Original Article

second language research

Second Language Research
1–22

DOI: 10.1177/02676583231181472
journals.sagepub.com/home/slr
Sage

Using allophonic variation in L2 word recognition: French listeners' processing of English vowel nasalization

Megan M Dailey
LSCP, Ecole Normale Supérieure – Paris Sciences et Lettres (ENS – PSL), Centre National de la Recherche Scientifique (CNRS), Ecole des Hautes Etudes en Sciences Sociales (EHESS), France

Camille Straboni
Ecole Normale Supérieure – Paris Sciences et Lettres (ENS – PSL), France

Sharon Peperkamp
LSCP, Ecole Normale Supérieure – Paris Sciences et Lettres (ENS – PSL), Centre National de la Recherche Scientifique (CNRS), Ecole des Hautes Etudes en Sciences Sociales (EHESS), France

Abstract
During spoken word processing, native (L1) listeners use allophonic variation to predictively rule out word competitors and speed up word recognition. There is some evidence that second language (L2) learners develop an awareness of allophonic distributions in their L2, but whether they use their knowledge to facilitate word recognition online, like native listeners do, is largely unknown. In an offline gating experiment and an online eye-tracking experiment in the visual world paradigm, we compare advanced French learners of English and a control group of L1 English listeners on their processing of English vowel nasalization during spoken word recognition. In the gating task, the French listeners' performance did not differ from that of the English ones. The eye-tracking results show that French listeners used the allophonic distribution in the same way as English listeners, although they were not as fast. Together, these results reveal that L2 learners can develop novel processing strategies using sounds in allophonic distribution to facilitate spoken word recognition.

Keywords
allophony, eye-tracking, gating, second language, speech perception, spoken word recognition, vowel nasalization

Corresponding author:
Megan M Dailey, Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, Paris, 75005, France
Email: megan.dailey@ens.psl.eu

I Introduction

Word recognition in connected speech is all but straightforward: there are no consistent acoustic markers of word boundaries, word forms are variably realized depending on surrounding phonological context, and there is additional variability along dialectal and individual talker axes. A lot of research has focused on how listeners deal with this variability, both in their native language (L1) – where word recognition is usually effortless – and in a second language (L2) (for a review, see Cutler, 2012). As for L2 listeners, some of this research has focused on how they deal with phonological variation, with mixed results. Advanced learners can range from native or near-native performance to failure to recover the intended word (Brand and Ernestus, 2018; Darcy et al., 2007; Gustafson and Bradlow, 2016; Tremblay, 2011; Tremblay and Spinelli, 2014; Tuinman and Mitterer, 2012; Tuinman et al., 2012).

Not all phonological variation, though, is harmful for word recognition. One example is allophony: two sounds whose distributions are complementary and conditioned by the following sound allow the listener to reduce the set of possible word candidates before perceiving that sound. A well-studied case is that of English vowel nasalization before nasal consonants.[1] Native English listeners are sensitive to the distribution of oral and nasal vowels: when a vowel preceding a nasal consonant lacks nasalization or a vowel preceding an oral consonant is inappropriately nasalized due to cross-splicing, they are slower to identify the consonant (Fowler and Brown, 2000). That English listeners use their knowledge of the allophonic distribution of oral and nasal vowels during word recognition is shown in several studies. For instance, Warren and Marslen-Wilson (1987) similarly used cross-splicing to create versions of minimal pairs such as *drought–drown*, such that in each word either the vowel or the final consonant (but not both) was nasal. In a gating paradigm, they found that participants needed larger fragments to recognize the correct words in the cross-spliced than in the original items. Lahiri and Marslen-Wilson (1991) likewise used a gating paradigm but with only natural stimuli; they found that for words containing a nasal vowel, English listeners increasingly guessed words with upcoming nasal consonants and ruled out words with an oral consonant before hearing the beginning of the nasal consonant. Similar results were obtained by Beddor et al. (2013) in an eye-tracking experiment: when presented with a word with a nasalized vowel (e.g. *scent*), participants fixated the target image before the vowel offset when the competitor contained no nasal consonant (e.g. *set*).

Native listeners have also been shown to rapidly integrate novel allophonic variation when processing words spoken in another dialect. For instance, Dahan et al. (2008) presented English listeners with words spoken in a different dialect of English than their own. Specifically, in the speaker's dialect, /æ/ is raised towards /ɛ/ before voiced /g/, but not before voiceless /k/. In an eye-tracking task, targets and competitors were pairs such as *bag* and *back*. It was found that the prior presence of /g/-final target words facilitated recognition of /k/-final target words. Thus, participants had adapted to the fact that when listening to this speaker, /g/-final candidate words can be ruled out as soon as an unraised /æ/ is being heard.

In the present study, we focus on L2 listeners. Whether these listeners, like L1 listeners, also rely on allophonic variation to rule out candidate words during word recognition has been the topic of some previous research. Three studies have investigated L2 learners' sensitivity to two allophonic word boundary cues in English: aspiration, which distinguishes, for instance, *keeps talking* from *keep stalking*, and glottal stop insertion, which distinguishes, for instance, *an ice man* from *a nice man*. Using a 2AFC (two-alternative forced-choice) task with minimal word pairs like the ones above, these studies focused on learners of English whose native languages were Spanish (Altenberg, 2005), Japanese (Ito and Strange, 2009), and French (Shoemaker, 2014), respectively. In Japanese, word-initial stops are weakly aspirated, and glottal stops can be inserted before an utterance-, phrase- or even word-initial vowel. In Spanish and French, by contrast, stops are never aspirated but a glottal stop can occur before a word-initial vowel in emphatic speech. All three studies found that the performance of L2 listeners was worse than that of native English controls, yet better than chance, and better for the glottal stop cue than for the aspiration cue. It is noteworthy to mention that while the Japanese participants had overall higher accuracy scores than the other two L2 groups, this difference was much larger for the aspiration than the glottal stop cue, suggesting that the Japanese listeners benefitted from the presence of weakly aspirated stops in their L1. Yet, their higher scores on the glottal stop cue suggests that L1 transfer is not the only factor driving performance.

Shea and Curtin (2010) investigated the perception of allophonic spirantization in Spanish by English-native learners of this language. In Spanish, the fricatives [β, ð, ɣ] are allophones of the stops /b, d, g/; they occur at the onset of unstressed syllables. In English, [β] and [ɣ] are not attested, while [ð] occurs phonemically and hence contrasts with [d]. In a stress perception test of Spanish non-words, including both legal ones (e.g. [ˈbɑβɑ], [βɑˈbɑ]: fricative in the unstressed syllable) and illegal ones (e.g. [ˈβɑbɑ], [bɑˈβɑ]: fricative in the stressed syllable), high-intermediate learners showed similar (although not identical) behavior to native Spanish listeners in that they were more likely to perceive syllables with fricative onsets as unstressed, even if those syllables carried stress. These results suggest that the participants had learned the allophonic stop–fricative distribution and used it in determining word stress in Spanish. Note that no separate analyses were carried out for items with [ð] and items with [β] or [ɣ]. It is therefore uncertain whether these participants were equally sensitive to the distribution of the two L2 allophones that are absent from English, i.e. [β] and [ɣ], as to that of the third one [ð], which in English stands in phonemic contrast with [d].

While the previous studies have shown that L2 listeners are sensitive to allophonic variation and can more or less successfully rely on it to rule out word candidates in offline tasks, no study has yet examined this capacity during online word recognition. As spoken word processing in L2 is more challenging and generally delayed compared to in the native language (Weber and Broersma, 2012), L2 listeners may have difficulty relying on an allophonic distribution in an online task even if they can use it in an offline task. The present research aims to shed light on this question.

Our case study concerns the processing of English vowel nasalization by speakers of Northern Metropolitan French (henceforth: French) that are self-reported advanced learners of English. In addition to 11 oral vowels, French has three phonemic nasal vowels, /ɛ̃/, /ɑ̃/ and /ɔ̃/. Sequences of a nasal vowel followed by a nasal consonant, though, only occur when the two are separated by a morpheme boundary. (This is due to the fact that nasal vowels historically derive from sequences of an oral vowel followed by a nasal consonant, with the nasal consonant being dropped; Sampson, 1999.) Consequently, while nasal vowels are necessarily followed by a nasal consonant in English, they are most often followed by an oral consonant in French, whereas oral vowels can only be followed by an oral consonant in English but by either an oral or a nasal consonant in French. Furthermore, nasal coarticulation in French is largely progressive, with regressive coarticulation being restricted to high vowels (Dow, 2020). Thus, the distributions and acoustic properties of nasal and oral vowels in French differ from those in English.

Despite these differences, there are several reasons why, compared to the L2 cases reviewed above, it would seem less challenging for French listeners to learn the complementary distribution of oral and nasal vowels in English and exploit it during word recognition. First, the English nasal–oral vowel distinction is an acoustically salient contrast (Beddor, 1993). Second, the token frequency of nasal vowels in French is high; they occur for instance in *dans* [dɑ̃] 'in', *en* [ɑ̃] 'of', *on* [ɔ̃] 'we', and *bien* [bjɛ̃] 'well'. Therefore, French listeners are highly sensitive to vowel nasality. Third, English vowel nasalization is phonetically natural, as it originates in gestural overlap; that is, lowering of the velum for nasal consonants necessarily begins during the preceding vowel.

In two experiments, we use an offline gating paradigm and an online eye-tracking paradigm, respectively, to examine whether French learners of English use the allophonic nasal–oral vowel distinction to speed up word recognition in English. In both experiments, we compare a group of advanced French learners of English with a control group of L1 English listeners.

II Experiment 1

In this experiment, we test a group of French advanced learners of English and a control group of L1 English listeners on their use of vowel nasalization during word processing in an offline gating paradigm. In order to avoid response biases due to lexical statistics (i.e. English has more words with a postvocalic oral than a postvocalic nasal consonant), we use a 2AFC task.

In the gating task literature, it is common to distinguish the isolation point of each word – the point at which responses become consistently correct – from its hypothesized recognition point – when listeners are confident they have identified the correct word (Grosjean, 1980). Once a listener has isolated a word, they are not yet expected to be confident that they have recognized the word, even though their behavioral responses indicate that they have distinguished it from its competitors. Confidence increases as more and more of the word is presented, and a listener reaches the recognition point when their reported confidence level, determined from a confidence rating provided at each stimulus presentation, reaches a certain threshold and does not dip below this threshold – in other words, when they are both consistently correct and consistently confident. In this experiment we follow the same reasoning, measuring both categorical responses and confidence ratings at each presentation to determine both isolation and recognition points for each item.

1 Methods

a Materials. Test items were 20 pairs of two- and three-syllable stress-initial nouns of American English (AmE). In each pair, words were phonemically identical until the onset of the consonant following the first, critical, vowel, which was a nasal consonant for one word and an oral one for the other word (e.g. *bunny–butter; grandfather–grasshopper*). Nasal and oral words did not differ in log frequency values taken from the CELEX corpus of AmE (Baayen et al., 1995) ($t < 1$).

In addition to the nasal–oral test pairs, there were 20 distractor pairs (i.e. *sofa–soda*) that were similar in structure to the test pairs: all consisted of nouns that were phonemically identical until a disambiguating consonant after the critical vowel. Distractor pairs did not differ in log frequency from test pairs ($t < 1$). Performance on these distractor pairs was not analyzed. For a full list of test and distractor items, consult Appendix 1.

Audio files for all test items and for one of each distractor pair were created by cutting the words from carrier sentences of the type 'Click on the X', where X represents the word.[2] Sentences were recorded by a male native speaker of AmE. Gating sets for each word were constructed following the procedure in Warren and Marslen-Wilson (1988): First, an alignment point, or zero gate, was marked at first vowel offset (Figure 1). Then, gates were marked from this alignment point every 20 milliseconds backwards through the vowel, marked −1, −2, etc., until its onset, and forwards through the following consonant, marked 1, 2, etc., until its offset. Stimuli were then cut from the original audio file containing the whole word, starting from word onset and until each successive gate, respectively, to create one stimulus presentation set. The resulting sets varied in size ranging from 5 to 13 stimuli.

For example, the first presentation of the word *bandage* (Figure 1) started at word onset and ended at gate −6, the second presentation started at word onset and ended 20 milliseconds later at gate −5, and so on. Stimuli were edited in Praat (Boersma and Weenink, 2019). Most nasal–oral pairs had an equivalent number of gates preceding the alignment point at first vowel offset (the area of focus for analysis); for those that did not the difference was a single gate. Appendix 1 lists the number of gates before the alignment point for each word. All sound files can be found online at the dedicated Open Science Framework (OSF) repository.

b Participants. The test group consisted of 16 L1 French self-reported advanced learners of English (age range: 21–39 years, median age: 24 years), living in France, who had started learning English in the French educational system in middle school. Their mean self-reported English verbal comprehension score was 7.7 on a scale from 1 (poor) to 10 (outstanding) ($SD = 1.09$). Seven participants reported having lived abroad in an English-speaking region, for on average 13 months total. Two additional French listeners were recruited but excluded from analyses as they did not fully understand the verbal and/or written instructions.

The control group consisted of 19 L1 AmE listeners (age range: 18–27 years, median age: 23 years) living in the United States. Henceforth, we refer to them as the English participants.

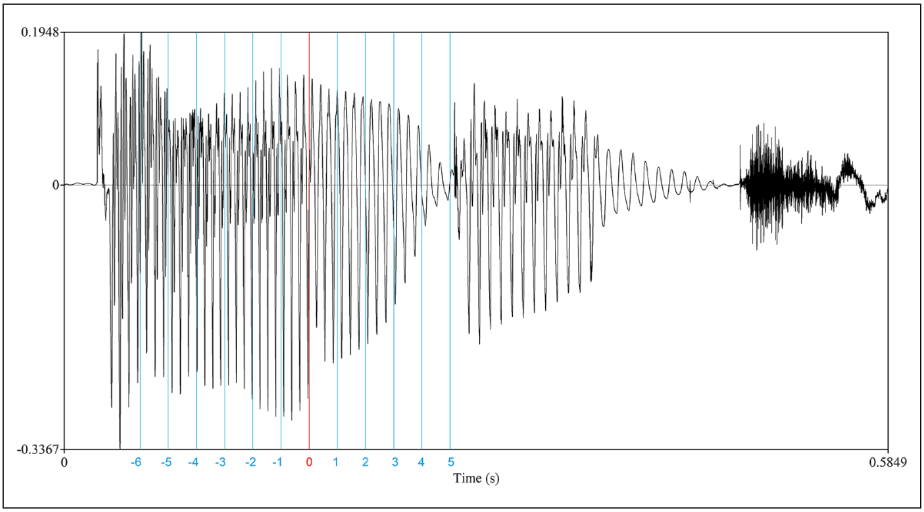

Figure 1. Gates for the word *bandage* imposed over its waveform.
Notes. for each word, the alignment point (marked as 0) occurs at the end of the first vowel. Image drawn in Praat Picture (Boersma and Weenink, 2019).

c Procedure. Due to the overlap of data collection with the beginning of the Covid-19 pandemic, data could not be recorded in the same way for all participants. Specifically, while 13 French participants were tested in testing booths in Paris in early March 2020, the remaining 3 French and the 19 English participants were tested online. Participants tested in the lab were paid 10 Euros for their time, while online participants volunteered to complete the experiment without monetary compensation.

Participants tested in the lab were seated in front of a computer in a quiet environment with headphones. Experiment sessions were conducted in English by an L1 speaker of AmE. Participants tested online were similarly asked to complete the experiment in a quiet environment with headphones. The online version of the experiment was created with PsychoPy (Peirce et al., 2019) and run on the experiment-hosting platform Pavlovia (https://www.pavlovia.org).

For each word, stimuli were presented in order from shortest to longest. At each presentation, two visual word forms corresponding to the relevant word pair (e.g. *bunny* and *butter*) were displayed on opposite sides of the screen. In addition, a four-interval confidence scale was presented above each word form, consisting of colored circles ranging from 'completely sure' (dark green) to 'completely unsure' (dark red), as shown in Figure 2.

Upon hearing each presentation, participants selected simultaneously the word they thought the stimulus corresponded to and how confident they were in their choice. To do so, lab participants pressed one of eight keys on the keyboard corresponding to their word choice and confidence level. Relevant keys were labeled with colored stickers that matched the colors and line-up of the circles on the screen. Online participants, by contrast, clicked on one of the eight colored circles.

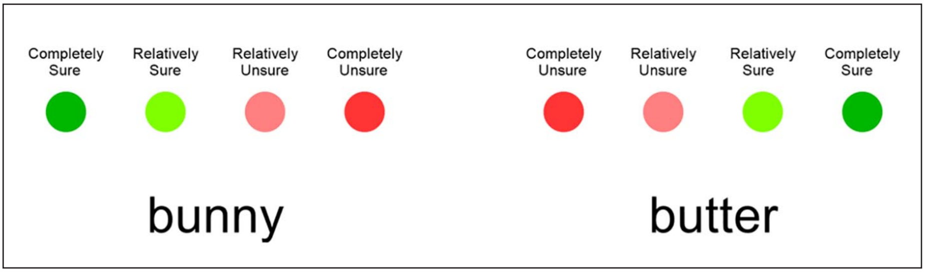

Figure 2. Visual display for the gating task, with sample test pair.

After the final presentation, participants heard the full word before moving on to the next one. Each participant heard 40 words total: 20 test words and 20 distractor words. The same distractor word in each pair was presented to all participants, but test word presentation was randomized and counterbalanced across participants, such that those who heard *bunny* did not hear *butter*, and vice versa. Visual word form order presentation was counterbalanced such that participants saw nasal and oral items on the left and right sides of the screen at equal rates.

Finally, all participants completed a language background and demographic questionnaire.

2 Results

We analysed performance on the test words only. All analyses were carried out in the lme4 package (Bates et al., 2015) in the R environment (R Core Team, 2020); parameters were estimated using maximum likelihood, fixed effects were contrast-coded, the random structure was the maximal one allowing for convergence and avoiding overfitting (as indicated by lme4's singular fit warning), the *bobyqa* optimizer (Powell, 2009) was used if it helped obtaining model convergence, and significance was established by means of model comparison using a likelihood ratio test.

a Isolation and recognition points. For each participant and for each word, the 'isolation point' is defined as the earliest presentation at which the participant gave the correct response and did not deviate from it until the end of the word, and the 'recognition point' is defined as the earliest presentation at which the participant's confidence rating additionally did not dip below 'relatively sure' for the rest of the word. All words were correctly isolated, but not all words were recognized, in the sense defined just above, before word offset. Table 1 shows the mean percentage of words with no recognition point.

We first performed a logistic mixed-effects regression on the data in Table 1, with fixed effects Group (French vs. English), Type (Nasal vs. Oral) and their interaction, random intercepts for Participant and Item, and a by-item random slope for Group. The model revealed neither a main effect nor an interaction (all $|z| < 1$). Thus, French and English participants did not differ in their ability to recognize nasal and oral items (or not) before the end of the word. Words without a recognition point were excluded from recognition point analyses.

Table 1. Mean percentage of words without a recognition point by participant group and trial type (standard errors in parentheses).

	Nasal	Oral
English	2.63 (3.67)	4.38 (5.12)
French	2.11 (3.30)	6.88 (6.35)

Table 2. Mean percentage of isolation points before or at vowel offset, and mean normalized isolation point (i.e. proportion of the way into the vowel at which they are located) (standard errors in parentheses).

	Mean percentage		Mean normalized isolation point	
	Nasal	Oral	Nasal	Oral
English	84.2 (8.40)	91.1 (6.56)	0.74 (0.04)	0.71 (0.04)
French	83.8 (9.25)	81.9 (9.65)	0.73 (0.04)	0.72 (0.05)

In order to determine both how often participants relied on vowel nasalization to isolate and recognize target words and the time frame in which they did so, the next set of analyses focuses on those isolation and recognition points that occurred before or at first vowel offset; that is, before participants heard the upcoming, phonemically disambiguating consonant that distinguished target and competitor words.

b Isolation points before or at vowel offset. We start with isolation points. Table 2 shows mean percentages of words that were isolated before or at vowel offset and mean proportions of the way into the vowel that these isolation points occurred, henceforth 'normalized isolation points'.[3] Note that both French and English participants isolated more than 80% of nasal and oral items before or at vowel offset, and that normalized isolation points for these items occurred on average well before vowel offset.

We performed a logistic mixed-effects regression on the percentage of isolation points located before or at vowel offset, with Group, Type and their interaction as fixed factors, and random intercepts for Participant and Item Pair. This model revealed no main effect of Group ($\beta = 0.23$, $SE = 0.13$, $z = 1.71$, $Chi^2(1) = 2.75$, $p = .1$) or Type ($\beta = -0.13$, $SE = 0.11$, $z = -1.20$, $Chi^2(1) = 1.40$, $p > .1$), but a trend towards an interaction ($\beta = -0.21$, $SE = 0.11$, $z = -1.87$, $Chi^2(1) = 3.39$, $p < .07$). Therefore, French and English participants did not reliably differ in their rate of isolation of nasal and oral items before or at vowel offset.

We then ran a linear mixed-effects regression on normalized isolation points, with Group, Type and their interaction as fixed factors, random intercepts for Participant and Item Pair, and a by-Item Pair random slope for Group. The model revealed no main effect of Group ($|t| < 1$) or Type ($\beta = 0.01$, $SE = 0.01$, $t = 1.48$, $Chi^2(1) = 2.24$, $p > .1$), and no interaction ($t < 1$), confirming that for those items that they isolated before or at vowel offset, French and English participants displayed a similar isolation timeline.

c Recognition points before or at vowel offset. Next, we turn to recognition points. Table 3 shows mean percentages of words recognized before or at vowel offset, and mean

Table 3. Mean percentage of recognition points before or at vowel offset and mean normalized recognition point (i.e. proportion of the way into the vowel at which they are located) (standard errors in parentheses).

	Mean percentage		Mean normalized recognition point	
	Nasal	Oral	Nasal	Oral
English	52.4 (11.5)	52.7 (11.5)	0.87 (0.03)	0.83 (0.03)
French	55.6 (12.5)	49.7 (12.6)	0.84 (0.04)	0.83 (0.04)

'normalized recognition points' (analogous to normalized isolation points, only for those recognition points occurring before or at vowel offset). French and English participants recognized approximately half of nasal and oral words before or at vowel offset.

We performed analyses analogous to those on the isolation point data above. A logistic mixed-effects regression on the percentage of recognition points located before or at vowel offset with Group, Type and their interaction as fixed factors, and a random structure consisting of intercepts for Participant and Item Pair revealed neither a main effect nor an interaction (all $|z| < 1$). Similarly, a linear mixed-effects regression on normalized recognition points with Group, Type and their interaction as fixed factors, random intercepts for Participant and Item Pair, and a by-Item Pair slope for Group revealed no main effects (both $t < 1$) and no interaction ($\beta = 0.01$, $SE = 0.01$, $t = 1.25$, $Chi^2(1) = 1.62$, $p > .1$). Thus, English and French participants differed neither in the percentage of nasal and oral items that they recognized before the beginning of the disambiguating consonant, nor in the proportion of the vowel they needed to hear to recognize these items.

3 Discussion

Taken together, the results from this experiment show that the performance of French participants did not differ from that of English participants. First, the two groups recognized words before hearing the word in its entirety at the same rate. Second, there were no group differences in the number of words isolated and recognized before or at vowel offset. Third, and most importantly, for words that were isolated/recognized before or at vowel offset, the groups did not differ in the proportion of the way into the vowel at which isolation/recognition was achieved.

These results add to the evidence that advanced L2 learners can use allophonic distributions in an offline task. The difference with respect to previous studies (Altenberg, 2005; Ito and Strange, 2009; Shoemaker, 2014) is that the relevant feature, i.e. nasalization, is used contrastively on vowels in the native language. Therefore, French listeners are used to distinguishing nasal from oral vowels for the purposes of word recognition. What they had to adapt to, though, is the different distributional pattern, with nasal and oral vowels being predictive of an upcoming nasal or oral consonant, respectively.

As self-rated advanced listeners residing in France, the L2 exposure and proficiency of the French learners are on par with the advanced L2 groups in Altenberg (2005), Shea and Curtin (2010) and Shoemaker (2014), all of whom exhibited learning of an L2 allophonic distribution to some degree. By contrast, Shea and Curtin (2010) also tested low-intermediate learners, who showed little to no learning. In addition, two of these previous studies reported positive correlations with proficiency: L2 listeners' performance on English word segmentation with stop aspiration cues in Ito and Strange (2009) was dependent on their length of residence in the United States, and, likewise, Shoemaker's (2014) 3rd-year student group outperformed 1st-year students on an English word segmentation task with the same cues. We did not test a mid- or low-proficiency group; moreover, our French participants had low to no length of residence in an English-speaking region and rated themselves similarly on all measures of proficiency, leaving little room for differences in these measures to correlate with task performance.[4]

To sum up, French advanced learners, like L1 English listeners, isolated and recognized nasal and oral words based on the presence or absence of nasality in the first vowel, demonstrating that they have learned the allophonic distribution of nasal and oral vowels in English, and that they can use this knowledge in an offline gating task to distinguish words from competitors. In Experiment 2, we use an eye-tracking procedure to examine if they can do the same in real time.

III Experiment 2

In this experiment, we test French advanced learners with a similar background as those in Experiment 1, and L1 English listeners, using an on-line eye-tracking procedure in the visual-world paradigm (Allopenna et al., 1998). In this paradigm, increased fixations to a target image in the presence of competitor images during auditory perception of a corresponding word indicate increasing activation of the target word candidate, and simultaneous increasing inhibition of competitor candidates, consistent with models of word recognition like the TRACE model, among others (Allopenna et al., 1998; McClelland and Elman, 1986). In addition, fixation latencies act as a proxy for the speed of activation of target words (e.g. Dahan et al., 2001). We utilize both of these measures of activation to characterize lexical competition for nasal–oral pairs, both initially and over the course of the word.

1 Methods

a Materials. Test items were the same 20 nasal–oral word pairs as in the gating experiment (e.g. *bunny–butter*), plus one additional pair of the same structure.[5] There were also 20 distractor pairs, constructed by replacing one word of each of the distractor pairs from the previous experiment with a phonologically unrelated word. For instance, *sofa* was paired with *kiwi* rather than with *soda*. All stimuli were imageable nouns. The full list of test and distractor pairs can be found in Appendix 1.

For the test phase of the experiment, a male L1 speaker of American English (AmE) recorded each word in the carrier sentence 'Click on the X', where X represents the stimulus word.[6] For a familiarization phase, another male L1 speaker of AmE recorded each word in isolation. Recordings took place in a sound-proof booth with state-of-the-art equipment.

Visual stimuli were 80 color images selected from pngtree.com, a free image-sourcing platform. All were resized to approximately 350 × 350 pixels. The full repository of images and auditory stimuli used in the current experiment are available on the OSF repository.

b Participants. The test group consisted of 24 L1 self-reported French advanced learners of English (age range: 19–40 years, median age: 25 years) living in France, who had started learning English in the French educational system in middle school. Their mean self-reported English verbal comprehension score was 7.7 on a scale from 1 (poor) to 10 (outstanding) ($SD = 1.35$). Seven participants reported having lived abroad in an English-speaking region for an average of 14 months.

The control group consisted of 24 L1 English listeners (age range: 18–43 years, median age: 25 years) living in the Paris region. Of these participants, fourteen were speakers of North American varieties of English and the remaining participants speakers of British English ($n = 9$) or Australian English ($n = 1$).

Ten additional participants were tested but discarded from the analyses because they were native bilinguals (French: n=3; English: $n = 4$) or because of trouble with eye-tracker calibration (French: $n = 3$). All participants were tested in Paris and were paid 10 Euros for their participation. None of them had participated in the gating experiment.

c Procedure. Experiment sessions were conducted in English by an L1 speaker of AmE. Participants were seated in a soundproof booth 60 cm in front of a 1,920 × 1,080-pixel computer screen and SR Research Eyelink 1000 eye-tracker. They donned a headset, placed their chin on a headrest for stability, and completed a standard calibration procedure.

The experiment started with a familiarization phase, during which no eye movements were tracked. On each familiarization trial, a stimulus image was shown on the computer screen with its written form presented beneath it, while participants heard the corresponding word in isolation. All items (42 test, 40 distractor) were presented in random order, and participants pressed a button to proceed from one item to the next.

In the test phase, participants were instructed to click on the image corresponding to the word they heard as fast as they could while their eye movements were recorded. In each trial, each image fit inside a 4-inch square. Images of a pair appeared on opposite sides of the screen, positioned halfway between the center fixation cross and the edge of the screen on a horizontal axis. After 2,000 milliseconds, the images disappeared and a fixation cross appeared in the center of the screen. One thousand milliseconds after the appearance of the fixation cross, participants heard the carrier sentence, ‘Click on the X’, and the two images reappeared at the onset of the target word. An example visual display is shown in Figure 3.

The test phase was divided into four blocks, and participants could take a short break between blocks. A recalibration procedure was performed after each break. In each block, all 41 item pairs (21 test and 20 distractor) were presented in a random order. Participants were instructed to click on one image of a pair in two blocks, and on the other one in the remaining two blocks, such that during the entire test phase, they heard each of the 82 words twice. The six possible orders of words within a pair across the four blocks, i.e. AABB, ABAB, ABBA, BBAA, BABA, and BAAB, occurred equally often, and participants were not told how many blocks they would complete. Thus, they could never predict which word they would hear upon seeing a given image pair.

At the end of the experiment, the French participants completed a language background assessment similar to the one in Experiment 1, with an added question to gauge their familiarity with items prior to the experiment. Twelve participants indicated that none of the words were new to them, 10 estimated that they had never encountered between 1 and 5 words (of $n = 82$), and 2 had been unfamiliar with between 5 and 10 words.

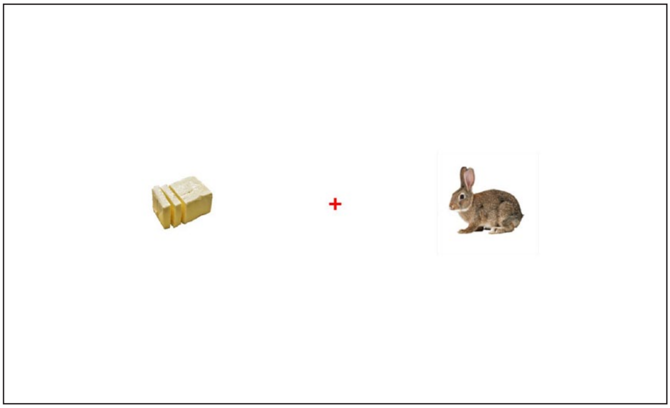

Figure 3. Visual display for the eye-tracking task (sample test pair: *bunny–butter*).

2 Results

For the analyses reported in this section, like for those in Experiment 1, parameters were estimated using maximum likelihood, fixed effects were contrast-coded, the random structure was the maximal one allowing for convergence and avoiding overfitting, the *bobyqa* optimizer (Powell, 2009) was used if it helped obtaining model convergence, and significance was established by means of model comparison using a likelihood ratio test.

a Accuracy. Performance on the behavioral component of the task was near ceiling for both groups. The percentage of trials in which participants clicked on the wrong image was 1.7% for French and 1.2% for English participants. A *t*-test revealed no difference in error rate between groups ($t < 1$). Errorful trials were excluded from further analysis.

b Proportions of fixations: Test items. To obtain a holistic picture of lexical activation of the crucial test items during listening, we calculated proportions of fixations to target, as shown in formula (1), in a time window that started 200 milliseconds after word onset and ended 700 milliseconds after first vowel offset. (The 200-millisecond delay accounts for the time it takes to execute an eye movement; Altmann, 2011.) A fixation was counted as a fixation to the target/competitor if it fell into the area on the screen in which the target/competitor picture was located.

$$\textit{Proportion of Target fixations} = \frac{\textit{Target fixations}\,(ms)}{(\textit{Target} + \textit{Competitor})\ \textit{fixations}\,(ms)} \tag{1}$$

Table 4 shows mean proportions of fixations to nasal and oral targets by English and French participants, and Figure 4 shows time course plots of mean proportions of fixations to target items minus mean proportion of fixations to competitor items.

Table 4. Mean proportions of fixations to target in nasal and oral test trials (standard errors in parentheses).

	Nasal	Oral
English	0.84 (0.04)	0.81 (0.05)
French	0.78 (0.05)	0.77 (0.05)

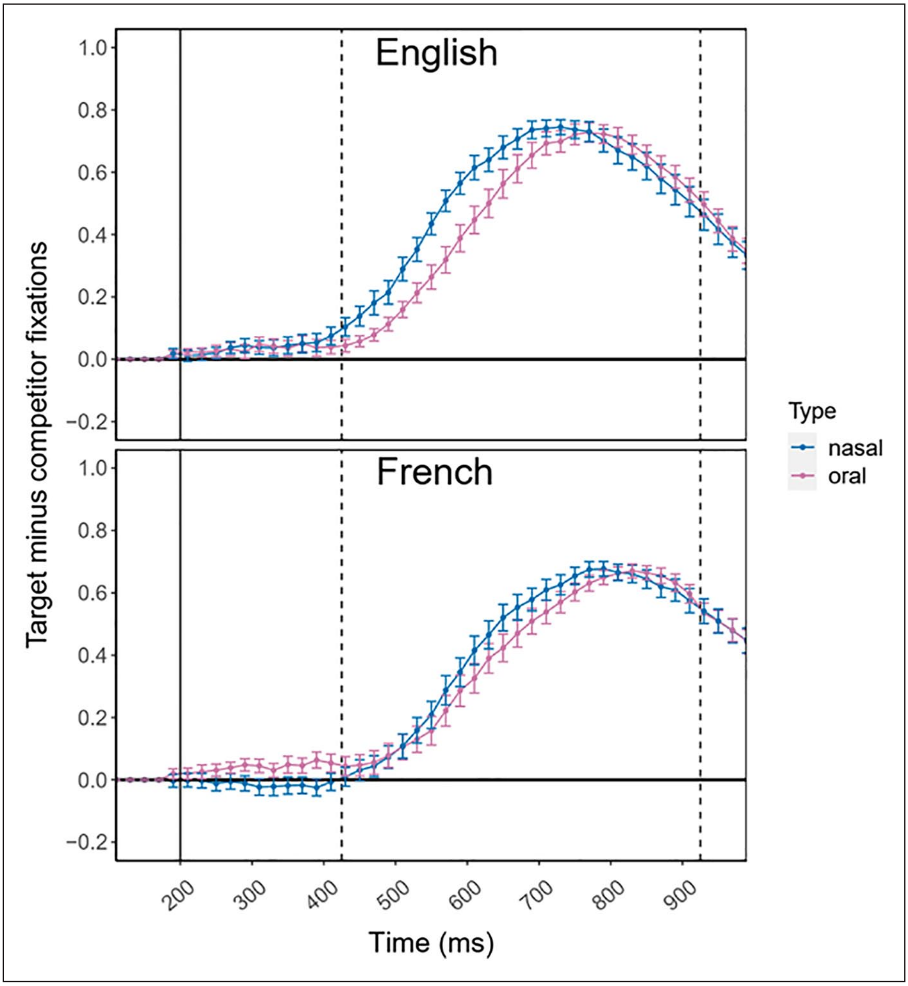

Figure 4. Mean proportions of target fixations minus mean proportions of competitor fixations by Group and Type.

Notes. The solid vertical black line marks the beginning of the time window for analysis, at 200 milliseconds after word onset. The leftmost dotted black line marks 200 milliseconds after mean vowel offset. The rightmost dotted black line marks mean analysis window offset.

Table 5. Mean target fixation latencies (ms) in nasal and oral test trials (standard errors in parentheses).

	Nasal	Oral
English	465 (37)	478 (36)
French	479 (40)	476 (38)

Note that fixations to targets seem to increase later in the French group compared to the English group. Additionally, an asymmetry in fixations to nasal vs. oral targets seems to be present for both groups, such that fixations to nasal targets increase earlier than those to oral ones. To quantify these differences, we performed a linear mixed-effects regression on proportions of fixations to target in the time window specified, with Group (English vs. French), Type (Nasal vs. Oral) and their interaction as fixed effects, and random intercepts for Participant and Item Pair. This model revealed main effects of Group (β=2.25, *SE*=0.694, t=3.24, $Chi^2(1)$=9.47, $p<.01$) and Type (β=1.01, *SE*=0.378, t=2.68, $Chi^2(1)$=7.15, $p<.01$), but no interaction between the two ($t<1$). French participants had overall lower proportions of target fixations than English participants, suggesting heightened competition between nasal and oral phones in the former. Both French and English participants had higher proportions of target fixations for nasal targets, indicating that the presence, but not absence, of vowel nasality facilitated the resolution of this competition in both groups.

c Target latency: Test items. Our second analysis of the test items focused on target fixation latencies in test items, an early measure of lexical activation considered a visual analog to reaction time (Dahan et al., 2001). Target latency was defined as the time to first fixation of the target from 200 milliseconds after target word onset. Mean target latencies to nasal and oral items for both English and French participants are shown in Table 5.

We ran a linear mixed-effects regression model on these data with fixed factors Group, Type and their interaction, random intercepts for Participant and Item Pair, and a by-pair random slope for Group. There were no main effects of Group ($t<1$) or Type ($\beta=-3.04$, *SE*=2.92, $t=-1.04$, $Chi^2(1)$=1.09, p=.30), and no interaction between the two ($\beta=-3.41$, *SE*=2.92, $t=-1.17$, $Chi^2(1)$=1.36, p=.24). Thus, there were no differences either between English and French participants or between nasal and oral targets in the time until first target fixation.

d Proportions of fixations: Test vs. distractor items. Next, we investigated whether the slower timeline of target-competitor resolution observed for the French group in the proportions of target fixation analysis (as shown in Section III.2.b) was restricted to nasal–oral pairs, or whether it was a direct consequence of slower L2 word processing in general. To this end, we performed an analysis on proportions of target fixations in test and distractor pairs. Proportions of fixations to target were calculated as in (1) above. The time window for analysis of the distractor pairs was equivalent to that of the test pairs, beginning at 200 milliseconds after word onset and ending 700 milliseconds after first vowel offset. Table 6 shows mean proportions of fixations to test and distractor targets for English and French participants.

Table 6. Mean proportions of target fixations in test and distractor trials (standard errors in parentheses).

	Test	Distractor
English	0.82 (0.05)	0.89 (0.04)
French	0.78 (0.05)	0.89 (0.05)

We submitted these data to a linear mixed-effects regression with fixed factors Group (English vs. French), Type (test vs. distractor) and their interaction, random intercepts for Item Pair and Participant, and a random by-participant slope for Type. There was no main effect of Group ($\beta=0.86$, $SE=0.548$, $t=1.56$, $Chi^2(1)=2.38$, $p=.12$), but a main effect of Type ($\beta=4.50$, $SE=0.488$, $t=9.21$, $Chi^2(1)=49.16$, $p<.001$) and an interaction ($\beta=-1.38$, $SE=0.324$, $t=-4.26$, $Chi^2(1)=15.38$, $p<.001$). Targets in distractor trials were fixated overall more than targets in test trials, but this difference was larger for French than for English participants. Specifically, while French participants fixated targets in test trials less than English participants, the two groups did not differ on distractor trials. Therefore, the heightened lexical competition for nasal–oral pairs observed in French participants in the proportions of target fixations analysis of the test trials was not due to a general L2 word processing delay.

3 Discussion

The group results of this experiment can be summarized as follows. In test trials, French participants had lower overall proportions of target fixations than English participants, yet the fixation pattern for oral vs. nasal targets was similar in both groups: nasal targets were fixated more than oral targets. Hence, the presence of vowel nasality facilitated the resolution of competition in nasal–oral pairs more than its absence. In addition, the two groups did not differ in proportions of target fixations in distractor trials (which were higher than those in test trials). There was also no difference between groups in target latencies of test trials, i.e. in the time it took to first fixate nasal and oral targets. In other words, the French participants only differed from the English participants in that they were slower to completely rule out the competitor in nasal–oral pairs.

That the French participants exhibited the same pattern of recognition of nasal and oral words as the English participants did is evidence that they have learned to use the allophonic distribution of English nasal and oral vowels to their advantage in online word recognition. Specifically, upon hearing a nasal vowel, they were quicker to rule out oral competitors than vice versa. In the previous experiment, though, neither group demonstrated such a nasal–oral response asymmetry. We will discuss this difference between the two sets of results in Section IV.

In contrast with the French group's delay in resolving competition in test pairs, we found no difference between groups in either initial target fixation latencies in test trials or proportions of target fixations in distractor trials, confirming that the French test pair delay arises from the particular challenge of processing the nasal–oral word pairs. What remains unclear is whether the observed delay was due to a processing cost associated with exploiting L2 allophonic cues specifically, or whether it was induced by the general phonetic similarity of nasal and oral words of a pair. This question is beyond the scope of the present study, but could be investigated in an experiment in which competition resolution of nasal–oral pairs is compared to that in phonologically similar distractor pairs.

Finally, note that, as in Experiment 1, we only tested a group of self-rated advanced learners, who, according to questionnaire data, were fairly homogeneous in terms of their language learning background and demographic characteristics.[7] The present results therefore do not allow us to provide insight into the developmental trajectory of L1-like word processing strategies and the factors that influence it.

IV General discussion

Using an offline gating paradigm and an online visual world paradigm with eye-tracking, we found that French advanced learners of English have not only learned the allophonic distribution of nasal and oral vowels, but also use it to their advantage in word recognition. In the gating task in Experiment 1, the learners did not differ from the native English participants either in the timeline of target word isolation or in reported confidence levels in word choice. This result is in accordance with previous studies on advanced learners' knowledge of L2 allophonic distributions (Altenberg, 2005; Ito and Strange, 2009; Shea and Curtin, 2010; Shoemaker, 2014), and extends it to a case where the L2 contrast concerns a phonetic feature that is phonemic in the native language. One may argue that the presence of the written word forms of both targets and competitors on the screen enhanced participant awareness of vowel nasality differences between words of a pair: in both French and English orthography, a nasal vowel is represented by a vowel followed by a nasal consonant. Perhaps, then, we would fail to observe native-like performance in a picture-based gating task. Yet, the results from the eye-tracking task demonstrate that our French participants do not need orthographic word forms to use vowel nasality to facilitate word recognition. In Experiment 2, they were just as fast as the English participants to initially fixate targets in test trials, and they showed the same asymmetric looking behavior, such that they resolved the nasal–oral competition faster for nasal than for oral targets. These eye-tracking data provide evidence that learners can use allophonic knowledge to develop facilitative strategies in online word recognition. It should be noted, though, that – contrary to what we observed in the gating task – the French participants in the eye-tracking task were not native-like: compared to the English participants, they showed overall smaller proportion target fixations in test trials. In other words, while they initially fixated the target as quickly as the English participants, they spent more time fixating the competitor during the entire time window, suggesting a higher level of uncertainty.

These results resonate with a recent study that, conversely, investigated English listeners' processing of vowel nasalization in French, or, more precisely, Canadian French (Desmeules-Trudel and Zamuner, 2023). Like Metropolitan French, Canadian French contrasts nasal with oral vowels, but oral vowels additionally exhibit nasalization when followed by a nasal consonant. Thus, the distinction between a nasal and an oral vowel in a pair such as *pain* /pɛ̃/ 'bread' – *peigne* /pɛɲ/ 'comb' can be quite subtle. Using eye-tracking and a visual world paradigm, Desmeules-Trudel and Zamuner (2023) found that, like native Canadian French listeners, advanced native English learners of Canadian French are sensitive to the amount of vowel nasalization, but they are not as fast as native listeners to resolve the competition between, say, *pain* and *peigne*. (The visual world display also contained a competitor with an oral vowel, e.g. *pêche* /pɛʃ/ ‘peach’, but looks to this type of competitor were not analyzed.) As noted by the authors, English listeners presumably performed well in this study because they are also sensitive to variation in vowel nasalization in their native language (Beddor et al., 2013).

A seemingly puzzling difference in the results of our two experiments concerns the nasal–oral asymmetry observed for both groups in the eye-tracking data only. That is, in the eye-tracking task, the competition between oral and nasal consonants was resolved more quickly upon hearing a nasal than an oral vowel, whereas in the gating task, there was no difference in oral and nasal targets in any of our measures. The presence of an asymmetry observed in eye-tracking can be interpreted straightforwardly as a result of the dynamics of word recognition in real time. Specifically, vowel nasality, while a reliable indicator of an upcoming nasal consonant, is variably realized in English: The timing of its onset as well as its strength in any given word are dependent upon such factors as the phonological environment surrounding the vowel, lexical factors such as neighborhood density, speech rate and style, as well as individual talker characteristics and group-level patterns of variation (Beddor et al., 2013; Kim and Kim, 2019; Krakow, 1993; Scarborough, 2013; Tamminga and Zellou, 2015). As a consequence, hearing no or little nasality in a vowel does not preclude the possibility of an upcoming nasal consonant, making it beneficial for listeners to maintain uncertainty when hearing an oral vowel and hence use a conservative threshold to rule out nasal competitors. In this way, the presence of nasality in the signal is a more reliable cue to an upcoming nasal consonant than the absence of nasality is to an oral consonant, whence the observed asymmetry. But why did we not observe an asymmetry in the gating results? Recall that our analyses concentrate on those stimulus presentations from which the response is consistently correct (isolation point) or from which listeners are additionally confident they have identified the correct word (recognition point). As these points occur for both groups on average after 70% and 80% into the vowel, respectively, they provide no insight into listeners’ interpretation of the earliest parts of the vowel, which in nasal targets may contain little or no nasalization. In other words, the responses to the early presentations may well have been biased toward oral responses,[8] which is irrelevant for the question of the use of the allophonic distribution of nasal and oral vowels to predict an upcoming consonant.

One may wonder at what level knowledge about English vowel nasality is represented in French learners. That is, have they integrated a productive, allophonic rule in their phonological *grammar* of English, or is their English phonological *lexicon* composed of acoustically detailed episodic memories of heard words, as in exemplar-based models (Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001)? The latter would mesh well with the finding that L2 listeners’ recognition of pronunciation variants of words is sensitive to the frequency with which they have heard the variant (Brand and Ernestus, 2018). Note that the two possibilities are not mutually exclusive; yet, it is only the presence of a productive rule in their phonological grammar that would allow them to isolate targets from competitors equally well in case of words they have never heard before. Experiments similar to the ones reported here but with nonce words could thus shed light on this issue. For the eye-tracking task, for instance, participants could be taught pairings of *written* nonce words with images of fantasy objects, before being tested on their recognition of the nonce words. As participants would have no prior oral exposure to these items, success in this task would indicate that the allophonic distribution is part of their L2 phonological grammar.

There are several additional open questions that can be investigated in future research. First, as we tested only advanced learners, it is unknown at what level of proficiency French listeners develop knowledge of English vowel nasality and incorporate it into a facilitative word recognition strategy. Accordingly, it would be of interest to test beginning and intermediate French learners. Second, and relatedly, we have no knowledge about the factors that may influence individual performance in L2 learners. Administering an objective assessment of English proficiency would allow for a more accurate characterization of differences among individual learners. There is substantial evidence for a link between the perception and production of L2 phonemic sound contrasts (for a recent overview, see Melnik-Leroy et al., 2022); it would therefore be especially interesting to have a measure of French learners' production of nasal and oral vowels in English, and hence to investigate its relation with the online processing of this allophonic contrast.

As we hypothesized in Section I, the fact that vowel nasality is contrastive in French could have played a role in the success of the French participants. But the high acoustic salience of the distinction between nasal and oral vowels in general (Beddor, 1993) could be sufficient in and of itself to promote the learning and use of their distribution in English. Testing L2 learners whose L1 has no nasal vowels phonemically, and additionally exhibits little regressive nasal vowel coarticulation, would allow for an examination of the role of acoustic salience. An example of such a language is Spanish (Solé, 1995). If advanced Spanish learners of English show the same performance as our French learners, we would conclude that acoustic salience alone can drive the development of a strategy that exploits the allophonic nasal–oral distribution during online L2 word recognition. However, if Spanish learners are less efficient than French learners in exploiting the cue, such results would lend support to our original hypothesis that contrastiveness gives French learners an additional advantage.

Yet another factor that could play to the advantage of the French participants is potential exposure to certain varieties of Southern French, in which a nasal vowel can be followed by a brief nasal appendix before the onset of the following, non-nasal, consonant (Boula de Mareüil et al., 2007; Delvaux et al., 2012). It is possible that, for speakers of these varieties and for listeners who have had exposure to them, the presence of such appendices would facilitate their association of nasal vowels with nasal consonants in English. It is unknown to what extent our participants, who were tested in Paris, had experience with nasal appendices, but future research should take this factor into account.

To conclude, we have shown that French advanced learners of English can rely on the allophonic distribution of nasal and oral vowels to rule out candidate words online and hence speed up L2 word recognition. These results contribute to our knowledge of lexical processing in L2, and open many avenues for future research, from the learning trajectory and the many factors that may influence performance, to the exact characterization of L2 learners' knowledge of allophony.

Acknowledgements

We thank Jeremy Kuhn and Junseok Lee for recording the stimuli, and Clara Lejeune for assistance with data collection. We also benefited from comments and questions by audiences at BUCLD 46, AMLaP 2021, and ISBPAC 2021.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research for this article was supported by the Agence Nationale de la Recherche, France (ANR-17-CE28-0007-01 and ANR-17-EURE-0017).

ORCID iD

Megan M Dailey https://orcid.org/0000-0002-1685-8015

Notes

1. Whether this process is better described as coarticulation than as allophony, as in some of the studies reviewed below, is irrelevant for the present purposes. We follow Solé (1992, 1995), who – based on speech production data – argues that vowel nasalization in American English is phonologized; but, for a production study with the opposite conclusion, see Cohn, 1993. We acknowledge that the degree of nasalization varies according to a host of factors (for an overview, see Beddor et al., 2013).
2. The full sentences were used for the eye-tracking task in Experiment 2.
3. *Normalized* isolation points (proportions) were analysed instead of *raw* isolation points to control for the variable vowel lengths in the items.
4. We did perform linear regressions on French participants' mean percentage of isolation points before or at vowel offset, with the following seven regressors: length of residence in an English-speaking region, and self-rated English oral comprehension, oral expression, reading, vocabulary, grammar, and pronunciation. These regressions revealed no correlations (before Bonferroni correction: $p = .07$ for English vocabulary, all other $p > .1$).
5. This pair, *donkey–dolphin*, was added to balance out the stimulus list, which is front-vowel heavy, with another pair with a low back vowel.
6. These were the same sentences recorded by the same speaker referred to in Experiment 1.
7. In the absence of an objective proficiency assessment, we used the questionnaire data to perform – like in Experiment 1 – linear regressions on French participants' individual mean target latencies in test trials. Among the seven predictors (length of residence in an English-speaking country and six self-rated measures of English proficiency), only self-reported English pronunciation yielded a significant correlation ($\beta = -7.41$, $SE = 2.14$, $t = -3.47$, adjusted $R^2 = 0.354$, $p = .002$, with a Bonferroni-adjusted significance level of $p < .007$), such that participants with higher self-rated pronunciation were faster to initially fixate nasal and oral targets. As pronunciation was the measure that exhibited by far the largest variance, it was also the one where a correlation between performance and self-rated proficiency was most likely to be observed in the first place. It is quite possible that with more variability in the other measures we would observe additional correlations.
8. Such a bias was indeed reported for native English listeners by Lahiri and Marslen-Wilson (1991), who likewise used a gating paradigm to examine the recognition of words with a nasal vs. an oral consonant; but, for contrasting results, see Ohala and Ohala, 1995.

References

Allopenna PD, Magnuson JS, and Tanenhaus MK (1998) Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–39.

Altenberg EP (2005) The perception of word boundaries in a second language. *Second Language Research* 21: 325–58.

Altmann GTM (2011) Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica* 137: 190–200.

Baayen RH, Piepenbrock R, and Gulikers L (1995) *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bates D, Maechler M, Bolker B, and Walker S (2015) Fitting linear mixed effects models using lme4. *Journal of Statistical Software* 67: 1–48.

Beddor PS (1993) The perception of nasal vowels. In: Huffman MK and Krakow RA (eds) *Phonetics and phonology V: Nasals, nasalization, and the velum*. San Diego, CA: Academic Press, pp. 171–96.

Beddor PS, McGowan KB, Boland JE, Coetzee AW, and Brasher A (2013) The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133: 2350–66.

Boersma P and Weenink D (2019) *Praat: Doing phonetics by computer* [computer program]. Available at: http://www.praat.org (accessed July 2023).

Boula de, Mareüil P, Adda-Decker M, and Woehrling C (2007) *Analysis of oral and nasal vowel realization in Northern and Southern French varieties*. In: Trouvain J (ed.) *Proceedings of the 16th International Congress of the Phonetic Sciences (ICPhS)*. Saarbrücken: Universität des Saarlandes, pp. 2221–24.

Brand S and Ernestus M (2018) Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *The Quarterly Journal of Experimental Psychology* 71: 1240–59.

Cohn AC (1993) Nasalisation in English: Phonology or phonetics. *Phonology* 10: 43–81.

Cutler A (2012) *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Dahan D, Drucker SJ, and Scarborough RA (2008) Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108: 710–18.

Dahan D, Magnuson JS, Tanenhaus MK, and Hogan EM (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16: 507–34.

Darcy I, Peperkamp S, and Dupoux E (2007) Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In: Cole J and Haulde JI (eds) *Laboratory phonology 9*. Berlin: Mouton de Gruyter, pp. 411–42.

Delvaux V, Huet K, Piccaluga M, and Harmegnies B (2012) *Inter-gestural timing in French nasal vowels: A comparative study of (Liège, Tournai) Northern French vs. (Marseille, Toulouse) Southern French*. In: *Thirteenth annual conference of the International Speech Communication Association (Interspeech)*. Baixas: International Speech Communication Association (ISCA), pp. 2682–85.

Desmeules-Trudel F and Zamuner T (2023) Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research* 39: 333–62.

Dow M (2020) A phonetic-phonological study of vowel height and nasal coarticulation in French. *Journal of French Language Studies* 30: 239–74.

Fowler CA and Brown JM (2000) Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception and Psychophysics* 62: 21–32.

Goldinger SD (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105: 251–79.

Grosjean F (1980) Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28: 267–83.

Gustafson E and Bradlow AR (2016) French speech segmentation in liaison contexts by L1 and L2 listeners. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7: 1–34.

Ito K and Strange W (2009) Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *The Journal of the Acoustical Society of America* 125: 2348–60.

Johnson K (1997) Speech perception without speaker normalization. In: Johnson K and Mullennix J (eds) *Talker variability in speech processing*. San Diego, CA: Academic Press, pp. 145–66.

Kim D and Kim S (2019) Coarticulatory vowel nasalization in American English: Data of individual differences in acoustic realization of vowel nasalization as a function of prosodic prominence and boundary. *Data in Brief* 27: 104593.

Krakow RA (1993) Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In: Huffman MK and Krakow RA (eds) *Phonetics and phonology V: Nasals, nasalization, and the velum*. San Diego, CA: Academic Press, pp. 87–116.

Lahiri A and Marslen-Wilson W (1991) The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38: 245–94.

McClelland JL and Elman JL (1986) The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.

Melnik-Leroy GA, Turnbull R, and Peperkamp S (2022) On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research* 38: 581–605.

Ohala JJ and Ohala M (1995) Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In: Connell B and Arvaniti A (eds) *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, pp. 41–60.

Peirce J, Gray JR, Simpson S, et al. (2019) PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51: 195–203.

Pierrehumbert JB (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee J and Hopper P (eds) *Frequency effects and the emergence of lexical structure*. Amsterdam, Netherlands: John Benjamins, pp. 137–57.

Powell MJ (2009) The BOBYQA algorithm for bound constrained optimization without derivatives. Report, DAMTP 2009/NA06, University of Cambridge, Cambridge.

R Core Team (2020) *R: A language and environment for statistical computing* [software]. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org (accessed July 2023).

Sampson R (1999) *Nasal vowel evolution in romance*. Oxford: Oxford University Press.

Scarborough R (2013) Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics* 41: 491–508.

Shea CE and Curtin S (2010) Discovering the relationship between context and allophones in a second language: Evidence for distribution-based learning. *Studies in Second Language Acquisition* 32: 581–606.

Shoemaker E (2014) The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition* 36: 709–31.

Solé M-J (1992) Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35: 29–43.

Solé M-J (1995) Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38: 1–23.

Tamminga M and Zellou G (2015) *Cross-dialectal differences in nasal coarticulation in American English*. In: *Proceedings of the 18th International Congress of Phonetic Sciences*. London: International Phonetic Association.

Tremblay A (2011) Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition* 14: 257–79.

Tremblay A and Spinelli E (2014) English listeners' use of distributional and acoustic -phonetic cues to liaison in French: Evidence from eye movements. *Language and Speech* 57: 310–37.

Tuinman A and Mitterer H (2012) The role of native-language knowledge in the perception of casual speech in a second language. *Frontiers in Psychology* 3: 249.

Tuinman A, Mitterer H, and Cutler A (2012) Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language* 66: 530–44.

Warren P and Marslen-Wilson W (1987) Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics* 41: 262–75.

Warren P and Marslen-Wilson W (1988) Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics* 43: 21–30.

Weber A and Broersma M (2012) Spoken word recognition in second language acquisition. In: Chapelle CA (ed.) *Encyclopedia of applied linguistics*. Chichester: Wiley-Blackwell.

Appendix 1. Full list of stimuli for Experiments 1 and 2.

Test pairs		Distractor pairs		
Nasal	Oral	Distractor 1	Distractor 2a**	Distractor 2b***
bandage (7)	backpack (6)	apple	actor	teapot
blanket (4)	blackboard (5)	basket	bathroom	suitcase
bunny (4)	butter (5)	bread	breath	guitar
camel (4)	cabin (4)	broccoli	broadcast	cherry
camera (4)	cactus (4)	coffee	costly	laptop
candle (4)	castle (4)	cupcake	custard	saxophone
candy (5)	casket (5)	hotdog	hockey	stapler
cinnamon (3)	cigarette (3)	moon	mood	pizza
dino (9)	diver (10)	notebook	nose ring	socks
donkey*	dolphin*	parrot	pallet	lighthouse
fingerprint (3)	fisherman (3)	panda	Pacman	flower
grandfather (5)	grasshopper (5)	pine tree	pie crust	grapefruit
lantern (6)	ladder (5)	puzzle	puddle	trophy
lemon (4)	lettuce (5)	razor	race car	mailbox
painting (3)	pastry (3)	seahorse	seagull	butterfly
pancake (4)	package (5)	skateboard	scalar	fire truck
peanut (3)	peacock (3)	sofa	soda	teddy bear
penny (3)	pepper (3)	toothbrush	toucan	giraffe
ping pong (3)	picture (3)	violin	violate	table
rainbow (5)	railroad (5)	wheelchair	weakness	piano
window (3)	whisky (3)			

Notes. The numbers in parentheses represent the number of gates per word before the alignment point at vowel offset in Experiment 1.
* This pair was only used in Experiment 2.
** In Experiment 1 only, as orthographic word forms displayed on the screen (participants heard only those distractor words in the Distractor 1 column).
*** In Experiment 2 only.