2] *and Mikko Kurimo*[3]

[1]CICERO Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [2]Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [3]Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, [4]Department of Digital Humanities, University of Helsinki, Helsinki, Finland, [5]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

OPEN ACCESS

Edited by:
Stamatis Papadakis, University of Crete, Greece

Reviewed by:
Ping Tang, Nanjing University of Science and Technology, China
Isabel Cristina Ramos Peixoto Guimarães, Escola Superior de Saúde do Alcoitão, Portugal

***Correspondence:**
Sari Ylinen
sari.ylinen@helsinki.fi

Specialty section:
This article was submitted to Digital Learning Innovations, a section of the journal Frontiers in Education

Received: *30 September 2020*
Accepted: *05 May 2021*
Published: *20 May 2021*

Citation:
Ylinen S, Smolander A-R, Karhila R, Kakouros S, Lipsanen J, Huotilainen M and Kurimo M (2021) The Effects of a Digital Articulatory Game on the Ability to Perceive Speech-Sound Contrasts in Another Language. Front. Educ. 6:612457. doi: 10.3389/feduc.2021.612457

Digital and mobile devices enable easy access to applications for the learning of foreign languages. However, experimental studies on the effectiveness of these applications are scarce. Moreover, it is not understood whether the effects of speech and language training generalize to features that are not trained. To this end, we conducted a four-week intervention that focused on articulatory training and learning of English words in 6–7-year-old Finnish-speaking children who used a digital language-learning game app Pop2talk. An essential part of the app is automatic speech recognition that enables assessing children's utterances and giving instant feedback to the players. The generalization of the effects of such training in English were explored by using discrimination tasks before and after training (or the same period of time in a control group). The stimuli of the discrimination tasks represented phonetic contrasts from two non-trained languages, including Russian sibilant consonants and Mandarin tones. We found some improvement with the Russian sibilant contrast in the gamers but it was not statistically significant. No improvement was observed for the tone contrast for the gaming group. A control group with no training showed no improvement in either contrast. The pattern of results suggests that the game may have improved the perception of non-trained speech sounds in some but not all individuals, yet the effects of motivation and attention span on their performance could not be excluded with the current methods. Children's perceptual skills were linked to their word learning in the control group but not in the gaming group where recurrent exposure enabled learning also for children with poorer perceptual skills. Together, the results demonstrate beneficial effects of learning via a digital application, yet raise a need for further research of individual differences in learning.

Keywords: learning game, gaming, language learning, foreign language, speech sound discrimination, automatic speech recognition, educational applications, digital learning applications

INTRODUCTION

Smart devices provide possibilities to produce easily obtainable applications not only for entertainment but also for learning. Whereas video games may improve perceptual, attentional and some other cognitive abilities in adults and adolescents (Eichenbaum et al., 2014; Moisala et al., 2017; Bediou et al., 2018), learning games seem to be a good means to expose children to potentially useful learning materials in an age-appropriate manner. A possibility to use animations, pictures, and sound and to make games interactive via touch and speech enable different applications ("apps") for learning foreign languages (for novel word learning, see Russo-Johnson et al., 2017; Junttila and Ylinen, 2020). However, in most cases their effectiveness has not been experimentally tested (Hirsh-Pasek et al., 2015) and, therefore, it may be difficult to assess their pedagogical value (see Papadakis, 2020 and Papadakis et al., 2020 for a review and evaluation tools). Moreover, while there are demonstrations of how children's learning from apps generalizes to real life situations outside the apps (see Lovato and Waxman, 2016; Russo-Johnson et al., 2017), there is only scarce evidence on how learning in language-learning apps generalizes to stimuli or tasks that are different from those used during learning (cf. Tremblay et al., 1997). Such generalization could be very useful, as skills in one language could, for example, facilitate the learning of other languages (e.g., Thomas, 1988; Cenoz, 2013).

Learning foreign or second-language (L2) speech sounds is a challenge for many learners, yet intensive training typically results in measurable learning gains (Logan et al., 1991; Lively et al., 1993; 1994). Although speech production and perception training improve most consistently in the trained domain, a number of studies have provided experimental evidence that training effects in speech production and perception transfer from one domain to the other. Specifically, speech perception training may facilitate sound production learning (Rvachew, 1994; Bradlow et al., 1997). In a similar vein, speech production training has been shown to improve the perception of speech sounds (Catford and Pisoni, 1970; Kartushina et al., 2015), yet the effects of speech production training seem to be less consistent than those of speech perception training (see, e.g., Baese-Berk, 2019). Although several factors, such as learners' prior abilities or differences between the native language and L2, may modify the extent of training gains, the across-domain transfer effects are anatomically plausible in the brain: Speech production areas in the frontal lobe have strong reciprocal neural pathways to speech perception areas in the temporal lobe. Consequently, the temporal-lobe speech areas receive forward predictions from the frontal areas during overt or covert speech production (Ylinen et al., 2015), whereas the frontal lobes contribute to speech perception as part of the dorsal stream of speech processing, for example in speech imitation or repetition tasks (Hickok and Poeppel, 2007). Possibly, then, repeated attempts to produce foreign speech sounds during an intensive training period may attune this network and sensitize it to the processing of acoustic features, such as sound frequency. Simmonds (2015) has put forth a hypothesis that a potential trigger for such sensitization is the variation of signals in the neural pathways of the language network, which may cause a shift from a stable state to a learning state in this network, resembling the learning state of song learning in songbirds. Articulatory experimentation during foreign-language speech training may increase variation in neural signals and increase the likelihood of switching to such hypothesized learning mode.

To this end, the current study aims to determine whether articulatory speech training with a digital language-learning game shows any increase in the ability to perceive speech-sound contrasts from other languages without prior training (i.e., whether training effects are generalized to untrained items) and whether a possible increase in sensitivity applies beyond the speech-sound domain to different phonetic features. Since previous studies show that discrimination skills are linked to word learning in adults (Silbert et al., 2015), we also explore whether perceptual sensitivity is linked to children's ability to learn word meanings. To study the generalization of sensitivity across languages after gaming, we exposed 6–7-year-old Finnish-speaking children to English during a four-week period of playing a version of a digital language-learning game called Pop2talk with tablet computers (see www.pop2talk.com). The duration of training was similar to our previous study (4. 3 weeks in Junttila et al., 2020) and this age group was chosen because Pop2talk is intended for beginning learners of English and the learning of English (or some other foreign language) starts at this age (i.e., in the first grade) in Finland. Before and after the four-week period, we compared perceptual skills and learning in children who had or had not played the game (the gaming and control groups, respectively). Specifically, we tested the learning of English words with a vocabulary task and the effects of gaming on the perception of other languages with a discrimination task. We tested the discrimination of two phonological contrasts from different languages, either of which was not trained in the game. Next, we will introduce the game, the contrasts used in the context of the models of phonetic learning, and our hypothesis.

Pop2talk game (and its predecessor, a digital board-game Say it again, kid! Ylinen and Kurimo, 2017; Junttila et al., 2020) is based on listening to and producing speech in English; first by imitation and later by free recall. No reading skills are needed in the game. An integral part of the game is its speech interface enabled by automatic speech recognition. The automatic speech recognizer used in the game can evaluate each speech sound from children's speech and give instant feedback about the accuracy of their utterances (Karhila et al., 2017, 2019). Feedback is expected to encourage children to make an effort to form accurate perceptual and articulatory representations of the words. Repetitive articulatory attempts may also increase the variability of signals in the speech networks of the brain. According to Simmonds (2015), such variability might result in the activation of the language-learning mode, which in turn may enable more native-like foreign-language articulation.

It is well-established that the perception of non-native phonetic contrasts is more effortful and error-prone than that of native contrasts [see Speech Learning Model (SLM) by Flege, 1995, Perceptual assimilation model (PAM) by Best et al., 1988, Best, 1994, and Native Language Neural Commitment (NLNC)

model by Kuhl, 2004]. Here we tested the discrimination of two different phonetic contrasts since we wanted to explore features that differ from the trained English features to a different degree. The first contrast we used was a Russian consonant contrast /z/(a voiced dental/alveolar sibilant) vs. /ʒ/(a voiced palato-alveolar/ retroflex sibilant, with a different place of articulation and a different position of tongue). Since there are no voiced sibilants in Finnish, these sounds were not familiar from the native language. /z/ is common in English, yet the children were not exposed to it in the game because no words with /z/ were included in its setup. /ʒ/ is less common and it did not occur in the game either (one word, *hedgehog*, included the affricate /dʒ/, but it is acoustically quite deviant from the Russian /ʒ/). Thus, the children were not exposed to sounds resembling the Russian /z/ vs. /ʒ/contrast in the native language or in the game. According to predictions by PAM (Best et al., 1988; Best, 1994), both Russian consonants may be poor exemplars of the Finnish sibilant (or, possibly, different enough from the Finnish sibilant to be uncategorized sounds). The discrimination of these consonants was, therefore, expected to tap the children's sensitivity to their acoustic features. The second contrast we used was a Mandarin lexical tone contrast between a flat vs. rising tone. Tone is used lexically neither in Finnish nor in English, so there was no exposure at all to this feature in a phonological sense. Although both languages naturally use pitch in their prosody, the Finnish intonation patterns do not typically include rising pitch (not even in question phrases; Iivonen, 1998). Therefore, the discrimination of these tones was again expected to tap the children's sensitivity to pitch as an acoustic feature. According to NLNC (Kuhl, 2004), these contrasts should lack neural commitment in native Finnish speakers' brain.

We hypothesize that if the variability of neural signals in the language networks (Simmonds, 2015) increases children's perceptual sensitivity to untrained sounds as a result of articulatory experimentation in a trained language, the effect may be generalized to different untrained phonetic features or it may be limited to new contrasts utilizing the trained feature. Specifically, as in the current study the children trained the production of English vowels and consonants that are cued by spectral information (e.g., formants and their transitions), we may hypothesize improvement of discrimination in Russian consonants that introduce a new center of gravity of spectral features and Mandarin tone that uses a different feature (pitch), if generalization takes place across features. An alternative hypothesis would be that improvement of discrimination takes place in Russian consonants introducing a new center of gravity of spectral features, but not in Mandarin tone that uses a different feature. Regarding the link between discrimination and word learning, we expect that children with good discrimination skills may also learn more words (Silbert et al., 2015).

Since our participants were children, we aimed to run the experiments as quickly as possible. To avoid multiple repetitions of all stimulus pairs of our stimulus continua, we used an adaptive task (up-and-down design) where a phonetic contrast gets more difficult after a correct response and easier after an incorrect response (for a review, see Treutwein, 1995). Nevertheless, psychometric tests are challenging for child participants with fluctuating and short attention spans, which should be taken into account in the interpretation of results in any psychometric experiment.

MATERIALS AND METHODS

Ethics Statement

Participants' caregivers signed a written informed consent form and the participants gave their oral consent before participation in the experiment. The study was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences.

Participants

Participants were 6–7-year-old children (mean age 7.13 years) who were assigned to gaming and control groups. All children studied in the first class in school. The children had started to learn English in school at the beginning of the first class. That is, the control group participated the usual English lessons and the gaming group played the game in addition to their English lessons. The inclusion criteria of the participants, whose background information was obtained from parental reports, were as follows: Finnish as the native language, normal hearing, no bilingualism, no diagnosis with language or learning deficits, and no diagnosed neuropsychological or neuropsychiatric deficits. However, we included children who had relatives with dyslexia or who had, at some stage, consulted a speech therapist, a psychologist, or a pediatric occupational therapist, yet did not have any diagnosis (their performance did not consistently deviate from the others; see Discussion). In line with STROBE statement (2021), see Results for further participant details.

Gaming

Our gaming intervention used a digital game app called Pop2talk, developed in-house by using Unity (Unity Technologies, San Francisco, United States). Pop2talk includes three in-house manufactured components: the language-learning game and automatic speech recognition and rating systems specially tailored for Finnish children who are learning English as a foreign language.

In the game, players popped geometric shapes on a touch screen where each tap triggered a replay of an English word. A word replay could also be triggered by the shapes popping spontaneously. After presentation of 4 (or more, depending on spontaneous popping) English stimuli, a card with a picture referring to the word to be learned opened on the screen. The word was first heard in Finnish and then in English. Then a microphone icon was lit up in the card, accompanied by a sound signaling the opening of the microphone. At this stage, children's task was to imitate the heard English word aloud (in a time window of 2 s). After the microphone had closed, the children heard back their own utterance and the English model they had heard. Then the children got one to five stars as feedback from the

automatic speech recognition (typically within a few seconds depending on the internet connection).

The gaming period lasted for four weeks, during which the children played on four days per week for about 15 min per day. Thus, the overall exposure to the game was about 4 h. The pre-tests were conducted one day before the gaming period and the post-test took typically place on the next day after the gaming period. In two cases, testing was delayed over the weekend (one child from the gaming group and one control child were absent from school on the test day), resulting in a three-day delay. In both cases, the participants' performance was above the group average, so it is unlikely that the delay deteriorated their performance.

The gaming-group children were exposed to 66 English words representing common English words and referring to things that were expected to be age-appropriate and familiar to children from their typical environment (e.g., *house, cat, read*). Typically, each word was repeated three times, but certain four words (*child, wash, mouth, feather*) were met more often (26, 22, 18, and 14 times, respectively) during the four-week period to see the effect of repetitive exposure. Each of these four words included a phoneme that does not belong to the Finnish phonology. The game proceeded from one level to another, so that children played two or three levels in each session. Each day, the children were introduced with two test cards where they were expected to demonstrate their word learning by producing the requested words without a model by free recall.

Testing Procedure

The participants were assigned to gaming (N = 42) and control groups (N = 38) semi-randomly so that children in the same class typically belonged to the same group. The assignment into groups was not fully random since for practical reasons it was possible to arrange gaming in a certain number of classes and schools only. A half of the participants within each group was tested with Russian and another half with Mandarin in the discrimination pre-test. However, as a result of unbalanced dropping out or exclusion of participants across the groups and languages, the final samples for each language were not equal (gaming with Russian: N = 17; gaming with Mandarin: N = 25; control with Russian: N = 21; control with Mandarin N = 17).

The participants attended five different public schools with the same curriculum. From three schools, participants were assigned to both gaming and control groups, one school had only gamers (because of providing few participants), and one school had only controls (because more controls were needed). We aimed to roughly balance the socioeconomic status (SES) on the level of the school neighborhood between the groups [at any rate, based on previous PISA reports (see e.g., Education GPS OECD, 2021), possible differences in quality of teaching between the schools were expected to be minimal in Finland]. Due to the practical arrangements of gaming, the tests were conducted in three samples. The first sample included 20 gamers and 18 controls (13 and 15 in the final sample, respectively) in September-October. After a one-week holiday in mid October, the second sample with 33 gamers and 12 controls (29 and 11 in the final sample, respectively) participated in October-November. Since we had not got enough controls, we started a recruitment process in November-December, yet it turned out to be unsuccessful because of approaching holidays. The third sample including only controls (a total of 13 children, 12 in the final sample) was tested in January-February. Since we were not able to test the groups at the same time, the learning of English in school had proceeded for several months longer for some of the controls (however, see Results for the comparison of scores between the groups).

Testing took place before and after a gaming period (or before and after the same time period for the control group) in a quiet room in school premises. For practical reasons, the children were tested in groups (maximum 6 children per group). A test session took less than 30 min, including instructions on each task, discrimination practice, discrimination test, word comprehension and word production tests. Total testing time could also be shorter, since the duration of the discrimination test depended on performance and production task was introduced to the gaming group only. The production task was performed first, because it was embedded in the game. This was followed alternatively by the word comprehension or discrimination task and then the other one.

Experimental stimuli and game sounds were presented via earphones (a headset with a microphone). Experimenters helped the participants to adjust the sound level and ensured that they considered the sounds clearly audible. A tablet computer with a touch screen was used to collect children's responses in discrimination and word comprehension tasks.

Discrimination Test

Stimuli

For the discrimination experiment, we chose lexical contrasts from the Mandarin Chinese and Russian languages, which were unfamiliar to the participants. The Mandarin contrast included two lexical tones [/ma/ with a flat tone (tone 1), "mother" vs. /ma/ with a rising tone (tone 2), "numb"]. The Russian stimuli were contrasted by two sibilant consonants [/muza/ "muse" vs. /muʒa/ "husband" (genitive)]. Note that although this contrast exists in English, /ʒ/ is infrequent (0.09%; Mines et al., 1978), it is observed in very limited contexts, and there are only a very few /z/-/ʒ/ minimal pairs. The Pop2talk version used did not include any words with /z/ or /ʒ/ although a word with the affricate /dʒ/ was included.

The original recordings were conducted in a sound-isolated recording studio with an AKG C2000B microphone, a Universal Audio Apollo Twin USB audio interface, and Audacity audio editor for the recording of the stimuli. The recordings were carried out by a female native speaker of Chinese and a female native speaker of Russian (bilingual Russian-Finnish) who uttered the requested words several times. From these recordings, we selected the exemplars where the target contrast was acoustically maximally salient yet the other acoustic properties were similar. With these stimulus pairs, stimulus continua were synthesized using TANDEM-STRAIGHT, a glottal and mixed/impulse (shaped pulse with noise) excited vocoder (Kawahara, 2006; Kawahara et al., 2008). Using TANDEM-STRAIGHT, the original speech stimulus pairs are decomposed into a set of real-valued parameters that can be

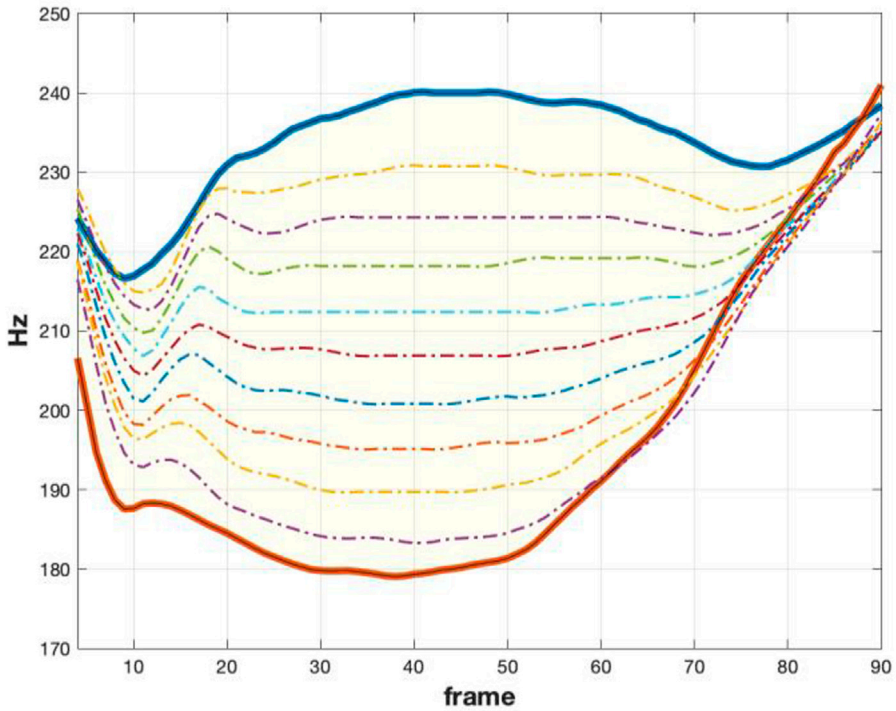

FIGURE 1 | F0 for /ma/ stimuli continua. Top and bottom F0 trajectories (marked with bold lines) are the exemplars with the highest target contrast (original recorded stimuli). Intermediate dash-dotted lines (in the shaded yellow region) are the synthesized trajectories between the two extrema.

manipulated without inconsistencies between the values. The intermediate variants of the stimuli can then be generated based on an interpolation of the STRAIGHT parameters (e.g., F0, STRAIGHT spectrogram, periodicity spectrogram) that can be modified to generate an equidistant division along a continuum between the two stimulus extrema. Specifically, TANDEM-STRAIGHT decomposes the information from speech into two components: (i) the source and (ii) the filter component. The source consists of the F0 and an aperiodicity spectrogram that captures the ratio between random and periodic components for each frequency band in the signal (see also Kawahara et al., 2009 for a description of the method and Kawahara et al., 2008 for the implementation details). The filter component consists of a representation of the spectrum of the signal in time. This is a two-step operation resulting in the STRAIGHT spectrogram (a smoothed time-frequency representation): first the TANDEM spectrogram is computed that is then used for the computation of the STRAIGHT spectrogram (see Kawahara et al., 2009)—note that TANDEM-STRAIGHT (Kawahara et al., 2008) is a refinement of STRAIGHT (Speech Transformation and Representation based on Adaptive Interpolation of weiGHTed spectrogram; Kawahara and Irino, 2005) where both tools use decomposition of the input to the same set of source and filter parameters that are referred to as STRAIGHT parameters. These parameters from the source and filter components can be modified to synthesize the variants of the input speech. For both contrasts, the variation was anchored to start just before the target phoneme and to end just after the target phoneme (where possible). The early onset and late cut-off of the anchor targeted a smooth transition to the target phoneme in order to obtain a natural sounding stimulus. The STRAIGHT parameters were then interpolated at equidistant steps to get the continua. For the tonal contrast, interpolation was performed in the F0, whereas for the sibilant contrast, interpolation was performed in the spectrogram, that is, the parameters of the filter component were modified (see **Figure 1** for /ma/ and **Figure 2** for /muza/-/muʒa/). From the resulting set of parameters, the stimuli were synthesized. The stimulus continua for both sounds were generated in 11 steps, because this was the optimal partitioning of the space between the two extrema, enabling us to get a fine-grained representation of the continuum while maintaining sufficient spacing between steps. Enough steps were needed to form a continuum of stimulus pairs with increasing difficulty and to avoid ceiling and floor effects.

Testing Procedure for Discrimination

Participants within each group were assigned to either language in the pre-test, but in the post-test both languages were tested. This was to enable seeing possible gaming effects on a language that had not been heard before (even in the pre-test).

The discrimination test had XXY design and it was implemented as a digital app designed with Unity. Specifically, in each trial, participants heard three stimuli, out of which two were the same and one was different. Simultaneously, they saw on the screen three note symbols. One of them trembled when a

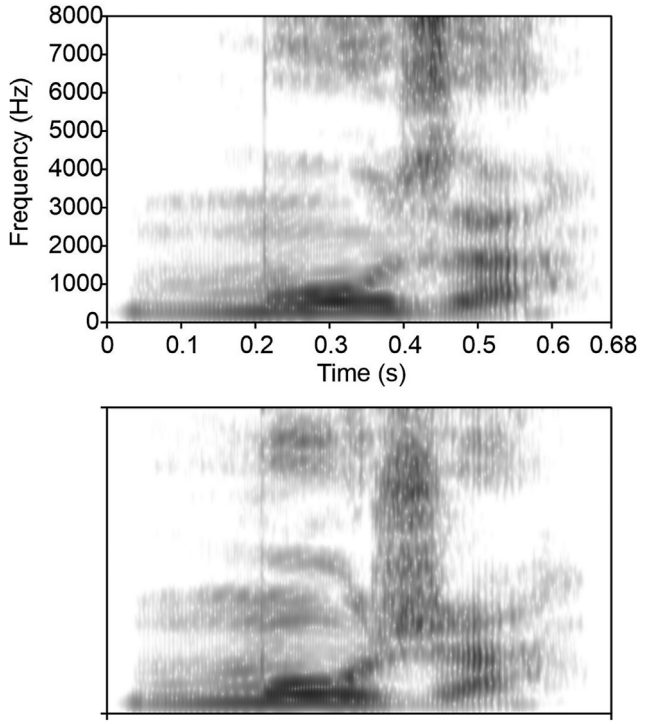

FIGURE 2 | Spectrograms of the two extremes of the Russian /muza/-/muʒa/. continuum. **Top:** spectrogram for muza (stimulus 1). **Bottom:** spectrogram for muʒa (stimulus 11). Note sibilant differences at higher frequencies around 400 ms.

stimulus was heard. The participants' task was to indicate the odd stimulus out by pressing its symbol. They had a chance to re-listen each trial once. The press of a symbol launched the next trial. The first stimulus was presented with a 1.5 s delay from the trial onset and inter-stimulus-interval was 1 s. Inter-trial-interval was 1.5 s. The position of the odd stimulus within each trial and the selection of the odd stimulus within a stimulus pair were random.

The items of the stimulus continua were paired as follows: steps 1 vs. 11 (level 1), 2 vs. 10 (level 2), 3 vs. 9 (level 3), 4 vs. 8 (level 4), and 5 vs. 7 (level 5). Since the acoustical difference was maximal at level 1 and got smaller across levels, the contrasting pairs formed continua from the easiest (level 1) to the most challenging (level 5). The easiest stimulus pair 1 vs. 11 was expected to be very easily discriminable, whereas the most difficult stimulus pair 5 vs. 7 was clearly more difficult. However, even the most difficult pair 5 vs. 7 was potentially perceptible, since it had a two-step difference (stimulus 6 was not used). We did not want to introduce the children with a task that was impossible to solve.

The task was adaptive with respect to the contrast difficulty. It always started with the easiest stimulus pair with maximal difference (level 1). If the response was correct, level 2 stimulus pair was presented. If the response was incorrect, the level 1 pair was presented again. If participants entered some other level than 1 and responded correctly, they got to the next level and heard a more challenging contrast, except for level 5 where they heard the same contrast again. If they responded incorrectly at levels 2–5, they got to the previous level and heard an easier contrast. Since children's task could not be overly long, the task was terminated after 10 turns in the response function (a turn from correct to incorrect or from incorrect to correct). The task was also terminated if there were five consecutive correct responses at level 5 or five consecutive incorrect responses at level 1, as this was interpreted to reflect ceiling or floor effects (or, in case of 5 incorrect responses after some correct ones for level 1, intention to respond incorrectly against instructions). Participants' responses were recorded.

Before the actual discrimination task, the children had four practice trials. These trials were otherwise identical to the test trial but the stimuli were different. Two of the stimuli were English words/ʃu:/ and the different stimulus was Finnish /su:/(used in Ylinen et al., 2019), which was expected to be quite easily discriminable and thus illustrate the nature of the task well.

Data Analysis for Discrimination

Pre- and post-test response functions were analyzed by calculating scores from the average of all turning points within a session for each language. In case of floor (all responses incorrect) or ceiling effects (5 consecutive correct responses at level 5), the score was set to 1 or 5, respectively. To enable comparability between languages, the data were normalized by calculating z scores based on pre-test averages and standard deviations across groups for each language (N = 48 for Mandarin, N = 43 for Russian). Thus, z score 0 represented the average performance, positive values better than the average and negative values poorer than the average. To analyze the change between pre- and post-tests, the pre-test z scores were subtracted from the post-test z scores.

Vocabulary Test

The vocabulary test, also implemented in digital format with Unity, included word comprehension and production tasks. The word comprehension task was conducted before and after the gaming period to observe learning effects, whereas the production task was conducted only after that as it was considered too difficult for the pre-test. In the word comprehension test, 15 pictures were shown on the screen (10 were matched with auditory stimuli and 5 were foils; see **Figure 3**). Then the participants heard 10 English words in random order (*child, wash, read, shirt, mouth, movie, river, feather, quiet, hedgehog*) one at a time and their task was to touch the picture corresponding to the word. After a small delay, the next trial was introduced. To proceed in the test, the children were instructed to guess if they did not know the meaning of the word. In the word production test, the participants played the game as usual, but no words were heard at any phase either in English or Finnish. This test game had four cards that were shown on the screen one at a time between the popping of the shapes, and the children were requested to say aloud the English word corresponding to the picture displayed in the card. The word production test is not reported here since it was not conducted with the control group. The control group was not specifically

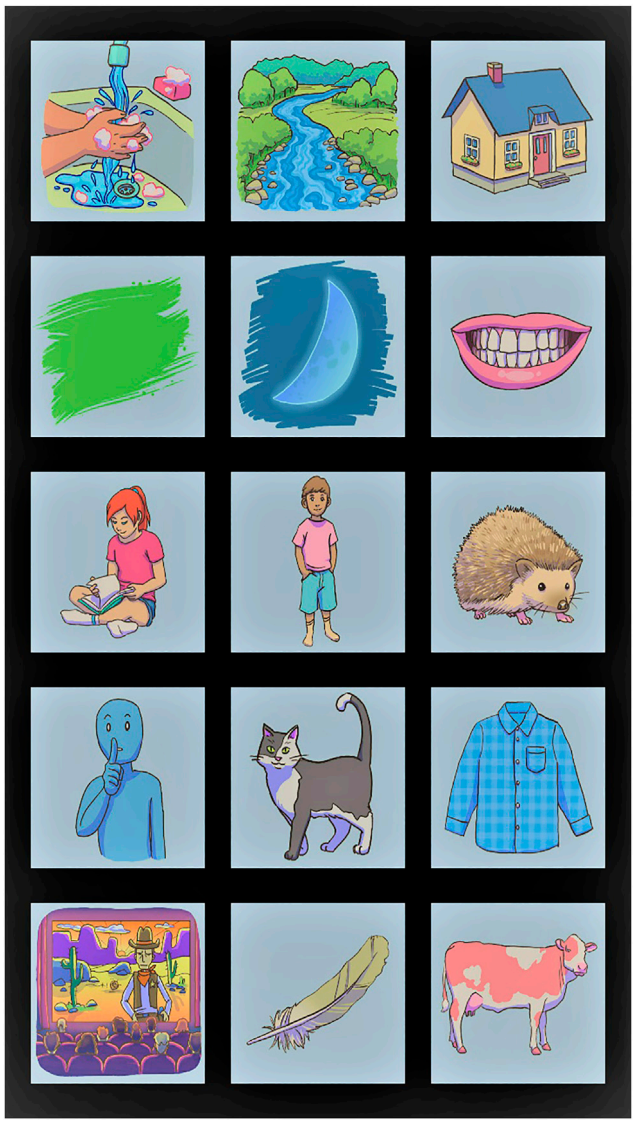

FIGURE 3 | A screenshot from word comprehension test.

exposed to the words and the word production task was expected to be too difficult for the controls (in the word comprehension test, the controls could just guess if they did not know the meanings of the words).

Statistical Analysis

Statistical analysis was carried out using a linear mixed model where independent variables were standardized pre-test score, group, language of the pre-test and language of the post-test. All possible interactions were included in the model. Based on Schwartz's Bayesian information criteria (BIC), random intercept model was observed to provide the best fitting error covariance structure. To further interpret these results and to find out whether there was a statistically significant increase in discrimination scores from pre-test to post-test, we used one-sample t-tests (one-tailed) comparing the difference z score (post-test minus pre-test) with zero, which corresponds to no change between pre- and post-test.

To determine whether children's perceptual skills were linked to vocabulary learning, we used Pearson's r for the post-test discrimination z scores (average across languages) and the scores of the word comprehension test (post-test minus pre-test difference scores). The alpha level was 0.05 throughout.

RESULTS

We tested 96 participants, but three participants did not finish the post-test, nine participants were excluded because of losing the post-test data due to a technical error, and one participant was excluded due to playing the game despite belonging to the control group. In addition, to avoid including children who performed intentionally poorly in the post-test despite showing better than average discrimination skills in the pre-test, participants were excluded from analysis if their z score had worsened over two standard deviations (SD) between the tests (i.e., post-test minus pre-test z score was -2 or lower). On this basis, three participants were excluded. The final sample was thus 80 participants (mean age 7.13 years; 39 girls, 41 boys).

A linear mixed model showed that the two-way interaction between group and pre-test score was statistically significant [$F(1, 72) = 4.82$, $p = 0.03$]. Also the main effects of pre-test score [$F(1, 72) = 12.47$, $p = 0.001$] and language were significant [$F(1, 72) = 10.08$, $p = 0.002$]. Interpretation of these results was that the post-test score for Russian was on average higher than the Mandarin score and that the pre-test scores were higher in the gaming group than in the control group, which is problematic for the interpretation of post-test results. To eliminate these pre-test differences, we calculated post-test minus pre-test z scores (i.e., we subtracted out the pre-test effect) and compared the difference scores to zero with one-sample t-tests. Although some improvement was observed for Russian in the gaming group (see **Figures 4**, **5**), all comparisons were non-significant [gaming group Russian: $t(16) = 1.675$, $p = 0.057$; control group Russian: $t(20) = 0.661$, $p = 0.258$; gaming group Mandarin: $t(24) = -0.153$, $p = 0.440$; control group Mandarin: $t(16) = 0.046$, $p = 0.482$].

A closer look at individual data reveals that the gamers' z scores for Russian divide into two clusters (see **Figure 5**). Specifically, in one cluster, six children show a marked improvement (2 SDs or more), whereas in the other cluster it is less so: four children show small improvement (0–1 SD), four children perform slightly more poorly after gaming (up to -1 SD), and three children show a clear drop (from -1 up to -2 SDs) in their performance after gaming. To clarify the role of children's background on the results, we looked at whether participants' background or aspects related to the experiment (possible risk factors, school, or timing of participation in the experiment) explained their performance. Z score ranges were from -1.58 to 2.79 in children who had relatives with dyslexia, from -1.89 to 2.29 in children who had, at some stage, consulted a speech therapist, from -1.58 to 1.27 in children who had consulted a

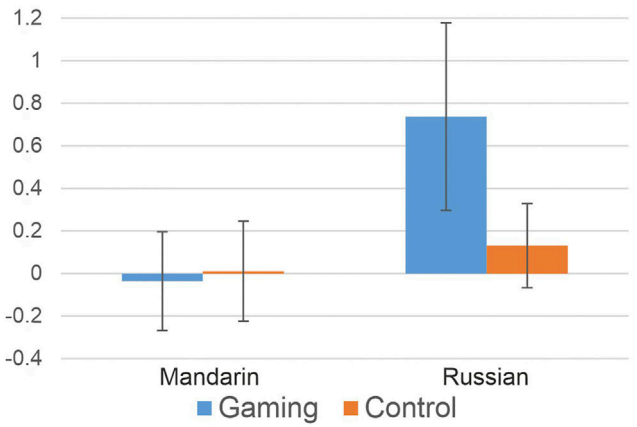

FIGURE 4 | Post-test minus pre-test difference discrimination z score for the gaming and control groups in Mandarin tone **(left)** and Russian sibilant **(right)** contrasts. Bars show the standard error of the mean. Zero denotes no change across sessions, positive values improvement and negative values denote poorer discrimination performance.

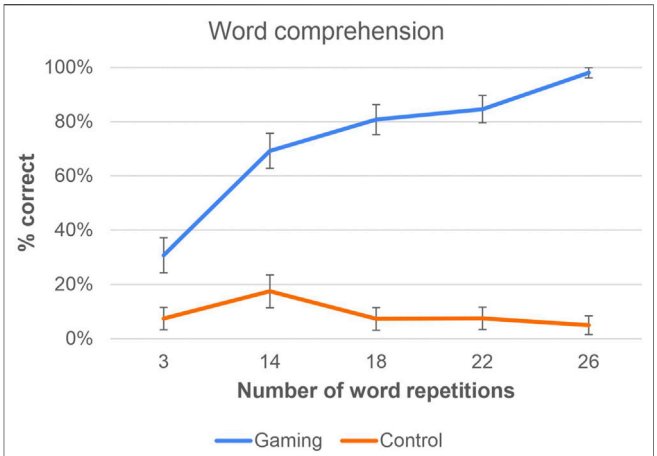

FIGURE 6 | The change between pre- and post-test in word comprehension (post-test minus pre-test percentages) as a function of number of times the words were repeated in the game. Note that the repetition concerns only the gaming group and the control group was not exposed to the words (their percentages reflect learning from other sources).

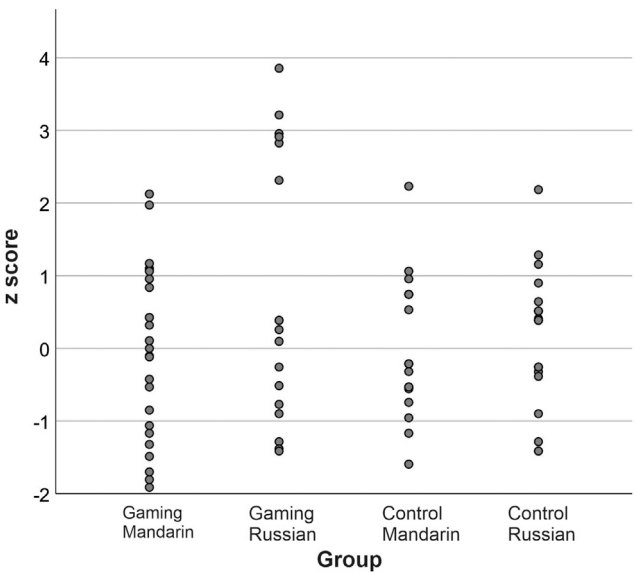

FIGURE 5 | Variation of post-test minus pre-test discrimination z scores across groups and languages in the discrimination task. Zero denotes no change across sessions, positive values improvement and negative values denote poorer discrimination performance.

psychologist, from -0.21 to 3.80 in children who had consulted a pediatric occupational therapist and from -1.47 to 3.17 in children with no such background factors. In participants attending different schools, z score ranges (and N) were as follows: 1) -1.89–2.1, N = 21; 2) -1–3.17, N = 7; 3) -0.25–2.92, N = 4; 4) -1.68–3.81, N = 36; 5) -1.58–2.21, N = 12). Z score ranges for participants, who were tested at different times, were -1.89–3.17 for the first sample, -1.68–3.81 for the second sample, and -1.57–2.21 for the third sample, which was also the smallest one.

To clarify the results, we were interested in the effort the children made in the discrimination task, and therefore we also looked at the frequency of re-listening stimulus trials for better performance. In the pre-test, the gaming and control group children used the chance of re-listening to the trials in 11 and 16% of the trials, respectively, whereas in the post-test the gaming and control group children re-listened to the trials in 7% and 10% of the cases, respectively.

The results of the word comprehension test are shown in **Figure 6** as a function of the number of word repetitions in the gaming group. The control group was not exposed to these repetitions, yet some learning was observed in this group, too. In the control group, post-test discrimination z scores correlated with word learning [$r(38) = 0.412$, $p = 0.01$]. No correlation was found in the gaming group [$r(44) = 0.114$, $p = 0.46$] (see **Figure 7**).

DISCUSSION

After eliminating pre-test differences between the groups, we found no significant generalization effects in the post-test although we observed some non-significant improvement in the discrimination of Russian consonants in the gaming group. Children's perceptual sensitivity was linked to spontaneous vocabulary learning in the control group, whereas no such link was found after gaming.

The finding of some improvement in the discrimination of the Russian contrast in the gaming group may suggest generalization effects in some individuals, yet they were not consistent across children and thus did not reach significance. Therefore, we explored whether participants' background information explained their performance. We found no consistent differences between children who had previously consulted a speech therapist, a psychologist, or a pediatric occupational

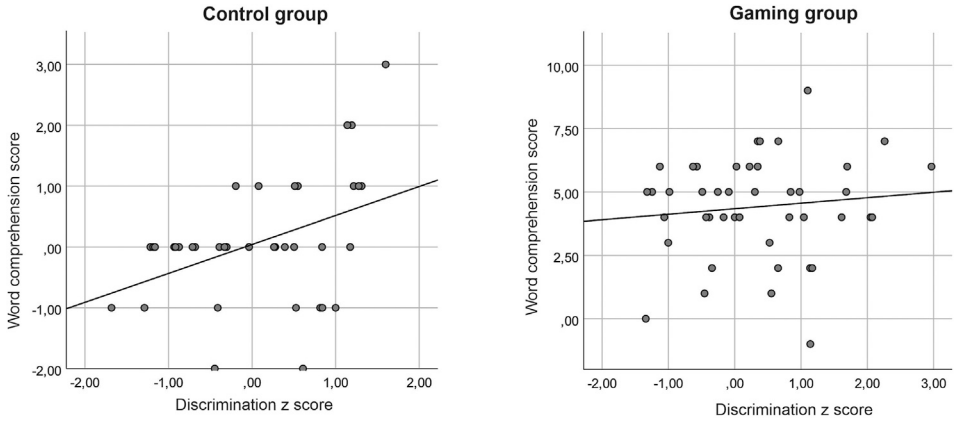

FIGURE 7 | Correlations between non-native discrimination z score (average across languages) and word comprehension score (post-test minus pre-test score) in the control **(left)** and gaming **(right)** groups. The word comprehension score may be negative, if there was a correct response in the pre-test but not in the post-test. Note different scales in the panels.

therapist or children who had relatives with dyslexia and children who did not have such background factors. Neither did we observe consistent differences between children attending different schools nor between participant samples that were tested at different times. Rather, variation seemed to be larger within than between these subgroups. Thus, the background factors may not explain the results. We can only speculate whether robust effects would have been obtained with longer training or without losing data because of technical problems. Nevertheless, some gamers' scores for Russian showed a marked improvement (2 SD or more), whereas some gamers show a clear drop in the post-test (up to -2 SD, see **Figure 5**). It does not seem likely that gaming would cause totally opposite effects to children's actual ability to discriminate the consonants. Rather, some other factors contributing to testing may account for this pattern.

Considering the causes of performance change in the post-test improvers, it is noteworthy that improvement for Russian was more modest in the top improvers of the control group than in the gaming group (**Figure 5**). Thus, it is plausible that the gaming group improvers increased their sensitivity to the foreign sound contrasts due to the game rather than due to some other factor, such as learning to discriminate the stimuli in the pre-test, spontaneous learning, or other development, which is expected to be similar in the control group. In the improvers, the training effects are likely promoted by connections between speech production areas in the frontal lobe and speech perception areas in the temporal lobe (see Hickok and Poeppel, 2007), enabling speech production training to improve the perception of speech sounds (Catford and Pisoni, 1970; Kartushina et al., 2015). However, we see no improvement for the Mandarin tone contrast in the gaming group (if anything, there is a slight decrease). Thus, it seems that the effects of articulatory gaming that trains non-native vowels and consonants might in some individuals (although not robustly) generalize to untrained speech-sound contrasts of a third language, increasing perceptual sensitivity to them (cf. Tremblay et al., 1997). However, the training effects do not seem to generalize to different phonetic features, such as tone cued by pitch in the current study.

In the lower end of the performance distribution, there are several possible accounts for poorer post-test performance. This is because children's behavioral performance in a discrimination task is not determined by their discrimination ability only, but also by their alertness, attentiveness, willingness, and motivation to perform the task. For example, possible accounts for poor performance include lack of effort (random responding) or willingness to terminate the task as quickly as possible (intentional incorrect responses, if they figured out that certain number of incorrect responses is a termination criterion). Since the discrimination task was not very interesting and did not include any rewards, it might be that some children were less motivated to focus on it in the second testing session than in the first testing session, when the task was new.

To better understand children's level of effort in the discrimination task, we calculated the proportion of re-listening the test trials for better performance and found that re-listening had decreased in the post-test for both groups. The smaller percentage of re-listening is likely not linked to improved perceptual skills because the adaptive task adjusted the difficulty level along with performance. Rather, it may indicate less effort in the post-test. Thus, it is not excluded that some children's lack of effort has exerted negative influence on our training effects. In addition to the test itself, also children's motivation for gaming may affect the results. According to experimenters' notes, two out of three children with the lowest (<-1) post-test minus pre-test difference z scores expressed at some point that they were not motivated to play although players typically seemed motivated.

Finally, we also looked at vocabulary learning. It was not reasonable to directly compare word learning statistically between the groups; it is clear that the gaming group learned considerably more than the control group that was not exposed to the words unless there was incidental exposure to the words in the classroom, media or equivalent thereof. Rather, we explored the link between children's discrimination skills and their word learning (here, associating English words with pictures) during the

four-week period between the pre- and post-tests. Although we did not expose the control group to the tested words, the vocabulary test showed some learning in this group as well. Correlation analysis in the control group showed that children's discrimination skills were linked to their vocabulary learning: those with better perceptual skills learned more words between the tests. This result is in line with previous findings in adults (Silbert et al., 2015). It is noteworthy that we could not control for children's exposure to English in this group. The control children may have learned the words in the classroom or learning may have occurred spontaneously from media or other games. Not surprisingly, targeted game training resulted in higher vocabulary scores in the gaming group. In contrast to the control group, however, we found no correlation between children's discrimination skills and vocabulary learning in the gaming group. This lack of correlation suggests that playing the language-learning game may have supported word learning particularly in the gaming-group children with the poorest perceptual skills. This is because it is plausible that in line with the control group, other exposure would have enabled them to learn fewer words compared with their peers with better perceptual skills. These results obtained by using an articulatory language-learning game emphasize the role of active speech production in children's word learning, including the learning of word meanings (see Icht and Mama, 2015; Junttila and Ylinen, 2020). Our current results on word learning also complement our previous findings on game-based language learning, which suggested better sensitivity to trained foreign speech-sound contrasts after game-based learning than after using a non-game application (Junttila et al., 2020).

The current study has, however, some limitations. For example, for practical reasons we were not able to test the groups at the same time, and therefore the learning of English in school had proceeded for several months longer for some of the controls. In a similar vein, we had to assign the participants into groups semi-randomly rather than randomly to arrange the training. Some data were lost because of technical problems and therefore groups were smaller than intended. In addition, the current study faced the challenges of conducting psychometric experiments with children. In particular, it is difficult to know a specific reason for poor performance because it may have several accounts, such as fluctuations in alertness and motivation, and it may or may not be intentional. Although psychometric experiments are, in general, a valid method to study perceptual sensitivity, the results may be somewhat distorted if participants do not cooperate and try their best. Further research is needed to clarify the factors underlying individual differences in the change of perceptual sensitivity. In line with Tremblay et al. (1997), further research with some other than psychometric methods would be interesting, if they were more sensitive in the study of children's perceptual abilities. For example, event-related potentials (ERP) measured with electroencephalography (EEG) in a passive paradigm that does not require active responding allows avoiding motivational effects on results. In addition, the measurement of the mismatch negativity (MMN) component of ERPs would allow avoiding the effect of children's attentional fluctuations, since the MMN reflects pre-attentive processing abilities and is elicited when attention is directed elsewhere (see Näätänen et al., 2007, for a review).

In sum, we found that game-based articulatory speech-sound training had no consistent generalization effect on perceptual sensitivity in untrained languages. However, in some individuals the sensitivity to a sound contrast tended to improve for spectral features (a consonant contrast). Word learning was significantly linked to perceptual skills in controls but not in gamers. This suggests that game-based language learning diminishes the effect of perceptual skills on word learning ability and that digital applications may thus support word learning particularly in children whose perceptual skills are not so good. In line with some previous studies, our results suggest that active speech production (Icht and Mama, 2015; Junttila and Ylinen, 2020) and game-based learning approach (Junttila et al., 2020) are beneficial for foreign-language learning in children. Together with the previous results, the current findings show that game-based learning including overt speech production may benefit both speech-sound and word learning.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SY, AS, MH designed research. SY, AS, RK, MK designed the pedagogic content of the game. RK, MK, AS, SY developed the speech technology features of the game. SK, AS, SY provided stimuli. SY, AS arranged data collection. SY analyzed data. JL conducted statistical analysis. SY wrote the original draft of the manuscript. All authors modified the manuscript.

FUNDING

The work was supported by Business Finland (project 5628/31/2018 to SY; 1634/31/2018 to MK). Helsinki University Library is paying open access publication fee.

ACKNOWLEDGMENTS

We wish to thank our participants, their caregivers, and their schools, principals and teachers for collaboration, City of Helsinki Education Division for access to schools, Vertti Viitanen for programming the game and the tests, and Elena China-Kolehmainen, Aino Hiltunen, Saana Hyttinen, Liisa Koivusalo, and Liisa Koponen for their assistance in data collection.

REFERENCES

Baese-Berk, M. M. (2019). Interactions between Speech Perception and Production during Learning of Novel Phonemic Categories. *Atten. Percept. Psychophys.* 81, 981–1005. doi:10.3758/s13414-019-01725-4

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., and Bavelier, D. (2018). Meta-analysis of Action Video Game Impact on Perceptual, Attentional, and Cognitive Skills. *Psychol. Bull.* 144, 77–110. doi:10.1037/bul0000130

Best, C. T. (1994). "The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model," in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Editors J. C. Goodman and H. C. Nusbaum (Cambridge, MA: The MIT Press), 167–224.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-speaking Adults and Infants. *J. Exp. Psychol. Hum. Perception Perform.* 14, 345–360. doi:10.1037/0096-1523.14.3.345

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. i. (1997). Training Japanese Listeners to Identify English/r/and/l/: IV. Some Effects of Perceptual Learning on Speech Production. *The J. Acoust. Soc. Am.* 101 (4), 2299–2310. doi:10.1121/1.418276

Catford, J. C., and Pisoni, D. B. (1970). Auditory vs. Articulatory Training in Exotic Sounds*. *Mod. Lang. J.* 54 (7), 477–481. doi:10.2307/32176710.1111/j.1540-4781.1970.tb03581.x

Cenoz, J. (2013). The Influence of Bilingualism on Third Language Acquisition: Focus on Multilingualismfluence of Bilingualism on Third Language Acquisition: Focus on Multilingualism. *Lang. Teach.* 46, 71–86. doi:10.1017/S0261444811000218

Education GPS OECD (2021). Finland. Student performance (PISA 2018). Available at: https://gpseducation.oecd.org/CountryProfile?primaryCountry=FIN&topic=PI&treshold=10 (Accessed January 20, 2021).

Eichenbaum, A., Bavelier, D., and Green, C. S. (2014). Video Games: Play that Can Do Serious Good. *Am. J. Play* 7, 50–72. Available at: https://www.journalofplay.org/issues/7/1 (Accessed September, 1 2020).

Flege, J. E. (1995). "Second Language Speech Learning. Theory, Findings, and Problems," in *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Editor W. Strange (Baltimore: York Press), 233–277.

Hickok, G., and Poeppel, D. (2007). The Cortical Organization of Speech Processing. *Nat. Rev. Neurosci.* 8 (5), 393–402. doi:10.1038/nrn2113

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting Education in "Educational" Apps. *Psychol. Sci. Public Interest* 16, 3–34. doi:10.1177/1529100615569721

Icht, M., and Mama, Y. (2015). The Production Effect in Memory: a Prominent Mnemonic in Childrenffect in Memory: a Prominent Mnemonic in Children. *J. Child. Lang.* 42, 1102–1124. doi:10.1017/S0305000914000713

Iivonen, A. (1998). "Intonation in Finnish," in *Intonation Systems. A Survey of Twenty Languages*. Editors D. Hirst and A. Di Cristo (Cambridge: Cambridge University Press), 314–340.

Junttila, K., Smolander, A-R., Karhila, R., Giannakopoulou, A., Uther, M., Kurimo, M., et al. (2020). Gaming Enhances Learning-Induced Plastic Changes in the Brain. Pre-print Available at: https://psyarxiv.com/k8dxp (Accessed November 19, 2020). doi:10.31234/osf.io/k8dxp

Junttila, K., and Ylinen, S. (2020). Intentional Training with Speech Production Supports Children's Learning the Meanings of Foreign Words: A Comparison of Four Learning Tasks. *Front. Psychol.* 11, 1108. doi:10.3389/fpsyg.2020.01108

Karhila, R., Smolander, A., Ylinen, S., and Kurimo, M. (2019). "Transparent Pronunciation Scoring Using Articulatorily Weighted Phoneme Edit Distance," in Proceedings of Interspeech 2019, Editors: G. Kubin, T. Hain, B. Schuller, D. El Zarka, P. Hödl, Graz, Austria, 15–19 September 2019 (International Speech Communication Association (ISCA)), 1866–1870. Available at: https://arxiv.org/abs/1905.02639 (Accessed September 27, 2019).

Karhila, R., Ylinen, S. P., Enarvi, S., Palomäki, K., Nikulin, A., Rantula, O., et al. (2017). "SIAK – A Game for Foreign Language Pronunciation Learning," in Proceedings of INTERSPEECH 2017, Editors: F. Lacerda, D. House, M. Heldner, J. Gustafson, Strömbergsson, M. Włodarczak, Stockholm, Sweden, 20–24 August 2017, (International Speech Communication Association (ISCA)), 3429–3430. Available at: https://www.isca-speech.org/archive/Interspeech_2017/pdfs/2046.PDF (Accessed September 27, 2019).

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The Effect of Phonetic Production Training with Visual Feedback on the Perception and Production of Foreign Speech Sounds. *J. Acoust. Soc. America* 138 (2), 817–832. doi:10.1121/1.4926561

Kawahara, H., and Irino, T. (2005). "Underlying Principles of a High-Quality Speech Manipulation System STRAIGHT and its Application to Speech Segregation," in *Speech Separation by Humans and Machines*. Editor P. Diveniy Boston, MA: Springer, 167–180.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). "Tandem-STRAIGHT: A Temporally Stable Power Spectral Representation for Periodic Signals and Applications to Interference-free Spectrum, F0, and Aperiodicity Estimation," in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, United States, 31 March–4 April 2008, (New York: IEEE), 3933–3936.

Kawahara, H. (2006). STRAIGHT, Exploitation of the Other Aspect of VOCODER: Perceptually Isomorphic Decomposition of Speech Sounds. *Acoust. Sci. Tech.* 27, 349–353. doi:10.1250/ast.27.349

Kawahara, H., Takahashi, T., Morise, M., and Banno, H. (2009). "Development of Exploratory Research Tools Based on TANDEM-STRAIGHT," in Proceedings: APSIPA ASC 2009: Asia-pacific signal and information processing association, 2009 annual summit and conference, Sapporo, Japan, 4–7 October 2009, (The Asia-Pacific Signal and Information Processing Association (APSIPA)), 111–120. Available at: http://hdl.handle.net/2115/39651.

Kuhl, P. K. (2004). Early Language Acquisition: Cracking the Speech Code. *Nat. Rev. Neurosci.* 5, 831–843. doi:10.1038/nrn1533

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese Listeners to Identify English/r/and/l/. II: The Role of Phonetic Environment and Talker Variability in Learning New Perceptual Categories. *J. Acoust. Soc. America* 94 (3 Pt 1), 1242–1255. doi:10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., and Yamada, T. (1994). Training Japanese Listeners to Identify English/r/and/l/. III. Long-term Retention of New Phonetic Categories. *J. Acoust. Soc. America* 96 (4), 2076–2087. doi:10.1121/1.410149

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese Listeners to Identify English/r/and/l/: A First Report. *J. Acoust. Soc. America* 89 (2), 874–886. doi:10.1121/1.1894649

Lovato, S. B., and Waxman, S. R. (2016). Young Children Learning from Touch Screens: Taking a Wider View. *Front. Psychol.* 7, 1078. doi:10.3389/fpsyg.2016.01078

Mines, M. A., Hanson, B. F., and Shoup, J. E. (1978). Frequency of Occurrence of Phonemes in Conversational English. *Lang. Speech* 21, 221–241. doi:10.1177/002383097802100302

Moisala, M., Salmela, V., Hietajärvi, L., Carlson, S., Vuontela, V., Lonka, K., et al. (2017). Gaming Is Related to Enhanced Working Memory Performance and Task-Related Cortical Activity. *Brain Res.* 1655, 204–215. doi:10.1016/j.brainres.2016.10.027

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The Mismatch Negativity (MMN) in Basic Research of Central Auditory Processing: a Review. *Clin. Neurophysiol.* 118, 2544–2590. doi:10.1016/j.clinph.2007.04.026

Papadakis, S. (2020). Tools for Evaluating Educational Apps for Young Children: a Systematic Review of the Literature. *Interactive Technol. Smart Educ.* Vol. ahead-of-print No. ahead-of-print. doi:10.1108/ITSE-08-2020-0127

Papadakis, S., Vaiopoulou, J., Kalogiannakis, M., and Stamovlasis, D. (2020). Developing and Exploring an Evaluation Tool for Educational Apps (E.T.E.A.) Targeting Kindergarten Children. *Sustainability* 12, 4201. doi:10.3390/su12104201

Russo-Johnson, C., Troseth, G., Duncan, C., and Mesghina, A. (2017). All Tapped Out: Touchscreen Interactivity and Young Children's Word Learning. *Front. Psychol.* 8, 578. doi:10.3389/fpsyg.2017.00578

Rvachew, S. (1994). Speech Perception Training Can Facilitate Sound Production Learning. *J. Speech Lang. Hear. Res.* 37 (2), 347–357. doi:10.1044/jshr.3702.347

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., and Tare, M. (2015). Non-native Phonemic Discrimination, Phonological Short Term

Memory, and Word Learning. *J. Phonetics* 50, 99–119. doi:10.1016/j.wocn.2015.03.001

Simmonds, A. J. (2015). A Hypothesis on Improving Foreign Accents by Optimizing Variability in Vocal Learning Brain Circuits. *Front. Hum. Neurosci.* 9, 606. doi:10.3389/fnhum.2015.00606

STROBE Statement (2021). Checklist of items that should be included in reports of cross-sectional studies. Availabe at: https://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_v4_cross-sectional.pdf (Accessed January 13, 2021).

Thomas, J. (1988). The Role Played by Metalinguistic Awareness in Second and Third Language Learning. *J. Multilingual Multicultural Dev.* 9, 235–246. doi:10.1080/01434632.1988.9994334

Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). Central Auditory System Plasticity: Generalization to Novel Stimuli Following Listening Training. *J. Acoust. Soc. America* 102, 3762–3773. doi:10.1121/1.420139

Treutwein, B. (1995). Adaptive Psychophysical Procedures. *Vis. Res.* 35 (17), 2503–2522. doi:10.1016/0042-6989(95)00016-x

Ylinen, S., Junttila, K., Laasonen, M., Iverson, P., Ahonen, L., and Kujala, T. (2019). Diminished Brain Responses to Second-Language Words Are Linked with Native-Language Literacy Skills in Dyslexia. *Neuropsychologia* 122, 105–115. doi:10.1016/j.neuropsychologia.2018.11.005

Ylinen, S., and Kurimo, M. (2017). "Kielenoppiminen Vauhtiin Puheteknologian Avulla," in *Oppimisen Tulevaisuus*. Editors H. Savolainen, R. Vilkko, and L. Vähäkylä, (Helsinki, Finland: Gaudeamus), 53–63.

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., et al. (2015). Two Distinct Auditory-Motor Circuits for Monitoring Speech Production as Revealed by Content-specific Suppression of Auditory Cortex. *Cereb. Cortex* 25 (6), 1576–1586. doi:10.1093/cercor/bht351

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.