

Memory-card vowel training for child and adult second-language learners: A first report

Cite as: JASA Express Lett. 3, 015202 (2023); <https://doi.org/10.1121/10.0016836>

Submitted: 18 October 2022 • Accepted: 21 December 2022 • Published Online: 12 January 2023

Paul Iverson, Begoña Pericas Herrero and Asano Katashima



View Online



Export Citation

ARTICLES YOU MAY BE INTERESTED IN

[Dysarthria detection based on a deep learning model with a clinically-interpretable layer](#)

JASA Express Letters 3, 015201 (2023); <https://doi.org/10.1121/10.0016833>

[An online implementation of a measure of spectro-temporal processing by cochlear-implant listeners](#)

JASA Express Letters 3, 014402 (2023); <https://doi.org/10.1121/10.0016838>

[Mode coupling in a coastal wedge due to quasi-crossing of range-dependent modal eigenvalues](#)

JASA Express Letters 3, 016001 (2023); <https://doi.org/10.1121/10.0016782>



Advance your science and career
as a member of the

ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE



Memory-card vowel training for child and adult second-language learners: A first report

Paul Iverson,^{a)} Begoña Pericas Herrero, and Asano Katashima

Department of Speech Hearing and Phonetic Sciences, University College London, Chandler House, 4 Wakefield Street, London WC1N 1PF, United Kingdom

p.iverson@ucl.ac.uk, begona.herrero.17@ucl.ac.uk, asano.katashima.13@alumni.ucl.ac.uk

Abstract: Japanese adults and Spanish-Catalan children received auditory phonetic training for English vowels using a novel paradigm, a version of the common children's card game *Concentration*. Individuals played a computer-based game in which they turned over pairs of cards to match spoken words, drawn from sets of vowel minimal pairs. The training was effective for adults, improving vowel recognition in a game that did not explicitly require identification. Children likewise improved over time on the memory card game, but not on the present generalisation task. This gamified training method can serve as a platform for examining development and perceptual learning. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Rachel M. Theodore]

<https://doi.org/10.1121/10.0016836>

Received: 18 October 2022 **Accepted:** 21 December 2022 **Published Online:** 12 January 2023

1. Introduction

High-variability phonetic training (HVPT; Logan *et al.*, 1991) has successfully improved phonetic perception by adult second-language learners across a range of contrasts (e.g., Lively *et al.*, 1993; Nishi and Kewley-Port, 2007; Wang *et al.*, 1999). Listeners use a computer program in which they hear a word with talker and phonetic variability between trials, identify the word or phoneme, and are given feedback. That being said, children do not seem able to make full use of HVPT; they can improve with training but typically do not have the same plasticity advantages over adults that are found in real-world language learning, and younger children often learn less than older children (Brekelmans, 2020; Heeren and Schouten, 2010; Shinohara and Iverson, 2021; Wang and Kuhl, 2003; cf. Giannakopoulou *et al.*, 2013). HVPT was not designed to be engaging for children. Moreover, adult work has concluded that HVPT is facilitated by pre-existing first- and second-language category knowledge (Iverson *et al.*, 2005; Iverson and Evans, 2009). Identification training may, therefore, be less effective for young children, who are typically lacking in second-language experience and also have incompletely developed categories in their first language.

To address this issue, the HVPT identification task has recently been combined with a high-variability category discrimination task (Shinohara and Iverson, 2018), which is often used as a proxy for identification assessment among listeners who may not know the category labels (Gottfried, 1984). A category discrimination task involves presenting listeners with two or more stimuli, having them give a discrimination judgment (e.g., which stimulus is different), but with *same* stimuli not being acoustically identical (e.g., the same word spoken by two different talkers). Such a task is considered phonetic/categorical rather than acoustic, because of the lack of an acoustic match among stimuli that are meant to be considered the same. Moreover, tasks involving stimulus variability and memory demands are thought to make it harder for listeners to rely on echoic-memory encodings and depend more on covert abstract encodings (e.g., phonetic categories or labels; e.g., Iverson and Kuhl, 2000). We have found that combined category and auditory discrimination training can produce similar results as identification training among Japanese adults learning English /r/ and /l/, although word or phoneme identification is not explicitly required (Shinohara and Iverson, 2018). Likewise, when discrimination and identification are combined in training, along with other elements, such as animations, to make the task more game-like, Japanese children can improve on English /r/ and /l/ more than Japanese adults, although learning is greater among teenagers than among the youngest children (Shinohara and Iverson, 2021).

We present here a further step toward making phonetic training more game-like and appropriate for children, by creating a computer-based audio version of a common children's card-matching game, often called *Concentration* or *Memory*. This game is played with a shuffled deck of cards laid face-down in a grid. Players turn over cards two at a time, take cards away if they match, and turn them face-down again if they do not match. The objective is to find matching pairs of cards by remembering the locations of cards turned over in previous rounds. Parents sometimes report that their

^{a)} Author to whom correspondence should be addressed.

children are better at this than adults, which may be true individually but has not been found across a sample; [Krojgaard et al. \(2019\)](#) found that completion times for standard games (visual card matching) were about the same for adults [mean (M) = 86 s] and 8-year-olds (M = 87 s) and slightly slower for 6-year-olds (M = 109 s). Our version is similar to the traditional game, except that a word is played when a card is flipped over on a computer screen. The face of the card can be unmarked (i.e., same design for every card, with matching based on sound only) or have a picture that can also be matched. A match in the present implementation is the same word spoken by two different talkers, selected among cards with words from the same cluster of vowel minimal pairs (e.g., *field*, *filled*, *failed*, *filed*). The game is similar to a category discrimination task in that it requires words to be matched across talkers without requiring the words to be identified, but it is performed as an interactive game rather than as a trial-by-trial test and with a higher memory load that has a visual-spatial component. Our aim is to provide an initial evaluation of whether this game can improve second-language phonetic perception, with a broader goal of developing tools for examining age-related plasticity in speech perception that have practical applications for child learning.

This research fits with trends to gamify language learning (see [Acquah and Katz, 2020](#)). These efforts have been mostly driven by practical learning outcomes and advances in speech technology rather than being designed to test scientific hypotheses, although a production training game based on speech technology has been used successfully in phonetic perception experiments ([Ylinen et al., 2021](#)). [Wade and Holt \(2005\)](#) developed an animated shooting game to examine implicit auditory category learning (i.e., when sound recognition is not an explicit requirement of the game); adults can improve on stimuli used in the game (e.g., Japanese learners of English /r/ and /l/), but this training effect has not robustly generalised to untrained stimuli ([Lim and Holt, 2011](#); [Saito et al., 2022](#)). The present memory-card training paradigm similarly allows implicit learning to be explored by manipulating the pictures on the cards and other feedback (e.g., no picture for explicit learning; matched pictures for implicit learning), except within a less visually and motorically complex game, compared to a shooting task, that many parents and children already like to play.

In separate experiments, we trained Japanese adults and Spanish-Catalan bilingual children (6–7 years old) on southern British English vowels. The two groups were a convenience sample, with Japanese adults being tested first because they could perform the tasks online without supervision, followed by testing in a school classroom in Spain, where we were permitted to test in post-lockdown stages of the COVID-19 pandemic. In each game, listeners had 14 cards with vowel minimal pair words spoken by two British English speakers. Participants clicked on pairs of cards, and an animal photo was revealed as the pairs were matched and removed. Learning was evaluated with pre/post-tests using English minimal pair words that were not part of the training set and were produced by different speakers.

2. Experiment 1: Japanese adults

Japanese has five spectrally distinct monophthongal vowels, /a, i, u, e, o/, and a quantity distinction; many English vowels assimilate into the same Japanese category ([Lengeris, 2009](#)), and HVPT has been shown to be effective for Japanese adults ([Nishi and Kewley-Port, 2007](#)). HVPT normally involves lab visits with structured training sessions. Our training program was designed instead for participants to use their own phones and laptops, with participants given a target of 200 games to complete within 10 days and flexibility with how to achieve this goal.

All participants completed pre- and post-training tests of vowel identification and category discrimination, with words and speakers that were not included in the training stimuli. Participants completed the pre/post-tests both in quiet and in noise, with the aim of reducing performance below ceiling for the most accurate listeners (see [Lengeris and Hazan, 2010](#)). We used two noise types, babble and single-talker maskers, to further examine whether learning differentially interacted with energetic and informational masking ([Brungart, 2001](#)). Our initial hypothesis was that the added memory load and attention required by this game may produce a larger effect for single-talker maskers, given that more attention is required to ignore an intelligible talker.

2.1 Method

2.1.1 Subjects

Eighteen adult native speakers of Japanese completed the pre-test, and 12 of these completed all training sessions and the post-tests. All subjects were aged between 18 and 59 years (M = 31 years). Twelve participants were residents of Japan at the time of the experiment, and six were recruited from the Japanese community in London, including one subject who had been in the Netherlands for 3 years.

2.1.2 Stimuli and apparatus

All tests were conducted using software written by us using JAVASCRIPT, PHP, and MYSQL, designed to run within a web browser on the participant's own phone or computer. Subjects were asked to wear headphones and test themselves in a quiet location without interruptions. All recordings had 44 100 16-bit samples/s but were compressed to MP3 format for faster internet delivery.

The stimuli were natural recordings of English vowels used in previous HVPT experiments ([Iverson and Evans, 2009](#); [Iverson et al., 2012](#)). There were 14 vowels, arranged into clusters of three or four based on a hierarchical clustering

analysis of second-language confusion matrices previously collected in our lab: /i, ɪ, aɪ, eɪ/, /ɛ, ɑ, a, ʌ/, /ɒ, əʊ, ɔ/, and /u, aʊ, ɜ/. This clustering was used to reduce the number of response options on each pre/post-trial and to allow for a wider range of minimal pair words within the memory card game. The pre/post-test stimuli comprised recordings of ten speakers (half male and female) of southern British English, producing these vowels in bVt words (*beat, bit, bet, Burt, bat, Bart, bot, but, bought, boot, bait, bite, bout, boat*). The memory card stimuli were produced by a different group of two female and two male speakers of southern British English, with nine minimal pair phonetic contexts recorded for each vowel cluster (e.g., *Ben-barn-ban-bun, field-filled-field-failed, cod-code-chord, shoot-shout-shirt*).

The pre/post-stimuli were presented either without noise, with babble, or with a single-talker masker. The noise was identical to that used in Song et al. (2020). The single-talker masker was a female speaker of Southern British English reading stories that were processed to remove any pauses; listeners heard a random segment of the story during each trial. The babble was created from 12 stories read by the same talker, with each story processed to remove pauses. The signal-to-noise ratio was set to +3 dB in both conditions, a level that we have found has a minimal behavioral effect for native speakers but begins to reduce performance by non-native speakers (Song et al., 2020).

2.1.3 Procedure

Each identification trial involved playing a single stimulus and displaying three or four word response options depending on the cluster (e.g., *beat, bit, bite, bait*). The listener clicked on the word they thought was correct, without an opportunity to hear it again, and received feedback indicating whether they were right or wrong. There were 84 trials: two repetitions × 3 noise conditions × 14 vowels, with the speaker chosen randomly on each trial.

Each category discrimination trial involved playing three words by three speakers in synchrony with three animated frogs. Two words were the same, and one was different. Listeners had to choose the different one and received feedback. There were 54 trials within each test: 18 vowel pairs (i.e., all within-cluster pairs) × 3 noise conditions, with the speakers and oddball positions chosen randomly.

For each memory card game, participants saw 14 cards placed in a 3 × 5 grid with one missing position. Each card had a different recording, comprising 2 talkers × 7 words (i.e., minimal pairs combined from three- and four-vowel clusters, such as field, filled, field, failed, cod, code, chord). The word clusters, speakers, and position of the words in the grid were chosen randomly. Participants clicked on pairs of cards, with a word played when each card was turned over. The underside of each card had identical markings such that matching was by sound only. Non-matching cards were turned back over. Matching cards were removed and progressively revealed an animal photo underneath the cards; the photo was randomly selected from a database such that each game had a different photo. Participants also saw the orthographic form of the word following a match. Participants were instructed to complete 200 games within 10 days, with their participation tracked by the software and a readout at the top of the display indicating the number of games played. In total, subjects spent 158 min, on average, actively playing the game to complete 200 games of training. Their best game thus far was also displayed (i.e., minimum number of moves and time to complete the game), and a button displayed the top ten scores from the participant group as a whole (anonymized with subject codes).

2.2 Results

Figure 1 displays the pre/post-results and performance on the memory card game during training. The statistical analyses were mixed models calculated using the lme4 package (Bates et al., 2015) in R, with the CAR package (Fox and Weisberg, 2019) used to obtain p-values within a type II analysis of deviance table; all fixed effects are described below, and random slopes were not included because the models did not converge in the pre/post-analyses. A logistic mixed-model analysis of

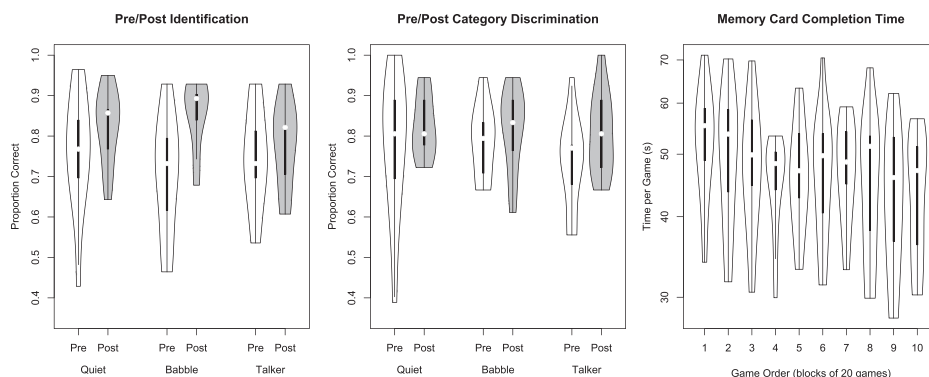


Fig. 1. Violin plots displaying improvements in Japanese adults' English vowel perception in terms of identification and category discrimination accuracy before and after training and progressively faster completion times during memory-card training. Memory card games were binned into blocks of 20 games for display purposes; game order was entered linearly in statistical analyses.

the identification trials, with by-subject and by-item random intercepts, demonstrated that there was a significant main effect of training, $\chi^2(1) = 23.94$, $p < 0.001$, and a significant interaction with the noise condition, $\chi^2(2) = 11.48$, $p < 0.001$. The participants averaged 0.74 proportion correct at the pre-test and 0.82 proportion correct at the post-test, with learning being greatest in the babble condition ($M_{Pre} = 0.70$, $M_{Post} = 0.85$). There was no significant main effect of noise, $p > 0.05$. A logistic mixed-model analysis of the category discrimination trials, with by-subject by-item random intercepts, also demonstrated that there was a significant main effect of training, $\chi^2(1) = 6.74$, $p = 0.009$ ($M_{Pre} = 0.77$, $M_{Post} = 0.81$). There was no significant main effect or interaction with noise, $p > 0.05$. It is notable that the improvement in identification was greater, although the memory card game was more similar to category discrimination than the identification task. Finally, a linear mixed-model analysis of the time to complete each game, with by-subject random intercepts, demonstrated that there was a significant effect of order, $\chi^2(1) = 57.442$, $p < 0.001$, with less time required to complete each game as training progressed. The time to complete a game was highly correlated with the number of moves, $r = 0.89$. In summary, there is clear evidence that memory-card training works for Japanese adults and English vowels, with evidence of learning within training as well as improvements for untrained tasks, words, and talkers.

3. Experiment 2: Spanish-Catalan children

Children, 6–7 years old, participated in this study as part of their school English classes on the island of Majorca (Spain). The Majorcan dialects of Catalan and Spanish are both official languages on Majorca, both are spoken at school, and one or both were spoken at home by the children. Spanish has a five-vowel system, /a, i, u, e, o/, and Majorcan Catalan has eight vowels, /a, ε, e, i, ə, u, ɔ, o/. The study was conducted in four 90-min classes over 2 weeks. The interface was the same as in experiment 1, except that children used school-provided laptops and headphones at their desks, and an experimenter had centralised control over the programs.

Pilot testing with children suggested that they were able to play the game with unmarked cards (i.e., *audio-only*), but they enjoyed the game more when there were symbols on the underside of the cards. In the *symbols + audio* version of the game, we used emoji-like symbols organised in themes (e.g., facial expressions, food). The symbols and audio words were not related in meaning, but matching words always had matching symbols. In addition to making the task easier and more engaging, the symbol condition was designed to test implicit learning (Lim and Holt, 2011); children heard the words but did not need to correctly perceive and remember the words to perform the task. Blocks of the category discrimination task were interleaved with the memory card games, allowing us to track performance throughout training rather than adopting the strict pre/post-test design of experiment 1. Within each session, participants thus had tasks changing between category discrimination, audio-only memory cards, and symbol + audio memory cards. There was no identification task, owing to the children being relatively new to English.

3.1 Method

3.1.1 Subjects

Thirty children (mean age = 6; 3 years, all born the same calendar year) were recruited from a primary school in the town of Inca in Majorca, Spain. These children are exposed to Spanish and the Majorcan dialect of Catalan at school on a daily basis. Parental responses to a language background questionnaire revealed that five children spoke a language other than Spanish or Catalan at home; they were allowed to participate in class, but their data were omitted from the analysis. One child was omitted because they were only present for two sessions. Of the remaining children, 13 spoke Catalan at home (typically Spanish as well), and 11 spoke only Spanish at home.

3.1.2 Stimuli and apparatus

The stimuli and apparatus were the same as in experiment 1, except that the testing was conducted on school-supplied headphones and laptops, and there were no conditions with added noise.

3.1.3 Procedure

The tasks were fundamentally the same as in experiment 1, except for the order and number of trials. The children were divided into two classes (i.e., about 15 students in a class), and each class played the games for four sessions of their English lessons. Given that category discrimination was tested throughout the 4 days, we used untrained words and speakers on the first and last days only for generalisation tests (CD-U) and had trained words and speakers on the second and third day (CD-T). As described above, there was an audio-only version of the memory card game with identical markings on the underside of each card (MC-A) and a symbol + audio version in which there were matching emoji-like symbols underneath the cards that were semantically unrelated to the simultaneous word (MC-SA). The participants completed the tasks in the following order: (day 1) CD-U, MC-A, CD-U, MC-SA, CD-U; (day 2) CD-T, MC-SA, CD-T, MC-A, CD-T; (day 3) CD-T, MC-A, CD-T, MC-SA, CD-T; (day 4) CD-U, MC-SA, CD-U, MC-A, CD-U. The children completed as many trials as they could during each test block, and the length of each block was determined by the experimenter at the time of testing (i.e., depending on the amount of time remaining in the class and the number of trials completed by most children, which was monitored via a control panel on the experimenter's phone). Each class lasted 90 min, but that

included time for other classroom activities, set-up time, and breaks; the aim was to have 30 min of testing time. Given the realities of testing classrooms full of young children, the numbers of trials were not entirely uniform. For example, one class was unable to complete the final two phases on the first day because they ran out of time, and seven children missed one session because of COVID-19 absences. This non-uniformity can be accommodated in mixed-model statistics. Over the entire course of testing, children spent an average of 85 min on category discrimination tasks (533 trials), 30 min on audio-only memory card games (14 games; 378 clicks on card pairs), and 28 min on symbol + audio memory card games (27 games; 414 clicks on card pairs), including active game-playing time only.

3.2 Results

The statistical analyses were mixed models calculated using the lme4 package (Bates et al., 2015) in R, with the CAR package (Fox and Weisberg, 2019) used to obtain *p*-values within a type II analysis of deviance table; all fixed effects are described below, and random slopes were not included because our models did not converge in the pre/post-analyses. For the category discrimination task, a logistic analysis with by-subject and by-item random intercepts demonstrated that there was a significant main effect of training day, $\chi^2(1) = 5.87, p = 0.015$, but as displayed in Fig. 2, children actually became worse when they reached the final day ($M = 0.56, 0.58, 0.55$, and 0.49 proportion correct, respectively, for days 1–4). Similarly, there was a significant main effect of test block within days, $\chi^2(1) = 15.45, p < 0.001$, with children becoming worse on this task as each day progressed ($M_{\text{Block } 1} = 0.59, M_{\text{Block } 3} = 0.51$). Children spontaneously reported that they became bored with the category discrimination task (e.g., asking whether they could play a different game). Thus, we do not have evidence that training improved the perception of untrained stimuli or words, and we think that the results were task related rather than reflecting a genuine decline in English vowel perception. There was also a main effect of home language, with children from Spanish-only homes scoring lower than the children from Catalan speaking homes, $\chi^2(1) = 7.87, p = 0.005$ ($M_{\text{Catalan}} = 0.60, M_{\text{Spanish-only}} = 0.48$), suggesting that the greater number of vowel categories in Catalan, or their more bilingual experience, aided vowel perception. There were no other significant main effects or interactions, $p > 0.05$.

In contrast, children improved over time on the auditory-only memory card game (Fig. 2) and spontaneously reported that they preferred the memory card games over category discrimination. A linear mixed model, with by-subject random intercepts and the log game-completion time as a dependent measure, demonstrated that there was a significant difference in the times to complete auditory-only and symbol + auditory games, $\chi^2(1) = 876.61, p < 0.001$, and a significant interaction between memory card condition and training day, $\chi^2(1) = 9.69, p = 0.002$. As displayed in Fig. 2, children were faster at completing the game when there were symbols ($M = 64$ s) than with audio only ($M = 131$ s). Given that there was a large difference between the game conditions, the analysis was rerun separately for audio-only and symbol + audio games. There was a significant effect of training day for the audio-only game, $\chi^2(1) = 12.31, p < 0.001$, with children becoming faster at the game over time ($M = 150, 132, 121$, and 121 s, respectively, for days 1–4), but no significant training effect for the symbol + audio game, $\chi^2(1) = 0.86, p = 0.357$. In this case, the symbol + audio game essentially acted as a control, suggesting that there was learning in the auditory-only condition beyond any changes in their abilities to perform the task itself and no evidence of incidental learning in the symbol + audio condition. Finally, there were no significant effects of language background in the memory card game, with the closest being a main effect of language background in the analysis with both game conditions, $\chi^2(1) = 3.22, p = 0.072$. The direction was the same as in the category discrimination game (i.e., faster for children who speak Catalan at home; $M_{\text{Catalan}} = 82$ s, $M_{\text{Spanish-only}} = 90$ s), but this measure may be less sensitive to language background given the additional task demands of the game (e.g., memory). As in experiment 1, the time to complete a game was highly correlated with the number of moves, $r = 0.90$.

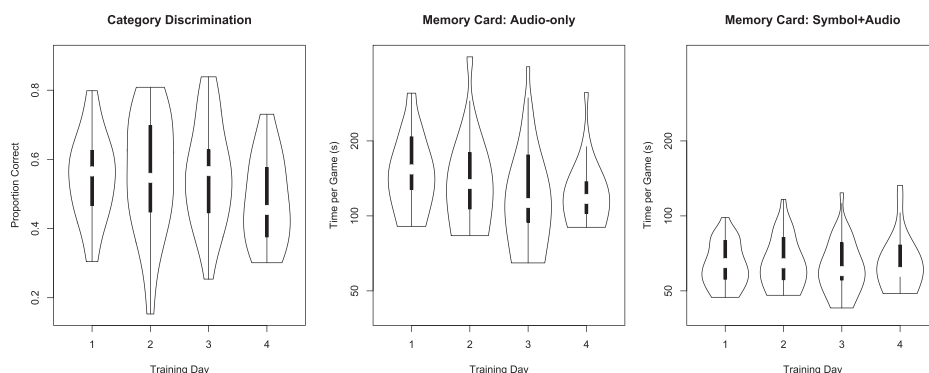


Fig. 2. Violin plots showing that Catalan-Spanish children decrease in category discrimination performance over the 4 days of training but improve in their time to complete audio-only memory card games. They do not change significantly in terms of their speed in performing the symbol + audio version of the memory card game.

4. General discussion

The results demonstrate that a memory-card trainer can improve English vowel perception by Japanese adults and that Spanish-Catalan young children are willing and able to play this game. Our evidence is weaker that this improved child performance, given that performance declined on the category discrimination task. However, there was evidence of learning for children within the memory card game for the audio-only condition and significantly less learning in the symbol + audio game that could be performed by matching symbols.

Earlier work found that children and adults perform similarly on memory card games (Krojgaard *et al.*, 2019). In the present study, we found that children playing symbol + audio memory card games (mean = 62 s per game) approached the speed of adults matching audio only (mean = 47 s per game), but the children were more than twice as slow on the auditory-only game (mean = 129 s per game). This could be due to the better vowel category discrimination skills of the adults in our study, but it is possible too that children have a specific difficulty with the auditory-only game because of their less mature verbal short-term memory and phonological awareness (e.g., Snowling and Hulme, 1994). We have speculated that this maturational factor may limit the effectiveness of auditory training for children (Shinohara and Iverson, 2021), and one way around this limitation may be the inclusion of symbols in the memory card game. We do not have evidence here of incidental learning in the symbols + audio condition, but there could have been a floor effect due to symbols being easy; it is conceivable that incidental auditory learning may have been found if our category discrimination task had been more effective. Additional experiments are required to test factors such as incidental learning, generalisation, and developmental changes. We can conclude, for now, that this memory-card training paradigm works for adults and is engaging for young children.

Acknowledgments

We are grateful to Magda Rojas Mora, Holly Bradley, and Yasuaki Shinohara for comments toward the development of the training game and to Maria Jose Rivero for facilitating child testing. This work was funded by the Economic and Social Research Council of the UK.

References and links

- Acquah, E. O., and Katz, H. T. (2020). "Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review," *Comput. Educ.* **143**, 103667.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**, 1–51.
- Brekelmans, G. (2020). *Phonetic Vowel Training for Child Second Language Learners: The Role of Input Variability and Training Task* (University College London, London).
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression* (Sage, Los Angeles).
- Giannakopoulou, A., Uther, M., and Ylinen, S. (2013). "Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English," *Child Lang. Teach. Ther.* **29**, 201–218.
- Gottfried, T. L. (1984). "Effects of consonant context on the perception of French vowels," *J. Phon.* **12**, 91–114.
- Heeren, W. F. L., and Schouten, M. E. H. (2010). "Perceptual development of the Finnish /t-t/ distinction in Dutch 12-year-old children: A training study," *J. Phon.* **38**, 594–603.
- Iverson, P., and Evans, B. G. (2009). "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers," *J. Acoust. Soc. Am.* **126**, 866–877.
- Iverson, P., Hazan, V., and Bannister, K. (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," *J. Acoust. Soc. Am.* **118**, 3267–3278.
- Iverson, P., and Kuhl, P. K. (2000). "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?," *Percept. Psychophys.* **62**, 874–886.
- Iverson, P., Pinet, M., and Evans, B. G. (2012). "Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels," *Appl. Psycholinguist.* **33**, 145–160.
- Krojgaard, P., Sonne, T., Lerebourg, M., Lambek, R., and Kingo, O. S. (2019). "Eight-year-olds, but not six-year-olds, perform just as well as adults when playing Concentration: Resolving the enigma?," *Conscious. Cogn.* **69**, 81–94.
- Lengeris, A. (2009). "Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese," *Phonetica* **66**, 169–187.
- Lengeris, A., and Hazan, V. (2010). "The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek," *J. Acoust. Soc. Am.* **128**, 3757–3768.
- Lim, S. J., and Holt, L. L. (2011). "Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization," *Cogn. Sci.* **35**, 1390–1405.
- 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. Am.* **94**, 1242–1255.
- 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. Am.* **89**, 874–886.
- Nishi, K., and Kewley-Port, D. (2007). "Training Japanese listeners to perceive American English vowels: Influence of training sets," *J. Speech Lang. Hear. Res.* **50**, 1496–1509.

- Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., and Tierney, A. (2022). "Incidental and multimodal high variability phonetic training: Potential, limits, and future directions," *Lang. Learn.* **72**(4), 1049–1091.
- Shinohara, Y., and Iverson, P. (2018). "High variability identification and discrimination training for Japanese speakers learning English /r/-/l/," *J. Phon.* **66**, 242–251.
- Shinohara, Y., and Iverson, P. (2021). "The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers," *J. Phon.* **89**, 101108.
- Snowling, M., and Hulme, C. (1994). "The development of phonological skills," *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* **346**, 21–27.
- Song, J., Martin, L., and Iverson, P. (2020). "Auditory neural tracking and lexical processing of speech in noise: Masker type, spatial location, and language experience," *J. Acoust. Soc. Am.* **148**, 253–264.
- Wade, T., and Holt, L. L. (2005). "Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task," *J. Acoust. Soc. Am.* **118**, 2618–2633.
- Wang, Y., and Kuhl, P. K. (2003). "Evaluating the 'Critical Period' hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age," in *Proceedings of the 15th International Congress on Phonetic Sciences*, August 3–9, Barcelona, Spain, pp. 1537–1540.
- Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). "Training American listeners to perceive Mandarin tones," *J. Acoust. Soc. Am.* **106**, 3649–3658.
- 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.