6-1-2022

Learning unfamiliar words and perceiving non-native vowels in a second language: Insights from eye tracking

Félix Desmeules-Trudel
Western University

Marc F. Joanisse
Western University, marcj@uwo.ca

Acta Psychologica 226 (2022) 103590

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Learning unfamiliar words and perceiving non-native vowels in a second language: Insights from eye tracking

Félix Desmeules-Trudel *, Marc F. Joanisse

Department of Psychology, Brain and Mind Institute, The University of Western Ontario, Canada

ARTICLE INFO

Keywords:
Spoken word recognition
Word learning
Vowel discrimination
Second language
Eye tracking

ABSTRACT

One of the challenges in second-language learning is learning unfamiliar word forms, especially when this involves novel phoneme contrasts. The present study examines how real-time processing of newly-learned words and phonemes in a second language is impacted by the structure of learning (discrimination training) and whether asking participants to complete the same task after a 16–21 h delay favours subsequent word recognition. Specifically, using a visual world eye tracking paradigm, we assessed how English listeners processed newly-learned words containing non-native French front-rounded [y] compared to native-sounding vowels, both immediately after training and the following day. Some learners were forced to discriminate between vowels that are perceptually similar for English listeners, [y]-[u], while others were not. We found significantly better word-level processing on a variety of indices after an overnight delay. We also found that training [y] words paired with [u] words (vs. [y]-Control pairs) led to a greater decrease in reaction times during the word recognition task over the two testing sessions. Discrimination training using perceptually similar sounds had facilitative effects on second language word learning with novel phonemic information, and real-time processing measures such as eyetracking provided valuable insights into how individuals learn words and phonemes in a second language.

1. Introduction

In the domain of speech perception, it has been shown that the presence of sounds in a target second language (L2) that do not exist in a listener's native language can pose a major challenge if listeners conflate unfamiliar phoneme contrasts. This is at the basis of many difficulties for discriminating L2 phonemes and learning new words in the target language. For example, French employs a series of front-rounded vowels [y, ø, œ] that English does not, leading to confusion with [u] and other back-rounded vowels by English listeners (Desmeules-Trudel & Joanisse, 2020; Flege, 1987; Flege & Hillenbrand, 1984; Gottfried, 1984; Levy & Strange, 2008; Rochet, 1995; Tyler et al., 2014). This process, known as *phonetic assimilation* (Best, 1995; Best & Tyler, 2007), has been the object of several studies using nonmeaningful stimuli such as isolated syllables. While these studies have shed some light on L2 phonetic assimilation during perception, its impact on *word learning* remains less well understood. Some studies have shown that L2 listeners asymmetrically process words that have sounds that are absent from their L1 (Cutler et al., 2006; Escudero et al., 2008; Weber & Cutler, 2004). For example, English words with [l] are more easily recognized than words with [ɹ] by Japanese listeners, although both these consonants are absent in Japanese (Cutler et al., 2006), and English words that had an [ɛ] sound were better recognized by Dutch listeners (proficient in English) than words with [æ] (Escudero et al., 2008). Thus, in these studies, L2 listeners were confused when asked to recognize L2 words that have sounds that are absent from their native language even when proficient in the target L2.

Another illustration of this relationship between sound perception and the lexicon in L2 listeners can be found in Darcy et al. (2012), who found that intermediate and advanced learners of French (L1-English) performed similarly on a speech identification ABX task, but that the advanced group had an advantage over the intermediate group on a lexical priming task using [y]-[u] vowels (among other vowel pairs). Thus, better L2 lexical representations did not motivate stronger sound identification abilities in advanced learners compared to intermediate learners. Furthermore, this suggests that although lexical representations and speech sounds are not entirely separated for learning and recognition, they are to some extent independent. In other words, listeners can establish new lexical representations (i.e., learn words)

* Corresponding author at: Department of Psychology, Brain and Mind Institute, The University of Western Ontario, Western Interdisciplinary Research Building, London, ON N6A 3K7, Canada.

E-mail address: fdesmeul@uwo.ca (F. Desmeules-Trudel).

https://doi.org/10.1016/j.actpsy.2022.103590
Received 14 January 2021; Received in revised form 7 April 2022; Accepted 9 April 2022
Available online 16 April 2022

without changing phonetic-phonological representations of the sounds (e.g., performance on a speech identification or discrimination task) that compose these newly-learned words. This is consistent with research focusing on task-specific use of different types of information (Ullman & Lovelett, 2018). Darcy et al.'s (2012) research is also consistent with previous 'lexicon first' work (Cutler et al., 2006; Escudero et al., 2008; Weber & Cutler, 2004) which suggests that lexical contrasts can be established even if L2 phonemes are not differently categorized in a speech perception task.

In the field of language learning, there has also been emerging interest in novel word learning as it relates to processes of memory consolidation in current psycholinguistic research, where memory is encoded or strengthened in the service of long-term retrieval of linguistic information. Of interest to the current paper, prior work on memory consolidation has shown mixed effects on speech perception in L2 learning. For example, studies found improvement in non-native sound discrimination and identification following a night of sleep under certain circumstances (Earle & Myers, 2015a; Earle & Myers, 2015b; Fuhrmeister, 2019). More concretely, this body of research has examined learning of non-native sounds (e.g., retroflex [ɖ] and dental [d̪] from Hindi with English listeners), but results differed across perception tasks (e.g., discrimination and identification; Earle & Myers, 2015a). Further studies also showed mixed results (see review by Fuhrmeister, 2019), where consolidation was found or not depending on the task (Eisner & McQueen, 2006), exposure to variability (Fuhrmeister & Myers, 2017) or production of new sounds and words (Baese-Berk & Samuel, 2016). This suggests that consolidation of procedural knowledge (i.e., generalization of a skill to a new input) and thus improvement of newly-learned speech sound *discrimination* relies on different neural mechanisms than consolidation of declarative knowledge (Ullman & Lovelett, 2018). The procedural nature of speech discrimination tasks and the declarative nature of lexical recognition or phonetic identification tasks likely explain the observed differences in the impact of memory consolidation on L2 (perceptual) learning. One of the interesting avenues for investigations thus focuses on the interactions between L2 speech perception abilities and how well listeners can learn words that contain foreign speech sounds.

With these considerations in mind, parallels can be drawn between the PRIMIR model of language development in children (Curtin et al., 2011; Werker & Curtin, 2005) and learning L2 words later in life. PRIMIR suggests that the encoding for phonology and lexical items are related but somewhat different. For instance, the model proposes separate levels of representation for acoustic and lexical items, and that the two types of information are thus learned separately. During acquisition, the various task demands require the use of the two types of information such that learning a new phoneme contrast may have different task affordances than learning a new word. Applying these claims to L2 acquisition, we propose that learners will attend to linguistic information in different ways depending on the processing situation or task. Concretely, that means that performance on an L2 speech discrimination task, and post-consolidation improvement or lack thereof, are likely to be relatively independent from a L2 word recognition task. In other words, we do not expect that listeners would improve their performance on both an L2 speech discrimination task and an L2 word recognition task after memory consolidation of newly-learned L2 information (e.g., phonemes, words).

Experimental evidence combining aspects of L2 speech sound perception, new-word learning and memory consolidation is scarce (Ullman & Lovelett, 2018). As mentioned above, the literature has primarily focused on how novel words are consolidated within the L1, and how consolidation impacts phonetic perception abilities in an L2. In this paper, we focus on how TRAINING to discriminate L2 sounds influences word learning and processing, with a special emphasis on eye tracking measures that can provide finer-grained information about processing by capturing real-time lexical access. We further include a second test session that was run following overnight delay, allowing us to consider the potential influence of consolidation effects on speech and word learning/processing abilities in an L2. Doing so, we aim to contribute to answering how L2 listeners learn, represent, and use newly-learned phonetic and lexical information in their target language, and if there are links between phonetic and lexical learning in an L2. All participants completed a speech discrimination task, a word learning task, and a word recognition task twice, over two days, in order to investigate the interactions between the sound-discrimination and word-learning/processing abilities within listeners, considering memory consolidation of new linguistic information.

Specifically, we investigated how Canadian French vowels [ɑ, e, u] and [y], the last of which is absent from the English inventory, are discriminated by native speakers of English with little or no knowledge of French. Following this discrimination task, we investigated how words that contain these vowels are learned and recognized: the word learning task was structured to promote (or not) appropriate categorization of assimilable [y], which is perceptually similar to the [u] category for English-native listeners as shown in previous investigations (Desmeules-Trudel & Joanisse, 2020; Levy & Strange, 2008). The sound discrimination task was completed at the very beginning of the procedure on day 1, and a second time on day 2 (i.e., after the day 1 learning and recognition tasks, and a night of sleep). The learning task was completed once on day 1. The word recognition task was completed twice over two days.

The design of our experiment allowed us to explore how the possibility of assimilating [y]-words to [u]-word representations impacts word processing in an L2, analogously to Kapnoula and McMurray (2016).[1] We focused on native English listeners who had no or little knowledge of French. During the learning phase, we manipulated the structure of the word learning process by promoting discrimination of perceptually "equivalent" vowels for L2 listeners: in one learning condition, assimilable [y]-[u] words were presented together and non-assimilable [ɑ]-[e] words were presented together, thus promoting discrimination of [y]-[u] during learning ([y]-[u] group). In the other learning condition, non-assimilable [y]-[e] words were presented together and non-assimilable [u]-[ɑ] words were presented together, thus this condition did not promote discrimination of assimilable [y] and [u] sounds ([y]-Control group). Combined with eye tracking, which offers the possibility to observe real-time processing abilities via proportions of fixations to target and competitor images (see below), we could more closely assess the (co)activation of newly-learned words.

Our study made use of the Visual World Paradigm (VWP; Allopenna et al., 1998; Dahan et al., 2001; Huettig et al., 2011), an experimental procedure which takes advantage of lexical competition effects while incorporating a naturalistic and dynamical task. In this paradigm, participants are asked to fixate to images on a display as they listen to speech and then provide an overt response such as a mouse click or button press. Measuring proportions of fixations to the target and competitor images within a time window of interest provides an idea of how each word is activated as speech unfolds, e.g., how words that contain [y] as well as other vowels are learned, consolidated and recognized, using both overt recognition and eye tracking.

On both testing days, offline measures of sound discrimination and word recognition (accuracy for correct responses) were recorded as well as eye movements for the word recognition task. Let us recall that participants performed a syllable discrimination (AX test) before learning

[1] Kapnoula and McMurray (2016) trained native listeners of English on a series of tasks, in which they were asked to discriminate close lexical neighbours (*cat* vs. *cap*; high-competition condition) or unrelated words (*cat* vs. *neck*; low-competition condition), and subsequently assessed their word recognition abilities depending on the training condition. They determined that listeners trained in the high-competition condition were faster at fixating to the target image (e.g., *cat*), such that word activation and competition processes were impacted by the type of training that was completed.

words on the first day and after consolidation (i.e., on the second day) in order to assess perception and improvement in performance post-consolidation of the key non-native vowel contrast. In our design, we used four talkers in order to evaluate consolidation of lexical items and favour generalization of word representations over variable phonetic information. This minimized the likelihood that listeners would focus on learning idiosyncratic acoustic details, and potentially enhancing their ability to generalize to novel exemplars during testing.

As mentioned above, we predicted an overall improvement in performance for the word recognition task after delay, since there is no specific reason to think that consolidation effects that have been previously shown for L1 would significantly differ for L2 word learning. We also expected learners who were confronted with [y]-[u] word pairs during learning to perform better on the word recognition task than learners who did not learn pairs of words that could not be phonetically assimilated, analoguously to Kapnoula and McMurray's (2016) results in an L1. This prediction is based on the fact that TRAINING to categorize a target word against a highly similar word (as in our [y]-[u] group) would inhibit lexical competition when compared to low-competition TRAINING (as in our [y]-Control group), therefore enhancing target recognition (i.e., faster and more robust) in the L1 (Kapnoula & McMurray, 2016). Participants in the [y]-[u] group were expected to demonstrate better recognition of these words on day 1, due to better competitor inhibition abilities post-training. Likewise, improvements observed on day 1 should be retained or improved on day 2 in the [y]-[u] group as well, perhaps reflecting memory consolidation effects previously found in similar studies of L1 novel word learning. Alternatively, it is possible that TRAINING effects in the [y]-Control group would improve following the delay period, causing group differences to disappear on the second testing DAY.[2]

2. Experiment

2.1. Methods

2.1.1. Participants

Twenty-four native speakers of English (15 female, 9 male, between 18 and 35 years old, $M = 21.75$ years, $SD = 4.18$) were paid or received partial course credit for their participation in the experiment. Of these twenty-four listeners, two were excluded from the analyses of the word-learning and word recognition tasks due to a technical problem (corrupted eye tracking data files, both from the [y]-Control group). All listeners reported having normal hearing, normal or corrected-to-normal vision, and reported having no history of language, hearing or speech impairment. All participants completed a background language questionnaire, and most self-reported knowing French as an L2 at a null to low level of proficiency, as well as other languages at a low or fair level of understanding and poor level of speaking proficiency (i.e., Italian, Klingon, Cantonese, German).[3] Four participants considered themselves bilingual (one in Punjabi, one in Hindi, one in Spanish, one in Croatian), but none of the reported languages had a front rounded vowel in its inventory. All participants completed two testing sessions, which were separated by 16 to 21 h ($M = 18.5$, $SD = 1.42$): the first session was conducted on day 1 in the afternoon, and the second session in the following morning. Participants self-reported sleeping between four and ten hours overnight ($M = 7$, $SD = 1.57$). Participants in the [y]-[u] group slept for 6.5 h on average and participants in the [y]-Control group for 7.5 h, and the difference in number of slept hours did not statistically differ across groups ($t = -1.37$, $df = 12.65$, $p = 0.196$).

2.1.2. Procedure and stimuli

On day 1, participants first completed a speech discrimination task, then a word-learning (i.e. training) task, and finished the session with a word recognition task (approximatively 60 to 90 min overall). On day 2, all listeners completed the speech discrimination task a second time, received a quick recall of the words learned the previous day (i.e., they were presented with the learned auditory words and associated pictures while passively listening, similarly to the passive learning procedure on day 1, see below for details), followed by the word recognition task (approximately 45 to 60 min overall), but no active-learning task on day 2. Stimulus in all three tasks were presented in (and the random orders controlled by) Experiment Builder (SR Research) version 2.1.140.

2.1.2.1. Sound discrimination task. Discrimination of the target vowel [y] was assessed using an AX task, to establish baseline perception performance on day 1 and to determine any improvement on day 2 following consolidation of newly-learned words. It is important to keep in mind that word learning, word recognition (day 1 only) and sleep all occurred before the speech discrimination task on day 2. Consequently, improvement on the discrimination task could be due to either consolidation of the new-L2 speech sounds or just practice with (different, i.e., words) stimuli as spoken by the same talkers.

Stimuli were CV syllables combining the onset consonants [b, d, ɡ, p, t, k] ([t] and [d] were not found before [y] since they are affricated [t^s] and [d^z] in this context in some varieties of Canadian French) and nucleus vowels [ɑ, e, u, y]. They were pronounced by four native speakers of Canadian French (3 males and 1 female, ages 24–31 years). Syllable duration was 190 ms on average (SD = 37.6 ms, *range* 111–279 ms). Note that training with variable stimuli has been found to promote generalization of newly-learned words (Barcroft & Sommers, 2005), but not speech discrimination (Earle & Myers, 2015b; Fuhrmeister & Myers, 2017, 2020). Speakers produced a reduplicated target syllables, e.g., for the target stimulus [ke] the speakers pronounced [keke]. The last syllable of each was then hand-segmented in Praat 6.0.43 (Boersma & Weenink, 2018) and normalized for amplitude at 70 dB.

The syllable pairs could include syllables pronounced by more than one speaker, and inter-stimulus interval was set at 500 ms (Werker & Tees, 1984). For each pair, participants were asked if they thought the vowels rhymed or not. Across syllable pairs, half contained the same vowel and half contained different vowels (72 each, 144 pairs total), so that participants heard each same-vowel pair 18 times ([ɑ-ɑ], [e-e], [u-u], [y-y]) but with different individual tokens (talkers) for each pair. For the different-vowel pairs, each possible combination was repeated twelve times with the order counterbalanced, using different tokens of the target syllables between repetitions.

2.1.2.2. Word learning task. Participants next completed the training task. Stimuli for the learning and the word recognition tasks were four sets of four nonwords composed of two CV syllables ($C_1V_1C_2V_2$), pronounced by the same speakers as the AX task, and produced with stress on the final syllable following French phonotactics. For each set, $C_1V_1C_2$ were held constant, and the final vowel V_2 was one of [ɑ, e, u, y] (Fig. 1). Total word duration was 438 ms on average (SD = 69 ms, *range* = 324–703 ms), with the first syllable lasting 185 ms ($SD = 52$ ms, *range* = 108–380 ms) and the second syllable lasting 253 ms (SD = 42 ms, *range* = 171–377 ms) on average. Note that duration of the second syllable was calculated from the nonwords' first syllable offset to the offset of the second syllable. The second syllable was significantly longer than the first syllable in our stimuli ($t = -9.69$, $df = 174.63$, $p < 0.001$), as expected given the tendency for noncontrastive word stress to fall on the

[2] Note that although participants were tested over two days, our design cannot unambiguously identify consolidation effects. Further information to this effect can be found in the Discussion of the paper.

[3] Some self-reported low proficiency in French was expected as the Canadian secondary school students are required to complete at least two years of French courses. Two learners also reported poor knowledge of Cantonese, and one reported poor knowledge of German. We do not expect poor knowledge of these languages to have an impact on performance, despite the fact that they have front-rounded vowels in their inventories.

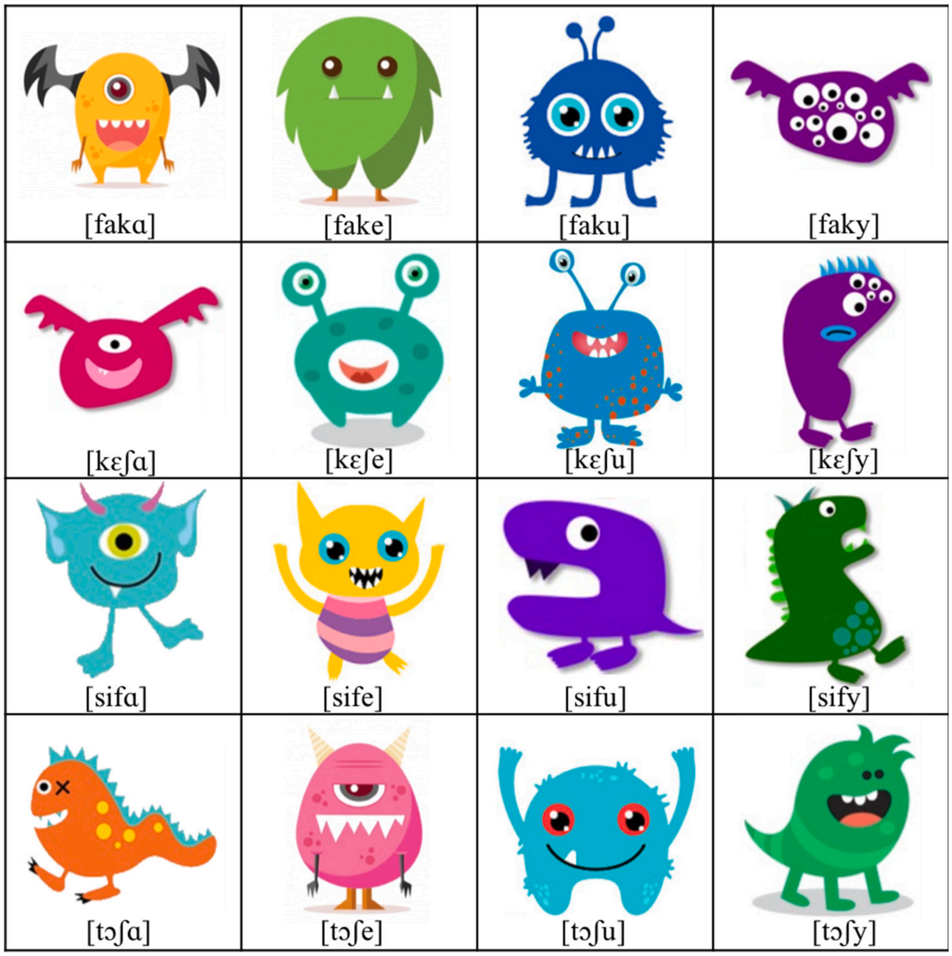

Fig. 1. Visual stimuli and associated words used in the word learning and word recognition tasks.

word's last syllable in French. Each pronounced nonword was hand-segmented in Praat 6.0.43 (Boersma & Weenink, 2018) and normalized for amplitude at 70 dB. The sixteen nonwords were associated with coloured pictures of alien-like cartoon images, sized 300 × 300 pixels on the display (Fig. 1).

Learning consisted of two phases. First, in the *passive* learning phase, participants saw each of the 16 images on a display for 1500 ms, then heard the corresponding auditory word twice separated by 1500 ms. The image then stayed on the screen for 2000 ms and disappeared, followed by a prompt to move on to the next trial. Participants completed passive exposure on day 1 before the active learning task, and on day 2 before the word recognition task. This task was conceived in order to introduce the words on day 1, letting participants know what kind of associations they would have to memorize later on. On day 2, we wanted to maximize the participants' recognition without further promoting active learning of the words, since the outcome of the active learning methodology was unknown prior to the experiment.

Following passive exposure on day 1, participants completed the *active* learning phase in which they chose which of two images on a display corresponded to an auditory stimulus, indicated via a button press. Images appeared on screen for 1500 ms followed by the auditory stimulus. For instance, in the [y]-[u] TRAINING condition, participants would see the images corresponding to [faky] and [faku] next to one another for 1500 ms, then would hear the audio clip corresponding to one of the images, e.g., [faky]. Participants were asked to indicate which image corresponded to the heard word (left vs. right) via a button box. Visual feedback was presented for 500 ms to indicate response accuracy (a green rectangle appeared around the correct picture, or a red rectangle around the incorrect picture). Trials were presented in 16-word blocks, with trials presented in pseudorandom order (i.e., the same pair of images had to be separated by at least one trial containing another pair) without repetition. Blocks repeated until listeners reached 14/16 correct responses within a block, requiring 10–40 min depending on their rate of learning. Performance was assessed by quantifying the number of blocks necessary to reach criterion. Of main interest was the extent to which the TRAINING condition influenced learning speed and word recognition in the subsequent task.

2.1.2.3. Word recognition task. Third, listeners all completed the word recognition task with the same items but novel auditory tokens. This task was completed immediately following the word learning task on day 1, and again on day 2 after the AX task and passive exposure (short word recall). Trial order was randomized on both days. For each experimental trial, participants saw a fixation point in the center of the display and had to fixate it for 200 ms. Then, the four 300 × 300-pixel images of one set were presented (i.e., rows in Fig. 1: e.g., [fakɑ], [fake], [faku] and [faky] together on the display), one in each corner of the screen and embedded within 350 × 350-pixel interest areas (invisible to the participant). Participants were free to look at the pictures for 1000 ms before an audio clip was played, corresponding to one of the images. Participants were instructed to push a button on a button box (that had one button in each corner) corresponding to the physical position of the chosen image on the display.

Eye tracking data was acquired during both the learning and recognition tasks with an EyeLink Portable Duo (SR Research) in remote mode using a chin rest, monocular recording. Listeners were placed such that their eyes were approximately 52 cm from the eye tracker camera ±5 cm. They completed a five-point calibration, then validation, keeping the maximum and average errors below 1° of visual angle for all participants, before moving on to the word recognition task. The tracker sampled at 500 Hz sample rate, but eye movements were resampled in 50-ms time bins, starting at auditory word onset, for analysis. Proportions of fixations to the target were calculated by dividing the number of samples within the interest area corresponding to the target image by the number of samples within all interest areas. Time bins that did not contain any fixations to any interest area on the screen were excluded from the analysis.

2.2. Results

All data and analysis codes are available in the OSF repository: https://osf.io/hp5sr/?view_only=94981621f55e4ea8ba9e2bbb361e808b. Statistical analysis of the button press and eye movement data was performed in R (R Core Team, 2019) with generalized additive mixed-effects models (GAMMs; Wood, 2017). GAMMs model the impact of parametric factors (i.e., fixed effects, in linear-models terminology) and can include random effects. They are equally appropriate for analyzing button presses, reaction times and eye tracking records. Models were built and visualized using the *mgcv* package version 1.8–31 (Wood, 2017) and *itsadug* package version 2.3 (van Rij et al., 2017) in R with the *gam()* function, and model comparisons to establish significance of factors were made with the *compareML()* function, starting with the maximal model and comparing it to simpler models throughout the procedure (i.e., backwards model fitting procedure). Inclusion of a factor yielding significantly better model fit warrants significance of the said factor (Porretta et al., 2016). For the current presentation, the final (i.e., better) model was refitted using the restricted maximum-likelihood method (REML; Porretta et al., 2018).

Among many advantages, GAMMs enable one to account for data without assuming normality of the distribution, since nonnormality is common in eye movement data, and considering autocorrelation for time series (i.e., one data point in a time series is necessarily correlated to the preceding point in the series; Baayen et al., 2018). GAMMs are also robust against missing cells, which can occur in eye movement data when listeners blink or look outside of the areas of interest on the display. These models have been recently and successfully used to analyze eye movement data for language processing (Desmeules-Trudel et al., 2020; Desmeules-Trudel & Zamuner, 2019, 2022; Porretta et al., 2018; van Rij et al., 2016).

2.2.1. Behavioural results

2.2.1.1. Speech discrimination task. For the speech discrimination task, we calculated d-prime scores for each participant to quantify a bias-free measure of sensitivity to all contrasts, using the *neuropsychology* package 0.5.0 in R (Makowski, 2016). We examined the impacts of DAY of testing, qualities of the VOWELS within a pair, and TRAINING group for [y]-[u] pairs. Fig. 2 shows d-prime scores for all *different* pairs per DAY of testing (Panel A) as well as d-prime scores for the [y]-[u] pairs (Panel B) across TRAINING groups and DAY of testing. Planned pairwise comparisons (Bonferroni-corrected *t*-tests) between paired VOWEL qualities importantly revealed, among many other significant differences (see OSF project), that d-prime scores for [y]-[u] pairs were significantly lower than all other pairs (range $t = -16.3$ to -9.68, $df = 47$, all $p < 0.001$). However, one-sample *t*-tests on d-prime scores for [y]-[u] pairs (Panel B) further showed that the scores were significantly above 0 on both day 1 ($t = 8.55$, $df = 23$, $p < 0.001$) and day 2 ($t = 6.78$, $df = 23$, $p < 0.001$), and thus suggest that listeners were sensitive to the target contrast.[4]

The GAMM analysis for all *different* pairs, which included random intercepts per participant, revealed that VOWEL qualities within the pairs significantly contributed to better model fit (comparison of model with DAY + VOWEL vs. DAY only: Δ in ML χ^2 score ($\Delta\ df = 5$) = 421.43, $p < 0.001$). Thus, the best model only included VOWEL qualities as a fixed effect and a random intercept of PARTICIPANT. Table 1 presents the model output (with VOWEL only as a fixed effect) and shows that the (log-likelihood) of responding correctly to [y]-[u] pairs was significantly lower than all other vowel pairs, as shown in Fig. 2A. This result is consistent with the prediction that English-native L2 listeners of French have difficulties discriminating /y/ and /u/.

Specifically for [y]-[u] pairs (Panel C in Fig. 2), neither DAY of testing nor TRAINING group significantly impacted the probability of giving a correct response. Thus neither of these factors improved or hindered their performance in the task for [y]-[u]. This suggests that potential improvement in word recognition abilities (i.e., on day 2) cannot be attributed to improved perceptual performance on the vowels themselves.

2.2.1.2. Word learning task. To assess learning speed during the training task (on day 1), we extracted the number of completed blocks during training for each participant (Fig. 3). Although [y]-[u] participants required more blocks ($M = 7.9$, $SD = 5.1$) to reach criterion than the [y]-Control group ($M = 6.1$, $SD = 3.4$), this difference was not statistically significant ($t = -0.986$, $df = 17.416$, $p = 0.338$).

2.2.1.3. Word recognition task. After training, listeners completed the word recognition task. Button-press accuracy data (correct responses were coded as 1 and incorrect responses were coded as 0) and reaction time data was trimmed by participant, based on reaction times that were longer than 7000 ms or shorter than 300 ms, and above or below 2.5 standard deviations of the participant's mean (8% data loss). The independent variables of interest were the TRAINING group and testing DAY, and random effects were flat intercepts for PARTICIPANTS and ITEMS.

Results suggest that all listeners had an overall higher probability of giving correct responses on day 2 than on day 1 (Fig. 4A). The GAMM analysis (Table 2) revealed that only DAY of testing was significant (comparison of model with DAY + TRAINING VS. TRAINING only: Δ in ML χ^2 score ($\Delta\ df = 1$) = 12.8, $p < 0.001$). This is consistent with the expectation that recall would be improved on day 2 in general.

We also analyzed (log-transformed) reaction times (RTs; Fig. 4B). Overall, RTs were shorter on day 2 than on day 1, and were shorter for the [y]-Control than for the [y]-[u] group overall. The GAMM analysis, exploring the impact of TRAINING group and DAY of testing on response speed, revealed that the interaction between the two factors of interest significantly contributed to better model fit (DAY*TRAINING VS. DAY + TRAINING; Δ in ML χ^2 score ($\Delta\ df = 1$) = 3.02, $p = 0.01$). Table 3 presents the GAMM output. In other words, the difference between listeners' RTs across day 1 and day 2 was greater in the [y]-[u] group than the [y]-Control group.

We further explored the impact of testing DAY and TRAINING group on

[4] Note that the outliers for all pairs (except for [u]-[e] pairs, for which two participants are outliers) originate from the same participant (for an illustration, see R code on OSF project, entitled *Discrimination_analysis.R*). It is possible that this participant generally performed poorly on the discrimination task (e.g., focusing on acoustic details rather than on vowel categories) or responded randomly during the task. Regardless, the significant pairwise comparisons demonstrate that overall group performance was significantly lower for [y]-[u] pairs than for all other pairs.

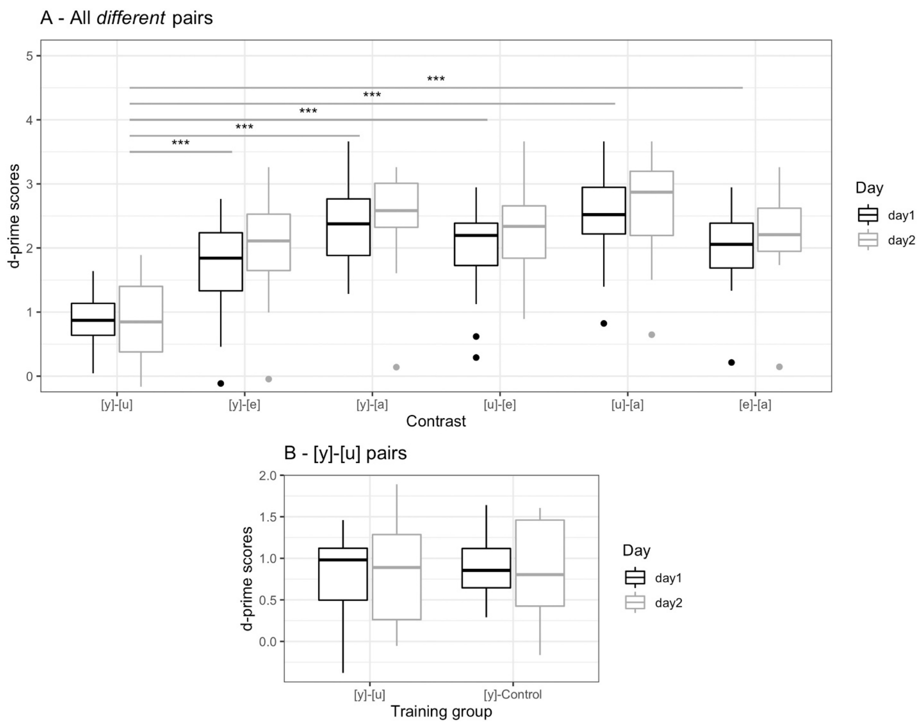

Fig. 2. D-prime scores (bias-free sensitivity to the contrasts) in the syllable discrimination (AX) task by VOWEL pair and DAY of testing (Panel A). D-prime scores by TRAINING group and DAY of testing for [y]-[u] pairs (Panel B). Box plots represent the median, 25th and 75th percentiles, lower and upper whiskers representing 1.5 times the interquartile range, and individual points considered outliers based on these criteria. Horizontal lines on Panel A indicate significant differences between [y]-[u] pairs and the other investigated vowel pairs, performance collapsed over the two testing days.

Table 1
GAMM output, accuracy analysis of the speech discrimination task.

Formula: response ~ vowel_pair + s(participant, bs = "re"), data = acc_discr, family = "binomial", method = "REML"

Parametric coefficients	Estimate	Std. error	*t*-Value	*p*-Value
Intercept	−0.206	0.192	−1.07	0.28
[y]-[e]	3.21	0.205	15.63	<0.001
[y]-[ɑ]	3.93	0.264	14.89	<0.001
[u]-[e]	3.579	0.232	15.42	<0.001
[u]-[ɑ]	3.623	0.236	15.37	<0.001
[ɑ]-[e]	3.344	0.215	15.6	<0.001
Smooth terms	**Edf**	**Ref.edf**	χ^2	***p*-Value**
Random (PARTICIPANT)	19.99	23	1	<0.001

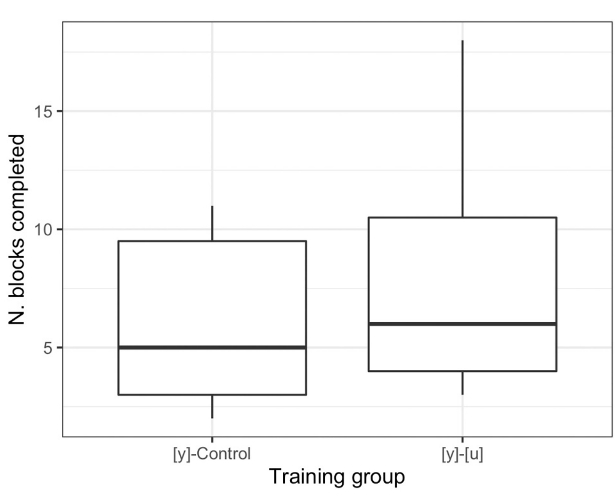

Fig. 3. Number of blocks completed by participants during the training phase by TRAINING group (in box plots representing the median, 25th and 75th percentiles, lower and upper whiskers representing 1.5 times the interquartile range, and individual points considered outliers based on these criteria).

[y] items only in a second set of model comparisons, and present in Fig. 5 the rates of [y] and [u] responses to [y] items. For theses analyses, when a participant was presented with a [y] auditory word, we compiled the number of times they gave [u] and [y] responses and submitted this data to our GAMM procedure. [y] responses were coded as 1 and [u] responses were coded as 0. Model comparisons revealed that neither DAY of testing nor TRAINING group significantly impacted the probability of giving a correct (i.e., [y]) response. This result can be due to the relatively low number of responses for [y] words (i.e., 681 responses recorded and analyzed).

2.2.2. Eye tracking results – word recognition task

The accuracy and reaction time data presented above are considered to reflect the end point of word processing, and therefore provide a general measure of processing difficulty from the onset of perception to button press. The eye movement measures in our Visual World Paradigm design can provide additional information by measuring recognition as it occurs. This can thus provide additional insights into real-time lexical access by reflecting online recognition, in what is termed the linking hypothesis (Allopenna et al., 1998; though see Teruya & Kapatsinski, 2019, for a critique).

With this idea in mind, proportions of fixations to the target images in the word recognition task were compiled every 50 ms and are plotted below in Figs. 6 and 7. An increase in proportions of fixations (y-axis) through time (x-axis) suggests that listeners fixated more to the target image during and after the word was heard, which in turn provides indications on how listeners interpret the auditory stimuli in relation to the images on the display. Using GAMMs, significance of the factors of interest through TIME was assessed by plotting difference curves (*plot_diff()* function) between two levels of a factor (e.g., DAY of testing and TRAINING group) on proportions of fixations to the target image, and observing the sections within the time window in which the confidence intervals do not overlap 0. Note that the time window of analysis started 200 ms after word onset, reflecting anticipated eye movement programming delay, until 3000 ms after word onset. The dependent variable was the empirical-logit-transformed proportions of fixations to the target (this transformation was necessary to make the data unbounded, see Porretta et al., 2018) for trials when participants gave a

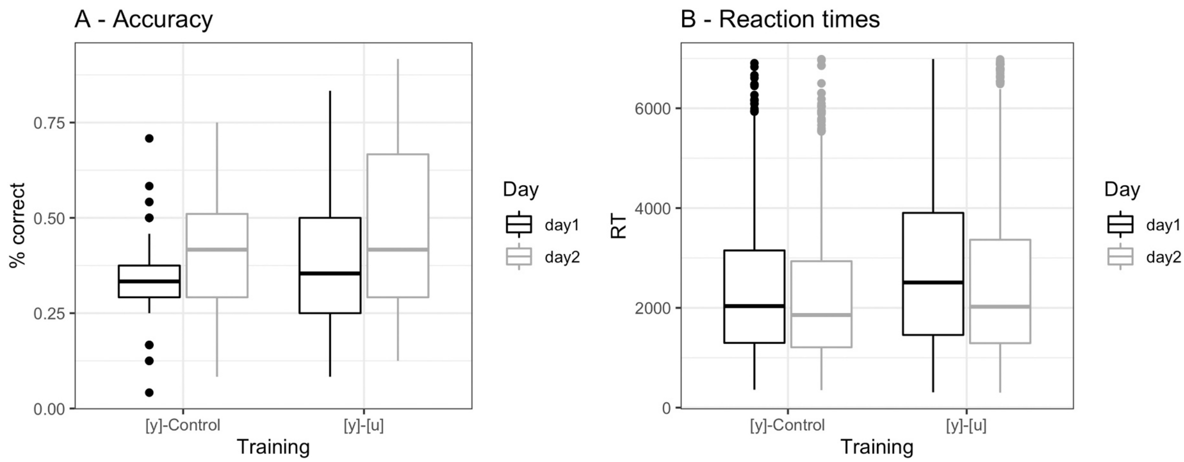

Fig. 4. Proportions of correct responses (in box plots representing the median, 25th and 75th percentiles, lower and upper whiskers representing 1.5 times the interquartile range, and individual points considered outliers based on these criteria) in the word recognition task by TRAINING group and DAY of testing (Panel A), and reaction times (Panel B) by TRAINING group and DAY of testing.

Table 2
GAMM output, accuracy analysis of the Visual World Paradigm task.

Formula: response ~ day + s(participant, bs = "re") + s(target, bs = "re"), data = acc_vwp, family = "binomial", method = "REML"

Parametric coefficients	Estimate	Std. error	*t*-Value	*p*-Value
Intercept	−0.414	0.188	−2.2	0.028
DAY 2	0.346	0.069	5.03	<0.001
Smooth terms	**Edf**	**Ref.edf**	***F*-value**	***p*-Value**
Random (PARTICIPANT)	18.92	21	215	<0.001
Random (TARGET IMAGE)	14.24	15	295.4	<0.001

Table 3
GAMM output, RT analysis of the Visual World Paradigm task.

Formula: logRT ~ training*day + s(participant, bs = "re") + s(target, bs = "re"), data = acc_vwp, method = "REML"

Parametric coefficients	Estimate	Std. error	*t*-Value	*p*-Value
Intercept	7.594	0.01	76.13	<0.001
[y]-[u]	0.184	0.139	1.33	0.185
DAY 2	−0.042	0.024	−1.71	0.087
[y]-[u]:DAY 2	−0.086	0.035	−2.46	0.014
Smooth terms	**Edf**	**Ref.edf**	***F*-value**	***p*-Value**
Random (PARTICIPANT)	19.68	20	61.7	<0.001
Random (TARGET IMAGE)	12.26	15	4.73	<0.001

correct response. We used the *bam()* function which is optimized for large data sets. Random nonlinear smooth components for PARTICIPANT and TARGET IMAGE through TIME, which are similar to random slopes but can nonlinearly vary over time, were included in the models. An autocorrelation correction factor of 0.89 was computed based on the data with the initial model and included in all subsequent models.

The fixations to the target on day 1 and day 2 are shown in Fig. 6A collapsed across TRAINING groups and items. Statistical analysis revealed that DAY of testing was significant (comparison of model with DAY + TRAINING VS. TRAINING only: Δ in ML χ^2 score ($\Delta\ df = 5$) = 13.9, $p < 0.001$), but not TRAINING. The best model is presented in Table 4. The difference curve between fixations on day 2 – day 1 (Fig. 6B) was computed following the GAMM analysis (the significant difference is found between dashed line on Fig. 6B). Together, data observation and GAMM analyses show that from 480 ms onwards, listeners fixated significantly more to the targets on day 2 than on day 1, regardless of TRAINING group.

We also analyzed proportions of fixations to images that corresponded to [y]-words on trials when participants gave correct [y] responses (Fig. 7A) and incorrect [u] responses (Fig. 7B). We found that a GAMM containing a parametric interaction between DAY and TRAINING had significantly better fit than one without the interaction (DAY*TRAINING VS. DAY + TRAINING Δ in ML χ^2 score ($\Delta\ df = 1$) = 3.54, $p = 0.008$), but also that individual factors did not significantly contribute to better model fit (comparison of model with DAY*TRAINING VS. DAY only: Δ in ML χ^2 score ($\Delta\ df = 6$) = 3.69, $p = 0.29$; comparison of model with DAY*TRAINING VS. TRAINING only: Δ in ML χ^2 score ($\Delta\ df = 6$) = 4.25, $p = 0.2$). This suggests that fixations to the [y] target when listeners gave correct responses

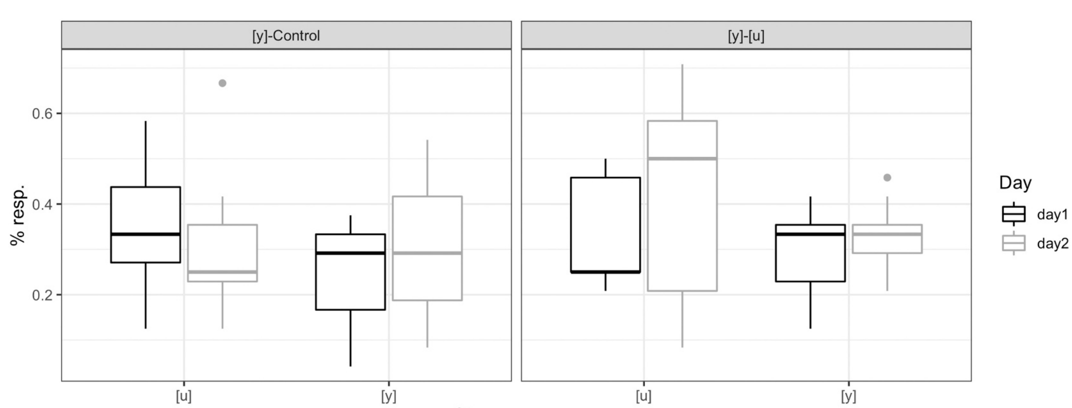

Fig. 5. Proportions of [u] and [y] responses in [y] trials by TRAINING group ([y]-Control in black, [y]-[u] in grey) and DAY of testing.

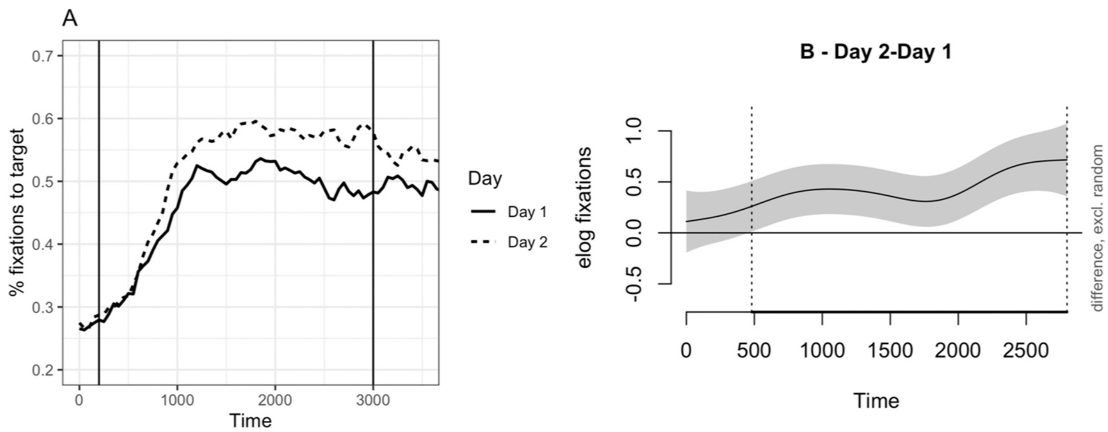

Fig. 6. Raw fixations (%) to the target on day 1 (solid line) and day 2 (broken line) in correct-response trials (Panel A) and difference curve for between day 2 and day 1 (Panel B). On the right panel, solid lines represent the average difference curves, shaded areas represent the confidence intervals around the mean, and areas between dotted lines represent the portions of the analysis window for which the two levels of the factor are significantly different.

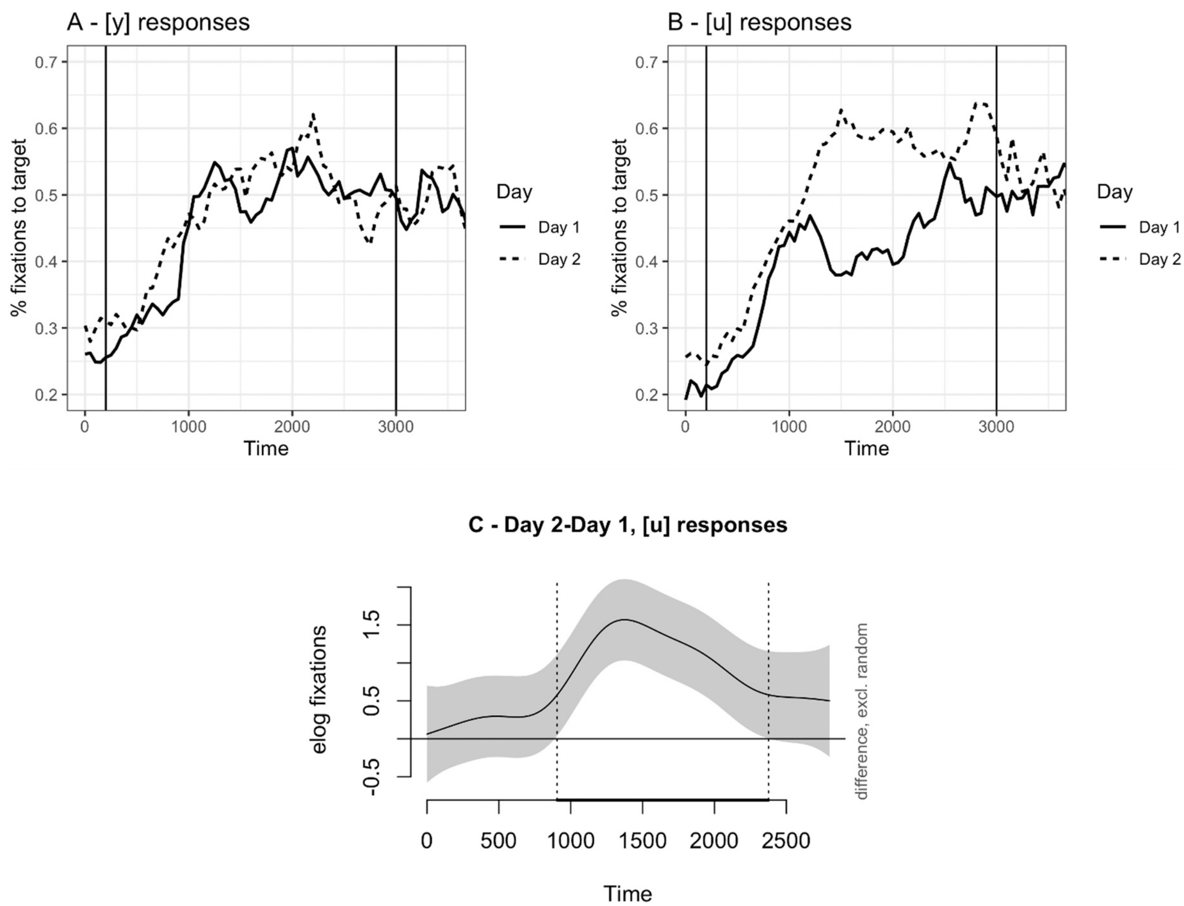

Fig. 7. Raw fixations (%) to the target on day 1 (solid lines) and day 2 (broken lines) in correct-response trials (Panel A) and incorrect-response (i.e., [u]) trials (Panel B). Difference curve between day 2 and day 1 for incorrect responses, as outputted by a GAMM (Panel C).

were not impacted by DAY of testing or TRAINING.

For [u] responses, we found that DAY of testing significantly contributed to model fit (Fig. 7C; comparison of model with DAY*TRAINING VS. TRAINING only: Δ in ML χ^2 score ($\Delta\ df = 5$) = 10.02, $p < 0.001$). Model output is presented in Table 5. This means that on day 2, all listeners fixated more to the [u] image when misinterpreting [y] auditory words as tokens of [u] words, thus reflecting listeners' tendency to phonetically assimilate [y] sounds to their [u] category, especially after consolidation. This tendency was not different across TRAINING groups.

3. Discussion

We investigated speech discrimination and word learning abilities in a simulated second-language learning task with English-speaking participants, using French-sounding stimuli. Of interest were the impact of training structure, i.e., practice with discriminating similar-sounding words for English native listeners. We also used of eyetracking to more accurately measure perception as it unfolds in real-time, which we expected would provide more detailed information about learning

Table 4
GAMM output, eye movement analysis of the Visual World Paradigm task (all targets).

Formula: elog ~ day + s(time) + s(time, by = day) + s(participant, time, bs = "fs") + s (target, time, bs = "fs"), data = eye_target, method = "REML", AR.start = eye_target $start.event, rho = AR1

Parametric coefficients	Estimate	Std. error	*t*-Value	*p*-Value
Intercept	−0.432	0.156	−2.77	<0.01
DAY 2	0.384	0.084	4.55	<0.001
Smooth terms	**Edf**	**Ref.edf**	***F*-value**	***p*-Value**
S(TIME)	5.716	6.263	3.33	<0.01
S(TIME):DAY 1	3.094	3.816	0.43	0.802
S(TIME):DAY 2	1.85	2.471	0.25	0.758
S(PARTICIPANT, TIME)	118.684	197	2.23	<0.001
S(TARGET, TIME)	89.494	143	2.11	<0.001

Table 5
GAMM output, eye movement analysis of the Visual World Paradigm task ([y] targets only, [u] responses).

Formula: elog ~ day + s(time) + s(time, by = day) + s(participant, time, bs = "fs") + s (target, time, bs = "fs"), data = eye_ytarget_uresp, method = "REML", AR.start = eye_target$start.event, rho = AR1

Parametric coefficients	Estimate	Std. error	*t*-Value	*p*-Value
Intercept	−0.879	0.21	−4.19	<0.001
DAY 2	0.772	0.173	4.46	<0.001
Smooth terms	**Edf**	**Ref.edf**	***F*-value**	***p*-Value**
S(TIME)	1.095	1.116	0.77	0.401
S(TIME):DAY 1	4.666	5.787	2.02	0.071
S(TIME):DAY 2	4.734	5.74	2.76	<0.05
S(PARTICIPANT, TIME)	86.433	197	0.9	<0.001
S(TARGET, TIME)	11.091	35	0.56	<0.01

outcomes compared to traditional overt recognition memory accuracy or reaction times. Further, participants also performed a follow-up session on a second day, to assess prior findings that novel word learning is potentially affected by memory consolidation. We first examined how TRAINING structure might promote discrimination of a non-native speech contrast. In the [y]-[u] learning condition, listeners were systematically presented with pairs of nonwords containing French [y]-[u], which are expected to be assimilated to the same category, i.e., English [u], and French [ɑ]-[e] nonwords that are also phonetically distinct in English and therefore unlikely to be assimilated. In the [y]-Control group, listeners were only presented with non-assimilable word pairs ([y]-[e] and [u]-[ɑ]). To the best of our knowledge, little evidence has been raised regarding consolidation in word learning as it pertains specifically to L2 learning that involves both unfamiliar words and phonemes.

Listeners thus completed both the phonetic discrimination and word recognition tasks on a first day of testing, and another time on a second day of testing. We predicted no effect of DAY for speech discrimination given previous findings that variability in stimuli (i.e., multiple voices) hinders improvement on speech discrimination after sleep (Earle & Myers, 2015a) and that speech discrimination relies on procedural memory processes that require extended exposure to stimuli in order to be consolidated (Ullman & Lovelett, 2018). Recall that Earle and Myers (2015a) found no effect of (sleep-related) consolidation on a discrimination task using variable phonetic input. In addition, past research has shown that training on phoneme discrimination does not necessarily have a positive effect on word recognition abilities (e.g., also across dialects, Dufour et al., 2010).

Results of the speech discrimination task showed that all participants were sensitive to the French [y]-[u] contrast, although significantly less than all other contrasts, but that neither DAY of testing nor the structure of TRAINING structure impacted performance. This is consistent with the idea that phoneme discrimination abilities are different from lexical recognition skills. That is, successful learning of words that have the unfamiliar [y] vowel (e.g., as seen via an increase in fixations to the target in [y] trials within the time window of analysis, Fig. 7A) did not generalize to improved performance in a speech discrimination task, consistent with the PRIMIR model's prediction (Curtin et al., 2011; Werker & Curtin, 2005) that phonemic discrimination learning occurs semi-independently from word-level learning. This explanation is also consistent with Darcy et al.'s (2012) results on a single-day speech perception study. Theoretically speaking, speech perception and word learning tap into distinct memory systems, with sound discrimination relying on procedural memory, and vocabulary acquisition engaging declarative memory. Ullman (2016) suggests that a procedural memory system is slower to be consolidated and requires repetition, as opposed to the declarative-memory system which is faster to consolidate. Note that given that testing sessions were separated by a relatively short period (16–21 h), it is possible individuals did not receive enough repetitions of the stimuli over multiple days to attain a consolidated representation, and thus improve performance on the more-procedural task (for a review on L2 grammatical rules and procedural memory, see Ullman & Lovelett, 2018). Further research on the relationship between speech perception abilities and word recognition will shed light on the exact mechanisms behind learning of a variety of (interacting) aspects of L2s.

For the word recognition task, we predicted that TRAINING in the [y]-[u] group would facilitate word recognition on day 1 and day 2, due to practice in inhibiting lexical competition when trained on similar lexical items (Kapnoula & McMurray, 2016). We also predicted overall improvement on day 2 based on past research on memory consolidation of linguistic information, a robust effect that has been found in the native language (Davis et al., 2009; Dumay & Gaskell, 2007). The analysis of accuracy data did not show a significant effect of TRAINING, but did show a clear effect of DAY of testing, i.e., participants gave significantly more correct responses on day 2 than on day 1. Furthermore, both TRAINING group and DAY of testing had an impact on reaction times (RTs) for this task, showing that RTs were significantly shorter on day 2 than day 1, and the difference between RTs on day 1 and day 2 was greater in the [y]-[u] group. Taken together, these results show that DAY had a significant impact on recognition of newly-learned words, with perhaps a more modest contribution of TRAINING group.

Other elements that could have contributed to improved performance on the second day of testing for the word recognition task include the fact that listeners were administered a quick (passive) recall of the words prior to the task on day 2. As mentioned above, due to the novelty of our paradigm and our aim to maximize the possibility of recognizing L2-sounding nonwords on day 2, we included such passive recall of the words. Unfortunately, that created a potential confound for assessing memory consolidation (although see below for an interpretation of the observed group differences on RT results). Since we cannot disentangle these factors when interpreting the improvements we found on day 2 of testing, future investigations will need to more closely isolate the effect of memory consolidation in L2 word learning. Given previous evidence that better word recognition and processing abilities are reported in the memory consolidation literature in an L1, we would predict that one of the forces behind the observed patterns (i.e., better recognition on day 2) is consolidation of these newly learned words.

Likewise, while all participants reported 7 h of sleep on average prior to attending the day 2 trial, we did not also include a no-sleep condition in this study, where some participants were re-tested following a similar delay but without intervening sleep. Thus, we leave open the possibility that sleep itself had no effect on learning and it remains possible that improvements are strictly due to the mere passage of time.

The observed effect of TRAINING on new-word processing RTs is a new finding in L2 research, consistent with previous literature on short-term impacts of training for lexical tasks (Kapnoula & McMurray, 2016). This interaction should be interpreted with care, since response times in the more challenging [y]-[u] condition were greater on day 1 than those of the [y]-Control condition. However, the absence of significant difference between the two groups on day 1 suggest that equivalent response times across groups on day 2 are possibly linked to consolidation effects, considering all other factors were equal including the re-familiarization of stimuli at the start of the second session.

Focusing more closely on [y] targets, which were of main interest in our study, no difference was observed between the rate of [u] and [y] responses, showing that the two word sets were not well discriminated. Thus, we cannot confirm the prediction that TRAINING (e.g., [y]-[u], in which listeners are trained on [y]-[u] pairs) improved learning or subsequent recognition of the learned words based on these data alone, but indications of this were found in the analyses of RT data.

Eye movement data was concordant with the behavioural data, showing a significant effect of DAY of testing which occurred early within trials (680 ms after word onset, considering the eye movement planning delay). The effect of TRAINING group was absent overall and on more fine-grained analyses, but we also found an effect of DAY of testing when listeners listened to [y] target but gave an incorrect [u] response. This result can be explained by the tendency of L2 listeners to assimilate sounds that are absent from their native language, in this case French front-rounded [y], to a pre-existing category, i.e., [u]. However, the significant increase in fixations to the incorrect target on day 2, as compared to day 1, shows that the wrong association between auditory word and image is strengthened, and could pose problems for L2 word learning. Overall, our results suggest that the demonstrated effect of consolidation in L1s also applies to an L2 (Ullman & Lovelett, 2018) and provides further information on the importance of (sleep-related) memory consolidation for learning.

The relatively weak differences across TRAINING groups can be explained by the fact that during training, manipulations on the simultaneous presentation of images ([y]-[u] and [ɑ]-[e] in the [y]-[u] group, and [y]-[e] and [ɑ]-[u] in the [y]-Control group) mainly involved visual recognition. In order to discriminate the images, listeners needed to encode the association between the [y] auditory words and corresponding images. However, they did not have to *auditorily discriminate* [y] and [u] words, therefore it might be difficult for them to establish new auditory-[y] representations. Consequently, it is possible that a learning task that involved discriminating competing auditory word pairs during might result in better representation of the auditory information by more strongly emphasizing this unfamiliar phonetic contrast. For example, presenting one image with two auditory stimuli and asking which one matches the picture could favour the establishment of a stronger [y]-sound representation during learning, and in turn increase recognition performance. Using further measures of lexical processing, such as electroencephalography, could also be highly beneficial to uncovering many aspects of the interaction between TRAINING, phonetic assimilation, and word learning.

In conclusion, we found that newly-learned words were significantly better recognized on a second day of testing in an L2 across all our measures, and more modestly impacted by the structure of learning (i.e., how closely the task promotes discrimination of assimilable sounds). The main finding suggests that the passage of time, potentially combined with the amount of exposure received across two sessions, had a significant impact on word learning but not sound discrimination. There was some evidence for an advantage when L2 learners focused on discriminating assimilable words (e.g., in RT data). Thus, an additional contribution of this work is the possible role of memory consolidation in second-language learning and processing, and how this might complements theories on the interplay between native-language phonetics and word recognition. Investigations with a variety of linguistic materials, such as different sounds, more complex words, or sentences, will also shed light on the role that consolidation processes play in learning, providing a step towards understanding the interplay of language-specific (e.g., phonetic assimilation) and domain-general (e.g., memory) factors in how adults learn a second language.

Declaration of competing of interest

We declare no conflict of interest for the project, reported in the manuscript entitled "**Learning unfamiliar words and perceiving non-native vowels in a second language: Insights from eye tracking**".
Félix Desmeules-Trudel and Marc F. Joanisse.

Acknowledgements

We are grateful to Krystal Flemming, Alyssa Yantsis, and members of the *Language, Reading and Cognitive Neuroscience Lab* at the University of Western Ontario for their significant contributions to the current project, as well as audiences at the *Canadian Linguistics Association Annual Meeting 2019*, *LabPhon 2020* for insightful comments. FDT benefited from funding from the *Fonds de recherche du Québec – Société et Culture* (postdoctoral fellowship) and from the *Canada First Research Excellence Funds* (*BrainsCAN* program at the University of Western Ontario; postdoctoral associate top-up), and MFJ benefited from a *Natural Sciences and Engineering Research Council of Canada* Discovery Grant for this research. Infrastructure support was provided by the *Canada Foundation for Innovation*.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.
Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-effect regression models in linguistics* (pp. 49–69). Cham: Springer.
Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language, 89*, 23–36.
Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27*, 387–414.
Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, NL: John Benjamins.
Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer software]. *Retrieved from*. http://www.fon.hum.uva.nl/praat/.
Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39*, 492–504.
Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269–284.
Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5–6), 507–534.
Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1-english-L2 french acquisition. *Second Language Research, 28*(1), 5–40.
Davis, M. H., Di Betta, A. M., MacDonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803–820.
Desmeules-Trudel, F., & Joanisse, M. F. (2020). Discrimination of four Canadian-French vowels by Canadian-English listeners. *The Journal of the Acoustical Society of America, 147*(5), EL391–EL395.
Desmeules-Trudel, F., Moore, C. E., & Zamuner, T. S. (2020). Monolingual and bilingual children's processing of coarticulation cues during spoken word recognition. *Journal of Child Language, 47*(6), 1189–1206.
Desmeules-Trudel, F., & Zamuner, T. S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics, 81*(5), 1654–1672.
Desmeules-Trudel, F., & Zamuner, T. S. (2022). Spoken word recognition in a second language: The importance of phonetic details. In *Second Language Research* (Online First, pp. 1–30).
Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2010). Does training on a phonemic contrast absent in the listener's dialect influence word recognition? *Journal of the Acoustical Society of America Express Letters, 128*(1), EL43–EL48.
Dumay, N., & Gaskell, G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39.
Earle, F. S., & Myers, E. B. (2015a). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America, 137*(1), EL91-EL97.

Earle, F. S., & Myers, E. B. (2015b). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1680–1695.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America, 119*(4), 1950–1953.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetic, 36*(2), 345–360.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47–65.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76*(3), 708–721.

Fuhrmeister, P. (2019). Interference in memory consolidation of non-native speech sounds. In S. Fuchs, J. Cleland, A. Rochet-Capellan, & P. Lang (Eds.), *Speech Production and Perception: Learning and Memory* (pp. 207–243).

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *Journal of the Acoustical Society of America Express Letters, 142*(5), EL448–EL454.

Fuhrmeister, P., & Myers, E. B. (2020). Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, & Psychophysics, 82*(4), 2049–2065.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12*(2), 91–114.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171.

Kapnoula, E. C., & McMurray, B. (2016). Training alters the resolution of lexical interference: Evidence for plasticity of competition and inhibition. *Journal of Experimental Psychology: General, 145*(1), 8–30.

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.

Makowski, D. (2016). *Package 'neuropsychology': An R toolbox for psychologists, neuropsychologists, and neuroscientists [R package]*. Retrieved from https://github.com/neuropsychology/neuropsychology.R.

Porretta, V., Kyröläinen, van Rij, J., & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent decision technologies 2017: Proceedings of the 9th KES international conference on intelligent decision technologies – Part II* (pp. 268–277). Springer International Publishing.

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listeners experience on word recognition. *Journal of Phonetics, 58*, 1–21.

R Core Team. (2019). *The R project for statistical computing [computer software]*. Retrieved from https://www.r-project.org.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

Teruya, H., & Kapatsinski, V. (2019). Deciding to look: Revisiting the linking hypothesis for spoken word recognition in the visual world. *Language, Cognition and Neuroscience, 34*(7), 861–880.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–968). Academic Press.

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research, 34*(1), 39–65.

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler, & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 267–293). Berlin: De Gruyter.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting time series and aucorrelated data using GAMMs [R package]*. Retrieved from https://rdrr.io/cran/itsadug/man/itsadug.html.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1–25.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866–1878.

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton: Chapman and Hall/RC.