

## Gaming enhances learning-induced plastic changes in the brain

Katja Junttila<sup>a,\*</sup>, Anna-Riikka Smolander<sup>a,b</sup>, Reima Karhila<sup>c</sup>, Anastasia Giannakopoulou<sup>d</sup>, Maria Uther<sup>e</sup>, Mikko Kurimo<sup>c</sup>, Sari Ylinen<sup>a,b,f</sup>

<sup>a</sup> Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, POB 21, FI-00014, University of Helsinki, Finland

<sup>b</sup> Logopedics, Welfare Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland

<sup>c</sup> Aalto University, Department of Signal Processing and Acoustics, P.O. Box 12200, FI-00076 Aalto, Finland

<sup>d</sup> School of Psychology, University of Bedfordshire, Park Square, Luton LU1 3JU, United Kingdom

<sup>e</sup> Centre for Psychological Research, Department of Psychology, Faculty of Education, Health and Wellbeing, University of Wolverhampton, M305, Millenium City Building, Wolverhampton WV1 1LY, United Kingdom

<sup>f</sup> CICERO Learning, Faculty of Educational Sciences, P.O. Box 9, 00014, University of Helsinki, Finland

### ARTICLE INFO

#### Keywords:

Digital game-based learning  
Gaming  
Foreign-language learning  
Automatic speech recognition

### ABSTRACT

Digital games may benefit children's learning, yet the factors that induce gaming benefits to cognition are not well known. In this study, we investigated the effectiveness of digital game-based learning in children by comparing the learning of foreign speech sounds and words in a digital game or a non-game digital application. To evaluate gaming-induced plastic changes in the brain, we used the mismatch negativity (MMN) brain response that reflects the access to long-term memory representations. We recorded auditory brain responses from 37 school-aged Finnish-speaking children before and after playing a computer-based language-learning game. The MMN amplitude increased between the pre- and post-measurement for the game condition but not for the non-game condition, suggesting that the gaming intervention enhanced learning more than the non-game intervention. The results indicate that digital games can be beneficial for children's speech-sound learning and that gaming elements *per se*, not just practice time, support learning.

### 1. Introduction

Technology can benefit learning in many ways, and digital language learning (DLL) opens new perspectives for language education (Li & Lan, 2021a). A number of studies also suggest that digital game-based learning can be an effective way to learn foreign languages (e.g. Aghlari & Tamjid, 2011; Chen et al., 2018; Franciosi, 2017; Grgurović et al., 2013; Hsu, 2017; Liu & Chu, 2010; Sandberg et al., 2014; Tsai & Tsai, 2018; see reviews: Acquah & Katz, 2020; Zou et al., 2019), even though playing a game may sometimes increase cognitive load and hinder learning from the game (deHaan et al., 2010). As compared to traditional teaching and other non-game learning approaches, the results on the effectiveness of digital game-based learning are somewhat mixed (Li & Lan, 2021a). However, learners often find learning by playing both enjoyable and motivating (Kober et al., 2020; Li & Lan, 2021b; Peterson, 2010a, 2010b; Yip & Kwan, 2006; Young & Wang, 2014) and prefer digital-game based learning over regular activity-based classroom learning (Yip & Kwan, 2006).

Although many studies demonstrate benefits of game-based learning, rigorous studies comparing the effectiveness of different game features or studies comparing learning with a game and learning the same content with other types of media are rarely conducted, yet they are much needed to better understand the best practices to support learning (Bavelier & Davidson, 2013; Mayer, 2015). Although positive effects of digital game-based language-learning have been shown (e.g. Aghlari & Tamjid, 2011; Hong et al., 2017; Liu & Chu, 2010), it is not yet understood whether the game elements themselves enhance learning or whether these effects are due to motivation increasing the practicing time (Sandberg et al., 2011; Tejedor-García et al., 2020). To this end, the aim of the current study was to determine the effectiveness of digital game-based language-learning approach on children's foreign-language learning, while controlling for the gaming time. This approach was expected to reveal the effect of gaming elements, such as visual game features, rewards, and freedom of choice, on learning.

To study gaming effects, we used a digital language-learning game called "Say it again, kid!" that is targeted for children who are beginning

\* Corresponding author at: Cognitive Brain Research Unit, Faculty of Medicine, POB 21, FI-00014 University of Helsinki, Finland.

E-mail addresses: [katja.junttila@helsinki.fi](mailto:katja.junttila@helsinki.fi) (K. Junttila), [anna.2.smolander@tuni.fi](mailto:anna.2.smolander@tuni.fi) (A.-R. Smolander), [Anastasia.Giannakopoulou@beds.ac.uk](mailto:Anastasia.Giannakopoulou@beds.ac.uk) (A. Giannakopoulou), [M.Uther@wlv.ac.uk](mailto:M.Uther@wlv.ac.uk) (M. Uther), [mikko.kurimo@aalto.fi](mailto:mikko.kurimo@aalto.fi) (M. Kurimo), [sari.p.ylinen@tuni.fi](mailto:sari.p.ylinen@tuni.fi) (S. Ylinen).

<https://doi.org/10.1016/j.bandl.2022.105124>

Received 5 May 2021; Received in revised form 5 April 2022; Accepted 8 April 2022

Available online 26 April 2022

0093-934X/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

to learn English (Karhila et al., 2017; Ylinen & Kurimo, 2017; see also Ylinen et al., 2021). Children are often interested in games and playing is a natural way for them to learn (Eskenazi, 2009). From the exposure to spoken English provided by the game, players are expected to acquire the English phonological system as well as to learn English words and their pronunciation. In particular, the current study addressed foreign speech sounds embedded in novel words. Learning to distinguish foreign-language speech sounds that do not exist in the learner's native language can be very difficult and depend on how similar they are to the native language speech sounds (Best & Tyler, 2007; Flege & Bohn, 2021). Even after years of training, perception of some foreign language contrasts can remain poor (Flege, 1988; Levy & Strange, 2008). Studies examining how training affects learning to discriminate speech sound contrasts in foreign language have shown contradictory results. Some studies have found that training has improved the ability to perceive foreign-language contrasts (Bradlow et al., 1999; Saloranta et al., 2020; Tamminen et al., 2015; Ylinen et al., 2010), while other studies have not found clear training effects (Hisagi et al., 2016; Peltola et al., 2007). A possible explanation could lie in the type of language training. Studies where training consists of general language immersion education or classroom teaching do not always show benefit for foreign sound-contrast discrimination (e.g. Hisagi et al., 2016; Peltola et al., 2007). Improvement in phonetic discrimination has, however, been found in studies using more targeted phonetic training (Bradlow et al., 1999; Saloranta et al., 2020; Tamminen et al., 2015; Ylinen et al., 2010).

Drawing from these previous studies, the current study addressed targeted training of foreign speech sounds, yet the main aim was to examine whether our gaming approach enhances learning more than the same amount of non-game training. Specifically, we compared the learning of non-native words with novel phonemes in two different learning conditions: a game condition and a non-game condition. By presenting equal number of repetitions of words to be learned with the game training condition and a non-game training condition, we wished to ensure that potentially better learning results could not be caused by more practice and a larger number of repetitions of the sounds or words to be learned.

Psychometric experiments requiring behavioural response are often challenging to conduct in children because they have short attention span and fluctuations in their willingness to co-operate (Ylinen et al., 2021). However, these shortcomings may be overcome by using brain research methods, such as the measurement of auditory event-related potentials (ERPs) with electroencephalography (EEG). In particular, the mismatch negativity (MMN) component of ERPs is an excellent tool to investigate auditory perception and perceptual learning especially in children, since MMN elicitation does not require direct attention towards the stimuli (Fitzgerald & Todd, 2020; Näätänen et al., 1978) or overt response (for a review, see Cheour et al., 2000). Therefore, we used MMN to investigate the effectiveness of game-based learning in children.

The MMN response originates primarily from auditory cortices located at the upper part of the temporal lobes and it is measured at scalp as having negative polarity (i.e., a larger or stronger response is more negative) due to its orientation in the cortex (Näätänen et al., 2007). It was first found to be elicited to auditory stimuli that deviated from the other sounds presented in a same sequence (Näätänen et al., 1978). This is typically termed an 'oddball' paradigm, where a stimulus sequence consists of a high-probability 'standard' sound that is repeated and occasionally replaced by a low-probability 'deviant' sound. The MMN is typically elicited 100–250 ms from the onset of acoustic deviance in adults. In children, the timing of MMN response can be later than in adults but the latency decreases with age (Shafer et al., 2000), ranging between 100 and 300 ms from the onset of acoustic deviance. The MMN does not reflect mere acoustic deviance detection, however: Näätänen and colleagues (1997) have shown that the MMN is enhanced by the activation of long-term memory traces for native-language speech sounds. Pulvermüller and colleagues (2001), in turn, have demonstrated

that the MMN reflects long-term memory traces for individual spoken words, as is shown by an increase of its amplitude. Since the MMN is an automatic response, this indicates that the memory representations for speech sounds and words are activated automatically even without focused attention.

The increased amplitude of the MMN response may also reflect learning effects (Winkler et al., 1999; for a review, see Näätänen et al., 2007). For example, a study by Ylinen and colleagues (2010) has shown that training can induce changes in the processing of foreign speech sounds due to the establishment of long-term memory representations for them, enhancing the ability to automatically distinguish foreign speech sounds that resemble each other. In particular, while native Finnish speakers' MMN responses to an English vowel contrast were significantly smaller than those of native English speakers' MMNs at pre-test, after training no such difference was found. The Finnish learners' MMNs were also significantly larger after training than before it. Importantly, the MMN increase was accompanied by improved behavioural categorisation of the English target vowels in the Finnish learners. Together, behavioural and brain indices suggested robust and consistent learning effects.

Based on previous game-based learning effects (see Baker et al., 2015; Hong et al., 2017; Yip & Kwan, 2006) and MMN training effects (Ylinen et al., 2010), we hypothesised that learning efficacy may be higher in the gaming condition than in the non-game condition, implying that gaming enhances plastic changes in the brain. Correspondingly, children's MMN responses are expected to show a larger increase with the game condition than with the non-game condition. Since we compared learning in a game and in a non-game condition and controlled for the gaming time, we expected this to reflect plastic changes in the brain that are due to the gaming elements *per se*. These elements include visual game features, rewards and feedback, and autonomy (freedom of choice). Thus, out of the key features through which digital learning games affect learning outcomes (Acquah & Katz, 2020) we kept, for example, ease-of-use and challenge constant, but varied rewards and feedback as well as control or autonomy. We hypothesized that the rewards provided by the game would support learning via the activation of fronto-striatal reward networks in the brain (for a review, see Nahum & Bavelier, 2020) that play an important role in language learning (Ripollés et al., 2014) or via the activation of frontal attentional networks (Kober et al., 2020; Nahum & Bavelier, 2020). In addition, we aimed to investigate the contribution of MMN sources at right and left temporal cortex to the learning of foreign speech sounds and words. Although the left hemisphere is typically associated with language-related processing (Näätänen et al., 1997), the right hemisphere has been shown to contribute more to foreign-language processing in children (Nora et al., 2017; Ylinen et al., 2019). Therefore, we expected that right temporal cortex might show more plastic changes than the left temporal cortex.

## 2. Materials and methods

### 2.1. Ethics

The study was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences. Participation was voluntary. Each participants' guardian signed a written informed consent and the participants gave their oral informed consent before participating. The participants were compensated with one cinema ticket per hour of participating. The study was conducted in accordance with the Declaration of Helsinki.

### 2.2. Participants

Participants were 7 to 11 year old children ( $M = 9.53$  years,  $SD = 0.77$  years,  $N = 37$ , 19 girls). To be included in the study, the participants had to meet the following criteria: they had to be between 7 and

11 years of age and monolingual native Finnish speakers with no developmental disorders, no language disorders, no learning disorders, no head injuries, and with normal hearing and normal vision or vision corrected to normal with eyeglasses. In addition, the participants were screened with behavioural pre-tests and had to meet the following criteria to be included in the study: a score of not below 1.33 standard deviations in each of the Wechsler Intelligence Scale for Children IV (WISC-IV; Wechsler, 2010) subtests included in the study, i.e. they needed to score a minimum of 6 standard points (age corrected score,  $M = 10$ ,  $SD = 3$ ) in Block Design, Digit Span, Coding, and Vocabulary. Out of the 42 children who volunteered, 37 fulfilled the inclusion criteria and were included in the study. These children were divided into two groups that received training and stimulation in a counterbalanced fashion. Specifically, 18 of the participants practiced the word containing the phoneme /ð/ with the game levels and the phoneme /θ/ with the non-game levels, whereas the other 19 participants practiced the word containing the phoneme /θ/ with the game levels and the phoneme /ð/ with the non-game levels (the stimuli are further described in section 2.3.2 EEG stimuli).

All participants attended the Finnish comprehensive school and 28 of them studied English at school. According to their guardians' reports, two children did not know any English, 18 knew a few English words, 17 could understand some simple sentences in English, and none of the children could understand and speak English fluently.

## 2.3. Experimental design

### 2.3.1. Behavioural tests

To ascertain whether the children knew the target words before playing the game, or whether they learned them during the gaming, each participant's knowledge of the words used in the EEG experiment was tested before the first EEG session and again after training. The words and pseudowords used as stimuli for the EEG experiment were presented one at a time through headphones. After each word the child was asked whether the word sounded familiar to them. If the child thought the word was familiar, they were also asked what the word means.

Participants' cognitive, phonological, and literacy skills were assessed with standardised psychological tests (the phonological and literacy skills will be reported elsewhere) to determine that they fulfilled the inclusion criteria. During the skill testing, the participant sat in a quiet room with the researcher. Wechsler Intelligence Scale for Children IV (WISC-IV; Wechsler, 2010) and in particular its subtests Block Design, forward and backward Digit Span, Coding, and Vocabulary were used to assess perceptual reasoning skills, auditory short-term memory, eye hand coordination and processing speed, and vocabulary, respectively. The testing took one hour.

### 2.3.2. EEG stimuli

For stimulus recordings, we chose English words with dental fricative phonemes (voiceless /θ/ and voiced /ð/) that do not belong to Finnish phonology (Sajavaara & Dufva, 2001) and were thus expected to be novel and relatively difficult for native Finnish speakers (Lintunen, 2005, 2013). Language learners are likely to face difficulties and make errors both in production and perception of phonemes that are not part of the phonemic system of their native language (Flege, 1988; Lintunen, 2005). We chose to use content words and, therefore, we chose words where the critical fricative is placed within the word (word-initial /ð/ is often used in function words). The recordings were conducted in a sound-attenuated recording studio where native British English speakers pronounced a list of English words and minimal-pair pseudowords several times. The most representative exemplars were chosen for the EEG experiment. The final stimuli presented during the EEG recording were spoken English words *healthy* ['helθi] and *feather* ['feðə] and their pseudoword counterparts *healty*\* ['helti]\* and *feder*\* ['fedə]\* (Fig. 1). The stimulus words were chosen so that they would have the target fricatives in the second syllable close to their recognition point, which enables the contribution of word recognition to the processing. In addition, we wished to ensure that most children in the target group of the study would not know the meanings of these English words in advance, whereas concepts should be familiar to them (i.e., the children understood the words in Finnish). Commonly used English textbooks for the age group of the study were checked and the chosen words were not included in their syllabus.

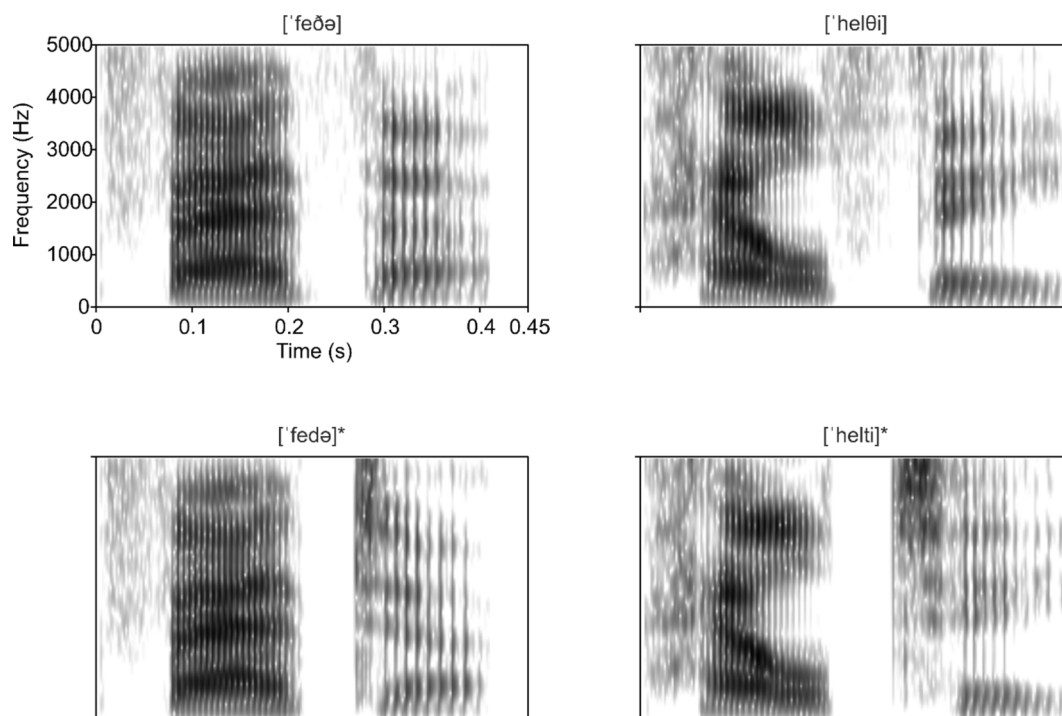


Fig. 1. Spectrograms of the stimuli.

The stimuli were modified using Praat software (Boersma & Weenink, 2010). The stimuli were cross-spliced at zero-crossings, and the first 200 ms of the word *healthy* and the first 220 ms of the word *feather* were used as the beginnings of the pseudowords *healty\** and *feder\**, respectively (Figs. 1 & 3). The pitch contour of *healthy* was adopted to *healty\** and the pitch contour of *feather* was adopted to *feder\**. Thus, the beginnings of *healthy* and *healty\**, as well as *feather* and *feder\** were identical, and therefore it was not possible to differentiate the pseudoword from the word before the last two phonemes of the words.

### 2.3.3. EEG recordings and data processing

All 37 children participated in at least two EEG measurements, which were conducted before and after game-based training with the language learning game. On average, the post-test measurements were carried on five weeks after the pre-tests. In addition, to control for the effects of repeated EEG measurements and out-of-game exposure to English, 22 of the children participated in an additional EEG measurement that took place on average five weeks (an equal amount of time as between the two last measurements) prior to the before-training measurement (this extra measurement is referred as a baseline measurement) (Fig. 2 A).

EEG was recorded from 64 channels and additional electrodes on the mastoids using BioSemi ActiveTwo system (BioSemi Inc., Amsterdam, The Netherlands) with Biosemi ActiView Version 7.07 EEG acquisition software. Horizontal eye movements were monitored with two electrodes attached near the outer canthi of the eyes. Vertical eye movements were monitored with an electrode placed under the left eye. One additional electrode was attached on the tip of the nose for online reference. During the measurement, the participants sat on a comfortable chair in a sound-attenuated and electrically attenuated room and watched a muted film of their choice with subtitles on. They were instructed to concentrate on the film and to ignore the auditory stimuli. The stimulus sequences were presented binaurally via headphones at a comfortable hearing level of 60 dB.

The EEG experiment included four different stimulus sequences. Two oddball stimulus sequences contained deviant and standard stimuli. In one oddball sequence, a repeating standard pseudoword *feder\** ( $p = 0.8$ ) was occasionally replaced by a deviant word *feather* ( $p = 0.2$ ). In the other oddball sequence, the standard pseudoword was *healty\** ( $p = 0.8$ ) and the deviant word was *healthy* ( $p = 0.2$ ). Each deviant stimulus was presented 120 times and each standard stimulus 480 times in the deviant sequences. In addition, to obtain ERP responses to oddball deviant stimuli *feather* and *healthy* in the position of repeating standards, these words were presented 180 times each in the two separate “standard-only” sequences ( $p = 1$  for each). Inter-stimulus interval (ISI; offset to onset) was 500 ms in both the oddball and the standard-only sequences. The order of the sequences was randomised for each participant and for each measurement session. All stimulus sequences were presented by using Presentation® software (Version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, <https://www.neurobs.com>). The EEG measurement took two hours, including preparation.

After the recording, the EEG data were processed with BESA Research 7.0 software (BESA GmbH, Gräfelfing, Germany). The data were re-referenced to the average of mastoids to improve the signal-to-noise ratio, and bad channels were interpolated. Then the data were filtered with a band pass of 1.5–20 Hz (slope 24 dB/octave; a strict high-pass cut off was used to reduce slow drifts in children’s EEG signal beyond the MMN frequency range, see Kalyakin et al., 2007; Nenonen et al., 2005; Picton et al., 2000; Sinkkonen & Tervaniemi, 2000). Eye movements were corrected using BESA Research Artifact Correction that is based on the spatial filtering method. Data were divided into epochs using a –100 to 800 ms time window (from stimulus onset), epochs with artifacts exceeding  $\pm 75 \mu\text{V}$  as well as 5 first epochs of each sequence were rejected, and the remaining epochs were averaged for each stimulus type separately. To illustrate the MMN, deviant-minus-standard difference waveforms were created by subtracting responses to identical stimuli in deviant and standard positions. Specifically, responses to

stimulus *healthy* presented in the standard-only sequence were subtracted from those to the deviant stimulus *healthy* of the oddball sequences, and responses to stimulus *feather* presented in the standard-only sequence were subtracted from those to the deviant stimulus *feather* of the oddball sequences.

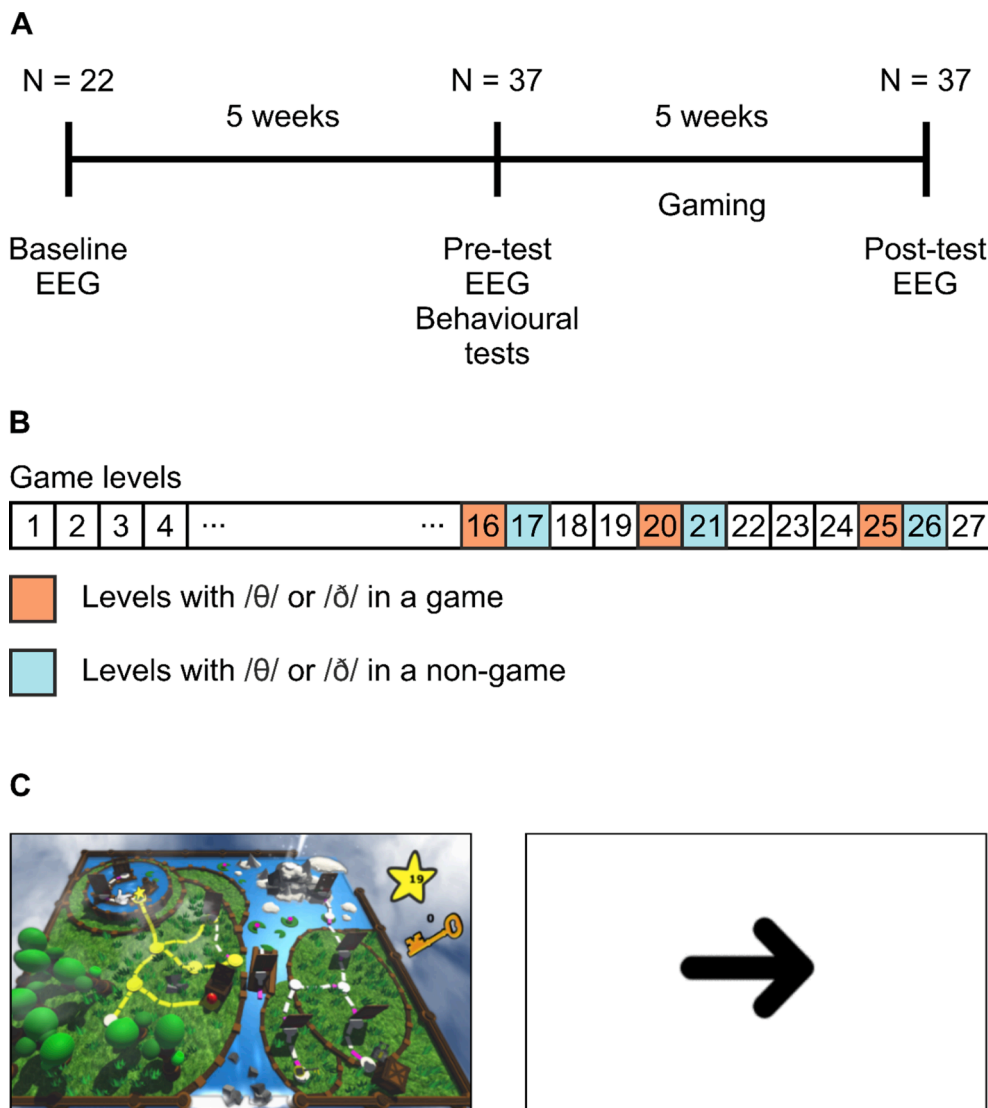
To examine the effects of game-based learning, we formed grand-average waveforms from deviant-minus-standard difference waveforms for game condition and from the deviant-minus-standard difference waveforms for the non-game condition. Note that to eliminate possible differences in difficulty between the two target phonemes, namely, the voiceless and voiced fricatives in *healthy* and *feather*, the participants were divided into two groups and stimulus words *healthy* and *feather* were assigned to game and non-game conditions in a counterbalanced manner in these groups (i.e., for one group, the game training and EEG game condition included *healthy* and the non-game training and EEG non-game condition included *feather*, whereas for the other group, the game training and EEG game condition included *feather* and the non-game training and EEG non-game condition included *healthy*). Consequently, the ERP waveforms for each condition (game vs. non-game) were constructed by combining responses to both kinds of target words (*healthy* and *feather*) for each condition across the two groups of participants (i.e., the grand-average ERPs include all 37 participants)<sup>1</sup>.

The MMN responses were quantified from deviant-minus-standard difference waveforms (created by subtracting responses to identical stimuli in deviant and standard positions). Fronto-central electrode sites F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4 were selected for analysis, because MMN responses are typically most prominent on fronto-central sites when referenced to the average of mastoids. To decrease the likelihood of false positive results, we created a grand-average waveform over all deviant-minus-standard responses of the nine pre-selected electrode sites to determine the peak latency of the MMN response (Luck & Gaspelin, 2017). A  $\pm 20$  ms time window centred at this average latency of 334 ms was then used to quantify the mean MMN amplitude from individual deviant-minus-standard difference waveforms.

### 2.3.4. Source localisation

To localise the sources of the MMN responses, the EEG data were pre-processed for source imaging with BESA Research 7.0 software (BESA GmbH, Gräfelfing, Germany) as described by Michel and Brunet (2019). The data were filtered with a band pass of 0.1–40 Hz (slope 24 dB/octave) and then divided into epochs using a –100 to 800 ms time window relative to the onset of the stimuli. Epochs exceeding  $\pm 100 \mu\text{V}$  during a 200 ms time window centred at the MMN latency (234–434 ms) were rejected. The remaining epochs were averaged separately for each stimulus type. Average reference was used for making deviant-minus-standard waveforms. Source localisation was conducted only for the participants who had a minimum of 80 epochs remaining for game stimuli, non-game stimuli, or both. For each remaining participant, the latencies of the most negative peaks in 284–384 ms time window were determined. A head model designed for 10-year-olds from BESA Research 7.0 was used for source modelling. Two dipoles were fitted over left and right auditory cortices in 40 ms time windows centred at these individual peaks of the game and the non-game condition. To be included in further analysis, the goodness of fit of the two-dipole model had to be at least 75% and at least one of the sources had to be in the auditory cortex or its proximity. For the game condition, 17 participants, and for the non-game condition, seven participants fulfilled these criteria. Since few participants showed an MMN response for the non-game condition, its

<sup>1</sup> Since we averaged the responses to the two items with slightly different points (20 ms) of divergence, the resulting MMN responses may be slightly wider and flatter than the responses to stimuli with identical timing would have been. This does not, however, affect the comparison of game and non-game conditions, because the effect is the same for both conditions.



**Fig. 2.** Timeline of EEG measurements, tests, and gaming (A). Schematic of the order of the game levels with the embedded game and non-game levels with target phonemes (B). Screen shots of game and non-game levels (C).

sources are not further examined. For the game condition, 14 participants showed sources in the left auditory cortex and 15 participants in the right auditory cortex. The source strengths were measured as absolute values in 40 ms time windows centred at the strongest individual peaks found in the MMN time range of 284–384 ms.

#### 2.4. Training

Between the pre- and post-test EEG sessions, the children practiced some frequent English words (e.g. child, drink, high) with the “Say it again, kid” (SIAK) language learning game (Karhila et al., 2017; Ylinen & Kurimo, 2017; see also Ylinen et al., 2021, for a newer mobile version of the game) on either Windows laptops or Android tablets using a headset microphone. Proceeding in the game requires listening to spoken English words and imitating them as accurately as possible. SIAK uses automatic speech recognition (ASR) technology optimised for children’s voices to evaluate children’s utterances (Karhila et al., 2019, 2017). The game is designed as a computer board game and it has 27 levels. Each level is presented as a game board which the players can explore by moving their avatar around it. Game boards contain cards, which pop up when the players move their avatar on them. Each card has a picture, and when it pops up, the children hear the corresponding

word first in Finnish and after that in English. When a microphone symbol is shown, the children are supposed to imitate the English word as accurately as they can. Players’ utterances are recorded and sent to a game server where the ASR assesses each speech sound of the utterance. Then the players hear their own utterances played back to them, which is followed by the original English speaker’s utterance for comparison. Finally, one to five stars are awarded to the player based on the ASR assessment. Although moving on the board is at first limited, getting more stars opens new paths to explore. These rewards are expected to encourage children to produce speech and try their best. In the end of each level, the players have the chance to test their learning with a test card where players try to produce some word learned at the level without a model. New levels of the game are unlocked when at least three stars are scored from the test card word. If the players do not score enough stars, they can go back to the learning card to practise the word and then try the test card again.

The gaming period lasted on average 4.3 weeks. The children played the SIAK game on average 15.5 min a day 2.9 days a week. They played through all 27 levels embedded in the game. Each level contained about 15 words to learn and typically focused on some topic such as *clothes*, *animals*, or *food*, yet some levels trained some phoneme that was expected to be difficult to Finnish speakers. Among the 27 levels, 21 game

levels included words not containing the target phonemes /ð/ and /θ/, whereas 3 game levels contained words with /θ/ for one half of the group (e.g., *healthy, mouth, three*) and words with /ð/ for the other half of the group (e.g., *feather, mother, this*) (Fig. 2 B). In addition, we embedded into the game 3 non-game levels that contained words with /ð/ for one half of the group and words with /θ/ for the other half of the group (Fig. 2 B). The non-game levels were stripped of all the game-like elements, that is, the colourful game board, freedom to explore, and gaining stars or any kind of feedback. Instead, the non-game condition had a white screen with a black arrow and had a forced presentation order of the target words (Fig. 2 C). Similar to the game condition, the task was to imitate the heard English words. Thus, the only difference between the game and non-game training was the presence or lack of gaming elements, which included the visual presentation of the game, freedom to explore the game and make choices, and rewards (i.e., stars provided by ASR for feedback).

Except for the non-game, the players were free to explore the game condition levels as they wished. Thus, different players could have different amount of practice. However, we had to ensure that the amount of practice would be the same between the critical game and non-game conditions. To match the extent of exposure to the target words and speech sounds in the game and non-game levels, we recorded the number of times the cards containing words with the target phonemes were opened in the game levels. Then the matching number of target words was presented in the non-game levels. As both conditions were embedded in the SIAK game and the number of trained words was equal between the corresponding game and non-game levels, the active training time in the game and non-game conditions was the same. To allow free exploration of the game levels, they had to be presented prior to the non-game levels, as we did not wish the non-game to restrict playing (therefore, the order of the game and non-game levels could not, unfortunately, be counterbalanced).

The order of game levels was set to proceed from simpler to more difficult contents. For example, the first level included greetings like *Hello!* and the 27th level proceeded to short sentences, such as *I speak Finnish*. The game levels with the target sounds were numbers 16, 20 and 25 (as these sounds were expected to be difficult). The three game levels with one target fricative (either voiced or voiceless) were followed by corresponding non-game levels with the other fricative (levels 17, 21 and 26) (Fig. 2 B). After the presentation of target game and non-game levels, some intervening levels with other sounds were presented before the next game level with the target sounds (e.g., a game level with /ð/, a non-game level with /θ/, two game levels with other sounds, the second game level with /ð/, the second non-game level with /θ/, etc).

## 2.5. Statistical analyses

The differences between children's learning of the words during training with the game and with the non-game (and used in the EEG experiment) were examined with McNemar's test. One test was conducted to compare the familiarity of the words to children and another to compare whether they knew the meanings of the words.

The game and non-game differences over the measurement time points were compared with linear mixed model analyses with IBM SPSS Statistics, Version 25. Two linear mixed model analyses were conducted separately. As the children were likely to hear English, and more specifically, the target phonemes of this study from other sources in addition to the game between the measurements, we compared the change in MMN amplitude between the baseline measurement and the pre-measurement. The effects of out-of-game exposure to English language and hearing the stimulus sequences during the first measurement were analysed for the 22 participants who participated in the baseline MMN measurement. The analysis included Treatment (game – non-game), Anteriority, and Laterality as factors and Time between the baseline measurement and pre-measurement in days as a covariate. Participant and treatment were included as random effects, and scaled identity was

chosen as the covariance structure based on Schwarz's Bayesian Criterion (BIC). To compare the effectiveness of the game and non-game training, a mixed model analysis for all 37 participants' MMN measures pre and post training was conducted. The analysis included the same factors as the previously described one (Treatment, Anteriority, and Laterality), and the Time between the pre and post measurements in days was included as a covariate. Participant and Treatment were included as random effects and unstructured covariance matrix was chosen based on BIC. Only the significant main effects and interactions at alpha level 0.05 including Treatment or Time are reported in the results.

The strength of left and right temporal sources of the MMN response was compared with linear mixed model analysis, where Hemisphere was included as a factor.

## 3. Results

Before participating in the first EEG recording session, one child found both stimulus words familiar-sounding and 18 reported that one of the stimulus words sounded familiar to them. However, none of the children knew the meanings of the stimulus words. After the children had played the game, 31 reported that the word learned with the game sounded familiar and 28 reported that the word learned with the non-game sounded familiar. According to the McNemar's test, no difference in familiarity between the words learned with the game and the non-game was found ( $p = 0.508$ ). After training with the game, 19 children had learned the meanings of the stimulus words correctly. Training with the non-game resulted in 16 children learning the stimulus word meanings correctly. Again, no difference between the training methods was found ( $p = 0.629$ ).

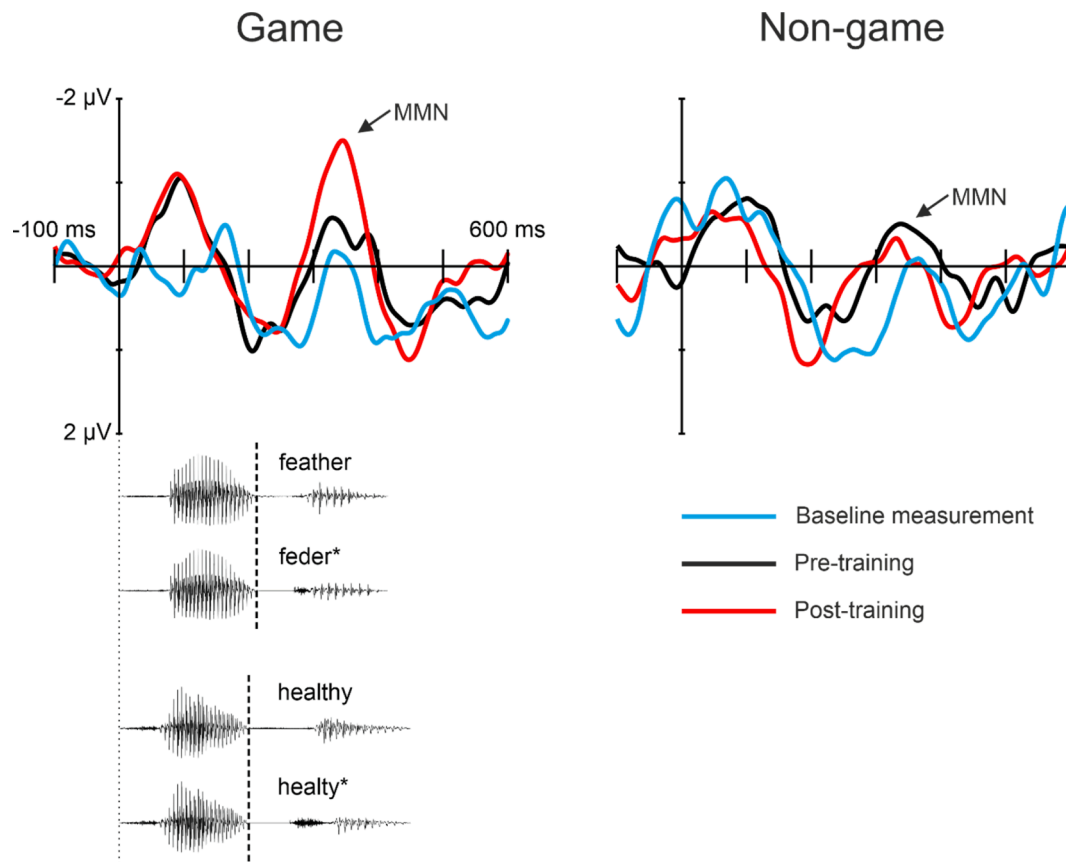
For MMN amplitude, the linear mixed model analysis for game vs. non-game effects revealed a significant interaction of Treatment and Time,  $F(1, 1195.08) = 26.81, p < 0.001$ . To adjust for the different number of days between the pre-measurement and the post-measurement between participants, the estimated marginal means for the game and non-game were evaluated at the day of the pre-measurement and at the mean of days between pre and post measurements (36.62 days), as presented in Table 1. Training with the game condition increased the MMN amplitude more than training with the non-game condition (Figs. 3 & 4, Table 1). Further pairwise comparison revealed a post-measurement difference between the game and the non-game condition's estimated marginal means ( $p = 0.008, 95\% \text{ CI} = 0.19, 1.20$ ). The MMN amplitude was larger (more negative) for the game condition ( $-0.99 \mu\text{V}$ ) than for the non-game condition ( $-0.30 \mu\text{V}$ ). However, none of the main effects or interactions were significant in the linear mixed model analysis for the effects of out-of-game exposure to English language and hearing the same stimuli from the first measurement repeatedly.

The linear mixed model analysis revealed no significant difference between the strengths of the left ( $M = 25.20 \text{ nAm}, SD = 10.38, 95\% \text{ CI} = 16.94, 33.45$ ) and right ( $M = 31.90 \text{ nAm}, SD = 18.36, 95\% \text{ CI} =$

**Table 1**

Estimated marginal means (EMM) of the MMN amplitudes in  $\mu\text{V}$  for the game and non-game conditions in the pre and post measurements. Time point pre-measurement represents the estimated mean at the day of the pre-measurement and the post-measurement represents the estimated mean at 36.62 days, which is the mean of days between the pre-measurement and the post-measurement.

Condition	Time point	EMM	S.E.	95 % CI	
				Lower	Upper
Game	pre-measurement	-0.32	0.22	-0.76	0.13
	post-measurement	-0.99	0.22	-1.44	-0.54
Non-game	pre-measurement	-0.25	0.22	-0.70	0.20
	post-measurement	-0.30	0.22	-0.74	0.15



**Fig. 3.** The deviant-minus-standard difference waveforms with the MMN responses at electrode FCz at baseline measurement, pre-training, and post-training, and the waveforms of the stimulus sounds presented on the same time scale. Deviation points between standard and deviant stimuli are marked with dashed lines. The average of both stimulus word-pseudoword pairs were used in the non-game and game conditions. Note that negativity is plotted up.

23.93, 39.88) temporal sources of the game condition MMN response. The left temporal source was found in 14 participants and the right temporal source in 15 participants. An example of the locations of the dipoles of one participant is presented in Fig. 5.

#### 4. Discussion

The present study aimed to investigate the effectiveness of game-based learning approach on children's foreign language learning by using the MMN component of ERP as an indicator of training-induced plastic changes in the brain (see Ylinen et al., 2010). More specifically, we compared the effects of game-based learning to the effects of non-game-based approach with equal number or repetitions during exposure. The MMN responses increased more for the gaming condition than for the non-game condition.

It is well-established that the MMN is sensitive to acoustic differences. The finding of larger MMN for words learned with the game than with the non-game cannot, however, be explained by differences in saliency of the acoustic stimulus properties, because our main finding is based on an interaction between Time and Treatment, and no differences were observed before training between the responses to the same words. The effect cannot be due to phoneme or word difficulty either, since the words that were learned with the game or with the non-game were counterbalanced between participants and the MMN responses were formed across both words in each condition. Therefore, there was, on average, no difference between the stimuli used for the game and the non-game MMN responses. Thus, stimulus properties cannot account for the pattern of results.

Many previous studies that have investigated the effectiveness of digital game-based learning have compared two participant groups: one

learning with a game and another learning with some other activity or another type of game (Aghlari & Tamjid, 2011; deHaan et al., 2010; Franciosi, 2017; Hsu, 2017; Liu & Chu, 2010; Yip & Kwan, 2006; Young & Wang, 2014). Although this is general practise, it is difficult to fully exclude the possibility that individual differences between the participants' learning abilities result in different learning effects between groups despite similar pre-test performance. This considers especially studies in children's learning because of the large differences in the maturation of children's skills. In the present study, however, the effect of individual learning-ability differences was controlled for: All participants learned a target word with the game and another word with the non-game. Hence, the different results for the game and the non-game do not stem from differences in individuals' learning abilities.

Taken together, since the properties of the stimuli or the children's learning abilities do not account for the larger post-test than pre-test MMN for the game condition, the most plausible explanation for the pattern of results is that the game is more effective for speech-sound learning than the non-game. The results of this study about beneficial effects of gaming on foreign-language learning are in line with previous studies (Aghlari & Tamjid, 2011; Franciosi, 2017; Hsu, 2017; Liu & Chu, 2010; Sandberg et al., 2014; Tsai & Tsai, 2018; Young & Wang, 2014). However, while some previous studies (Franciosi, 2017; Young & Wang, 2014) did not control for practising time that is likely to enhance learning (e.g. Tejedor-García et al., 2020), it is not fully clear whether their gaming benefits are due to the increase of practising time or gaming elements. To this end, the present study sought to investigate the effect of gaming compared to similar amount of non-game practise to determine the effect of the gaming elements. In the present study, the number of times each target word was trained was equal between the game and the non-game conditions. Further, the number of times other

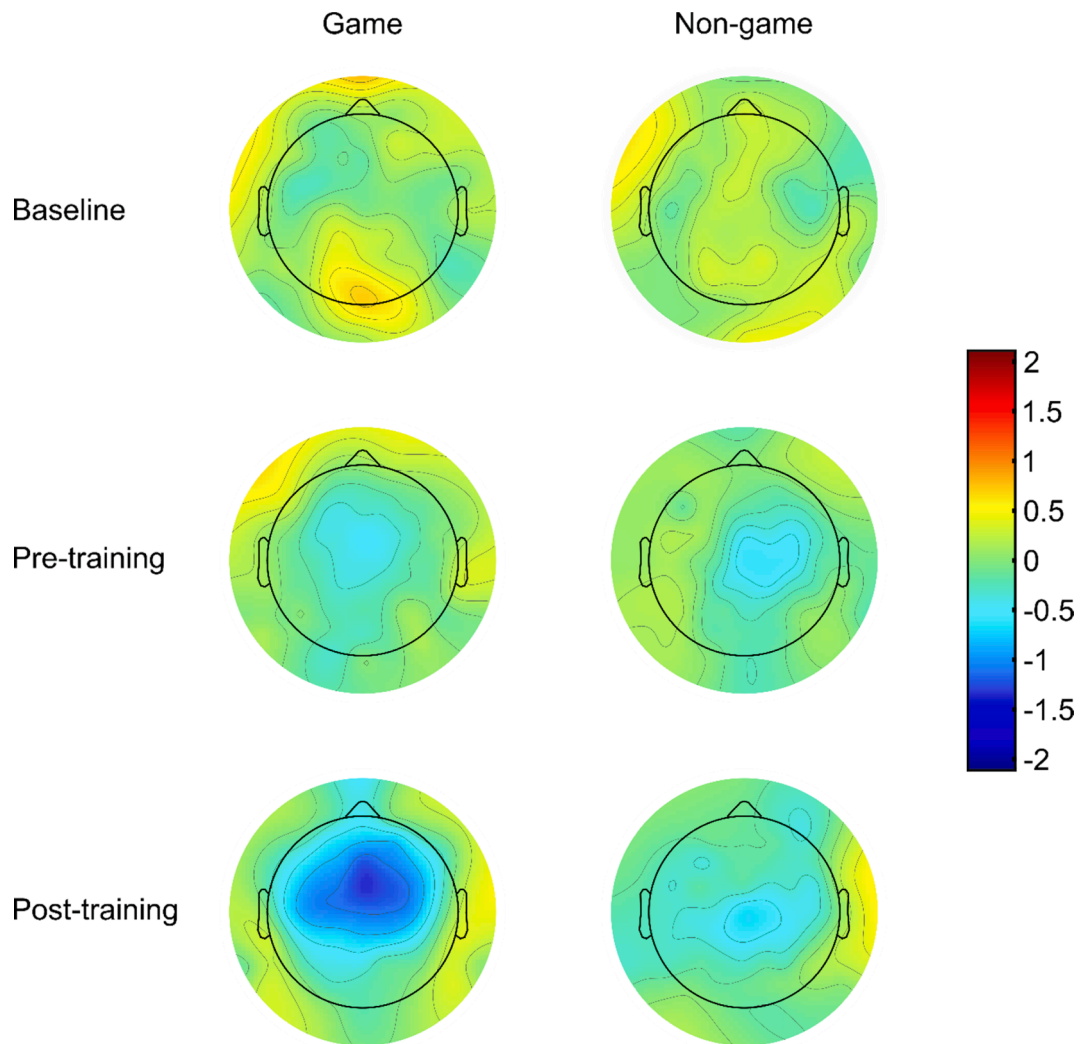


Fig. 4. Scalp topography maps for the MMN responses at baseline measurement, pre-training, and post-training. Amplitude is shown in  $\mu\text{V}$ .

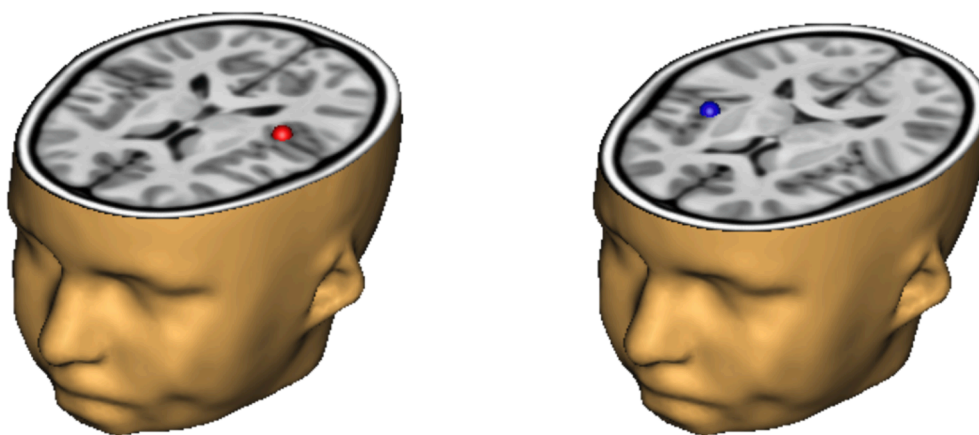


Fig. 5. Left and right temporal sources of a representative participant for post-training game condition depicted on an age-appropriate MRI template.

words containing the target phonemes were trained was also equal between the conditions. The enhancement in the MMN amplitude for game-trained words cannot, therefore, be explained by different playing times of game and non-game conditions. Rather, the effect must be due to the gaming elements *per se* benefiting learning.

The conclusion that gaming elements benefit learning raises a further

question, specifically which elements are the most important ones. The present study investigated the effects of game-based learning by comparing it to learning with a non-game condition that had been stripped of all game-like elements, that is, game graphics, freedom to explore (autonomy), and feedback rewards. That is, any of these could drive learning or their effects might accumulate. For example, feedback



is one of the elements that have been suggested to be relevant for player enjoyment in games (Sweetser & Wyeth, 2005). Previous studies have also shown that feedback contributes positively to learning results (e.g. Erhel & Jamet, 2013). Further, the potential of ASR has been noted in enhancing digital spoken foreign-language learning by providing feedback on the pronunciation accuracy (see Li & Lan, 2021a). Feedback and rewards activate striatum as part of dopaminergic reward networks and have been suggested to increase brain plasticity (for a review, see Nahum & Bavelier, 2020). The real-time feedback given to the children in the game condition may, therefore, have contributed notably to learning. In addition, gaming has been found to improve attention (Oei & Patterson, 2013), which in turn has been shown to drive brain plasticity and affect learning positively (for a review, see Nahum & Bavelier, 2020). Unfortunately, the current study design did not allow us to reliably tease apart the effects of feedback and the other gaming elements. Therefore, in the future it would be interesting to further explore the effectiveness of different game elements separately to determine their contribution to the learning effects of gaming. As noted by Li and Lan (2021a, 2021b), cognitive, social, affective, and neural dimensions are intertwined in DLL and further studies are needed to find out how these factors contribute to the benefits of game-based learning.

As compared with the previous studies on language learning, the present study expands the knowledge of the learning effects of gaming by demonstrating gaming-induced plastic changes in the brain. Whereas previous studies have shown that gaming has benefits on pronunciation (Young & Wang, 2014), vocabulary (Aghlari & Tamjid, 2011; Hsu, 2017; Sandberg et al., 2014), and listening and speaking skills (Liu & Chu, 2010), this study demonstrates how gaming can benefit learning by inducing plastic changes in auditory cortex. The MMN reflects, among other processes, the pre-attentive activation of long-term memory traces for familiar words (Pulvermüller et al., 2001; Shtyrov et al., 2008) and speech sounds (Näätänen et al., 1997; Ylinen et al., 2010). Therefore, a larger MMN response after than before gaming likely indicates that the novel speech sounds or word forms (or both) were learned, enabling their pre-attentive recognition. The post-test difference between conditions thus suggests that gaming induced stronger long-term memory traces than training with the non-game.

Based on the sole MMN results it is difficult to disentangle the contribution of speech-sound and word learning to the observed change in neural activity, because the novel non-native speech sounds co-occurred with the recognition point of the words. Our behavioural learning results may, however, clarify this issue. The vocabulary test suggested that the children were able to name the word meanings almost equally often after the non-game training (16 vs. 19 for non-game vs. game, respectively) and also recognised the spoken words almost equally often (28 vs. 31 for non-game vs. game, respectively). A significant difference in MMN responses but no significant difference in behavioural vocabulary learning suggests that our gaming effects are likely more strongly induced by foreign speech-sound learning than by word learning. This is somewhat expected, since the game training included different words with voiced or voiceless fricatives and thus during training the children were exposed to these speech sounds more often than to the particular words *feather* and *healthy* included in the vocabulary test.

Another possible account for more prominent speech-sound learning over word learning is that getting rewards and feedback on speech production in the game may have drawn the children's attention to the articulation of speech sounds (for attention effects, see Nahum & Bavelier, 2020; Oei & Patterson, 2013). Since it has been suggested that attention control and dopaminergic reward processing are key factors to increase brain plasticity during gaming (Nahum & Bavelier, 2020), this may have particularly fostered speech-sound learning enhancing the MMN. Accordingly, the fact that we did not observe significant MMN enhancement after same kind of speech production task in the non-game lacking feedback may be interpreted to imply that feedback provided by automatic speech recognizer may have played a significant role in

speech-sound learning. However, as stated above, here the effect of feedback could not be reliably separated from the effect of other gaming elements.

In a previous study (Ylinen et al., 2021), we attempted to induce generalisation effects across phonetic contrasts from different languages with articulatory training. Although a data pattern interpreted as a possible generalisation effect was observed in some individuals, the effect was not robust. Similarly, in the current study, the improved processing did not generalise across voiced or voiceless fricatives. Rather, the enhancement was larger for phonemes learned with the game condition even though the fricative types were counterbalanced across conditions. This implies that the training effect was specific to the targeted phoneme trained. Overall, however, our results are in line with previous studies showing that targeted phonetic training (in this case articulatory training with the game) may, at its best, improve phonetic processing in a relatively short time (see Bradlow et al., 1999; Saloranta et al., 2020; Tamminen et al., 2015; Ylinen et al., 2010), although in general the task of learning foreign phonetic contrasts is not trivial (Flege, 1988). Contrasting with some studies addressing general language immersion education or classroom teaching (e.g., Hisagi et al., 2016; Peltola et al., 2007), our findings emphasize the importance of targeted phonetic training that can be further boosted by using gaming elements.

Although we expected to find stronger MMN source in the right hemisphere (Nora et al., 2017; Ylinen et al., 2019), source localisation revealed no significant differences between the strength of the left and right temporal sources of the MMN response in the game condition. Even though the right hemisphere has been shown to be related to foreign word processing in children (Nora et al., 2017; Ylinen et al., 2019), language processing in children has also been demonstrated to shift to the left hemisphere as age increases (McNealy et al., 2011). Individual differences in this shift process could thus explain why no hemispheric difference was found. It was not possible to compare the sources for the game responses and the non-game responses, because too few participants showed an MMN response for the words learned with the non-game. The absence of significant MMN response in this condition suggests, firstly, that our short non-game training was not as effective as the game training in establishing long-term memory representations for foreign speech-sounds or words, and, secondly, that subtle acoustic differences between the stimuli were not acoustically salient enough to elicit MMNs consistently.

To conclude, the results of this study show that gaming may have positive effects on speech-sound learning. Gaming induced more robust learning effects as reflected by enhanced activity of auditory cortex than training with the non-game. The long-term memory traces for foreign speech sounds learned with the game were stronger, as revealed by the increased MMN amplitude. Although in many studies gaming effects are due to more extensive practice, the current results suggest that also gaming elements themselves support speech-sound learning, since we controlled for the amount of exposure and practice.

#### *CRediT authorship contribution statement*

**Katja Junttila:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Funding acquisition. **Anna-Riikka Smolander:** Software, Investigation, Data curation. **Reima Karhila:** Methodology, Software. **Anastasia Giannakopoulou:** Investigation, Writing – review & editing. **Maria Uther:** Investigation, Writing – review & editing. **Mikko Kurimo:** Methodology, Funding acquisition. **Sari Ylinen:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

#### **Declaration of Competing Interest**

The authors declare the following financial interests/personal

relationships which may be considered as potential competing interests: [After the closure of the research project where the experiments were conducted, Anna-Riikka Smolander and Reima Karhila founded a company called Hokema Oy, that attempts to commercialize the game and speech technology developed in this and the preceding research projects. In addition to Smolander and Karhila, Sari Ylinen and Mikko Kurimo have been listed as inventors. This has had no effect on the results of the current study or their interpretation].

### Acknowledgements

We wish to thank our participants, their caregivers, and their schools, principals, and teachers for collaboration; City of Helsinki Education Division, City of Espoo Education and Cultural Services, and City of Vantaa Education Department for access to schools; Vertti Viitanen, Aleksander Nikulin, Olli Rantula, and Krupakar Dhinakaran for programming the game; Seppo Enarvi and Kalle Palomäki for participating in developing the ASR technology; Nella Moisseinen, Mari Falcon, and Pietari Nurmi for their assistance in data collection; Cintia Widmann Etchemaite, Kari Lehikoinen, Anniina Åkerla, and Marika Pajunen for their assistance in data processing; Tommi Makkonen for technical assistance, and Jari Lipsanen for statistical assistance.

### Funding

This work was supported by the Academy of Finland (projects no 274058 and 274075), the Doctoral Programme in Psychology, Learning and Communication, and the University of Helsinki Research Funds.

### Appendix A

The trained words with the target phoneme /θ/:

healthy  
thick  
thin  
thing  
think  
teeth  
mouth  
north  
theatre  
month  
three  
thank you  
thirsty  
throw  
bath  
maths

The trained words with the target phoneme /ð/:

feather  
brother  
father  
mother  
their  
they  
then  
there  
this  
that  
these  
those  
than  
the  
with

### References

- Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers and Education*, 143. <https://doi.org/10.1016/j.compedu.2019.103667>
- Aghlari, L., & Tamjid, N. H. (2011). The effect of digital games on Iranian children's vocabulary retention in foreign language acquisition. *Procedia - Social and Behavioral Sciences*, 29, 552–560. <https://doi.org/10.1016/j.sbspro.2011.11.275>
- Baker, J. M., Martin, T., Aghababayan, A., Armaghanyan, A., & Gillam, R. (2015). Cortical Activations During a Computer-Based Fraction Learning Game: Preliminary Results from a Pilot Study. *Technology, Knowledge and Learning*, 20(3), 339–355. <https://doi.org/10.1007/s10758-015-9251-y>
- Bavelier, D., & Davidson, R. J. (2013). Games to do you good. *Nature*, 494, 425–426. <https://doi.org/10.1038/494425a>
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). <https://doi.org/10.1075/llt.17.07bes>
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer program]. (Version 5.1.45). <http://www.praat.org/>.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify english /r/and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology*, 49(1), 69–77. <https://doi.org/10.1111/bjet.12526>
- Cheour, M., Leppänen, P. H. T., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology*, 111(1), 4–16. [https://doi.org/10.1016/S1388-2457\(99\)00191-1](https://doi.org/10.1016/S1388-2457(99)00191-1)
- deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning and Technology*, 14(2), 74–94. <http://llt.msu.edu/vol14num2/dehaanreedkuwada.pdf>
- Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers and Education*, 67, 156–167. <https://doi.org/10.1016/j.compedu.2013.02.019>
- Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, 51(10), 832–844. <https://doi.org/10.1016/j.specom.2009.04.005>
- Fitzgerald, K., & Todd, J. (2020). Making Sense of Mismatch Negativity Frontiers in Psychiatry (Vol. 11). <https://doi.org/10.3389/fpsy.2020.00468>
- Flege, J. E. (1988). The production and perception of foreign language speech sounds. *Human Communication and Its Disorders, A Review*, 224–401.
- Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning* (pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>
- Franciosi, S. J. (2017). The Effect of Computer Game-Based Learning on FL Vocabulary Transferability. *Source: Journal of Educational Technology & Society*, 20(1), 123–133. <https://doi.org/10.2307/jeductechsoci.20.1.123>
- Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2), 165–198. <https://doi.org/10.1017/S0958344013000013>
- Hisagi, M., Shafer, V. L., Miyagawa, S., Kotek, H., Sugawara, A., & Pantazis, D. (2016). Second-language learning effects on automaticity of speech processing of Japanese phonetic contrasts: An MEG study. *Brain Research*, 1652(April 2015), 111–118. <https://doi.org/10.1016/j.brainres.2016.10.004>
- Hong, J. S., Han, D. H., Kim, Y. I., Bae, S. J., Kim, S. M., & Renshaw, P. (2017). English language education on-line game and brain connectivity. *ReCALL*, 29(1), 3–21. <https://doi.org/10.1017/S0958344016000173>
- Hsu, T. C. (2017). Learning English with Augmented Reality: Do learning styles matter? *Computers and Education*, 106, 137–149. <https://doi.org/10.1016/j.compedu.2016.12.007>
- Kalyakin, I., González, N., Joutsensalo, J., Huttunen, T., Kaartinen, J., & Lyytinen, H. (2007). Optimal Digital Filtering versus Difference Waves on the Mismatch Negativity in an Uninterrupted Sound Paradigm. *Developmental Neuropsychology*, 31(3), 429–452. <https://doi.org/10.1080/87565640701229607>
- Karhila, R., Smolander, A. R., Ylinen, S., & Kurimo, M. (2019). Transparent pronunciation scoring using articulatorily weighted phoneme edit distance. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2019-Sept* (pp. 1866–1870).
- Karhila, R., Ylinen, S., Enarvi, S., Palomäki, K., Nikulin, A., Rantula, O., Viitanen, V., Dhinakaran, K., Smolander, A. R., Kallio, H., Junttila, K., Uther, M., Hämäläinen, P., & Kurimo, M. (2017). SLAK - A game for foreign language pronunciation learning. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017-August* (pp. 3429–3430).
- Kober, S. E., Wood, G., Kiili, K., Moeller, K., & Ninaus, M. (2020). Game-based learning environments affect frontal brain activity. *PLoS ONE*, 15(11 November). <https://doi.org/10.1371/journal.pone.0242573>
- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36(1), 141–157. <https://doi.org/10.1016/j.wocn.2007.03.001>
- Li, P., & Lan, Y.-J. (2021a). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, 1–18. <https://doi.org/10.1017/S1366728921000353>
- Li, P., & Lan, Y.-J. (2021b). Understanding the Interaction between Technology and the Learner: The Case of DLL. *Bilingualism: Language and Cognition*, 1–4. <https://doi.org/10.1017/S1366728921000948>

- Lintunen, P. (2005). Aloitavat yliopisto-opiskelijat englannin kielen ääntäjinä. In L. Kuure, E. Kärkkäinen, & M. Saarenkunnas (Eds.), *Kieli ja sosiaalinen toiminta – Language and Social Action* (pp. 375–386). AFinLA Yearbook. Publications de l'association finlandaise de linguistique appliquée 63.
- Lintunen, P. (2013). The Effect of Phonetic Knowledge on Evaluated Pronunciation Problems. *Proceedings of the Phonetics Teaching and Learning Conference 2013, August, 104*.
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers and Education, 55*(2), 630–643. <https://doi.org/10.1016/j.compedu.2010.02.023>
- Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157. <https://doi.org/10.1111/psyp.12639>
- Mayer, R. E. (2015). On the Need for Research Evidence to Guide the Design of Computer Games for Learning. In *Educational Psychologist* (Vol. 50, Issue 4, pp. 349–353). Routledge. <https://doi.org/10.1080/00461520.2015.1133307>
- McNealy, K., Mazziotta, J. C., & Dapretto, M. (2011). Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science, 14*(6), 1261–1282. <https://doi.org/10.1111/j.1467-7687.2011.01075.x>
- Michel, C. M., & Brunet, D. (2019). EEG source imaging: A practical review of the analysis steps. *Frontiers in Neurology, 10*(APR). <https://doi.org/10.3389/fneur.2019.00325>
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313–329. [https://doi.org/10.1016/0001-6918\(78\)90006-9](https://doi.org/10.1016/0001-6918(78)90006-9)
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432–434. <https://doi.org/10.1038/385432a0>
- Nahum, M., & Bavelier, D. (2020). Video games as rich environments to foster brain plasticity. In *Handbook of Clinical Neurology* (Vol. 168, pp. 117–136). <https://doi.org/10.1016/B978-0-444-63934-9.00010-X>
- Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language, 92*(1), 26–32. <https://doi.org/10.1016/j.bandl.2004.05.005>
- Nora, A., Karvonen, L., Renvall, H., Parviainen, T., Kim, J. Y., Service, E., & Salmelin, R. (2017). Children show right-lateralized effects of spoken word-form learning. *PLoS ONE, 12*(2), 1–20. <https://doi.org/10.1371/journal.pone.0171034>
- Oei, A. C., & Patterson, M. D. (2013). Enhancing Cognition with Video Games: A Multiple Game Training Study. *PLoS ONE, 8*(3). <https://doi.org/10.1371/journal.pone.0058546>
- Peltola, M. S., Tuomainen, O., Koskinen, M., & Aaltonen, O. (2007). The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. *Journal of Psycholinguistic Research, 36*(1), 15–23. <https://doi.org/10.1007/s10936-006-9030-y>
- Peterson, M. (2010a). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation and Gaming, 41*(1), 72–93. <https://doi.org/10.1177/1046878109355684>
- Peterson, M. (2010b). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning, 23*(5), 429–439. <https://doi.org/10.1080/09588221.2010.520673>
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. In *Audiology and Neuro-Otology* (Vol. 5, Issues 3–4, pp. 111–139). S. Karger AG. <https://doi.org/10.1159/000013875>
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., ... Näätänen, R. (2001). Memory Traces for Words as Revealed by the Mismatch Negativity. *NeuroImage, 14*(3), 607–616. <https://doi.org/10.1006/nimg.2001.0864>
- Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H. J., ... Noesselt, T. (2014). The role of reward in word learning and its implications for language acquisition. *Current Biology, 24*(21), 2606–2611. <https://doi.org/10.1016/j.cub.2014.09.044>
- Sajavaara, K., & Dufva, H. (2001). Finnish-English Phonetics and Phonology. *International Journal of English Studies (IJES), 1*(1), 241–256. <https://doi.org/10.6018/ijes.1.1.47781>
- Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology, 147*(November 2019), 72–82. <https://doi.org/10.1016/j.ijpsycho.2019.11.005>
- Sandberg, J., Maris, M., & De Geus, K. (2011). Mobile English learning: An evidence-based study with fifth graders. *Computers and Education, 57*(1), 1334–1347. <https://doi.org/10.1016/j.compedu.2011.01.015>
- Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers and Education, 76*, 119–130. <https://doi.org/10.1016/j.compedu.2014.03.006>
- Shafer, V. L., Morr, M. L., Kreuzer, J. A., & Kurtzberg, D. (2000). Maturation of mismatch negativity in school-age children. In *Ear and Hearing* (Vol. 21, Issue 3, pp. 242–251). <https://doi.org/10.1097/00003446-200006000-00008>
- Shtyrov, Y., Osswald, K., & Pulvermüller, F. (2008). Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. *Cerebral Cortex, 18*(1), 29–37. <https://doi.org/10.1093/cercor/bhm028>
- Sinkkonen, J., & Tervaniemi, M. (2000). Towards Optimal Recording and Analysis of the Mismatch Negativity. In *Audiol Neurootol* (Vol. 5). <http://www.karger.com>
- Sweetser, P., & Wyeth, P. (2005). GameFlow: A Model for Evaluating Player Enjoyment in Games. In *ACM Computers in Entertainment* (Vol. 3, Issue 3).
- Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception - New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology, 97*(1), 23–29. <https://doi.org/10.1016/j.ijpsycho.2015.04.020>
- Tejedor-García, C., Escudero-Mancebo, D., Cardeño-Payo, V., & González-Ferreras, C. (2020). Using Challenges to Enhance a Learning Game for Pronunciation Training of English as a Second Language. *IEEE Access, 8*, 74250–74266. <https://doi.org/10.1109/ACCESS.2020.2988406>
- Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers and Education, 125*, 345–357. <https://doi.org/10.1016/j.compedu.2018.06.020>
- Wechsler, D. (2010). *WISC-IV – Wechsler Intelligence Scale for Children – IV*. Psychologien Kustannus Oy.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*(5), 638–642. <https://doi.org/10.1017/S0048577299981908>
- Yip, F. W. M., & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43*(3), 233–249. <https://doi.org/10.1080/09523980600641445>
- Ylinen, S., Junttila, K., Laasonen, M., Iverson, P., Ahonen, L., & Kujala, T. (2019). Diminished brain responses to second-language words are linked with native-language literacy skills in dyslexia. *Neuropsychologia, 122*. <https://doi.org/10.1016/j.neuropsychologia.2018.11.005>
- Ylinen, S., & Kurimo, M. (2017). Kielenoppiminen vauhtiin puheteknologian avulla. In H. Savolainen, R. Vilkkö, & L. Vähäkylä (Eds.), *Oppimisen tulevaisuus* (pp. 57–69). Gaudeamus.
- Ylinen, S., Smolander, A.-R., Karhila, R., Kakouros, S., Lipsanen, J., Huotilainen, M., & Kurimo, M. (2021). The Effects of a Digital Articulatory Game on the Ability to Perceive Speech-Sound Contrasts in Another Language. *Frontiers in Education, 6* (May), 1–12. <https://doi.org/10.3389/educ.2021.612457>
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience, 22*(6), 1319–1332. <https://doi.org/10.1162/jocn.2009.21272>
- Young, S.-S.-C., & Wang, Y.-H. (2014). The Game Embedded CALL System to Facilitate English Vocabulary Acquisition and Pronunciation. In *Educational Technology & Society* (Vol. 17, Issue 3).
- Zou, D., Huang, Y., & Xie, H. (2019). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning, 1–27*. <https://doi.org/10.1080/09588221.2019.1640745>