Spoken word recognition in a second language: The importance of phonetic details

Félix Desmeules-Trudel
University of Toronto Mississauga; University of Ottawa, Canada

Tania S. Zamuner
University of Ottawa, Canada

Abstract
Spoken word recognition depends on variations in fine-grained phonetics as listeners decode speech. However, many models of second language (L2) speech perception focus on units such as isolated syllables, and not on words. In two eye-tracking experiments, we investigated how fine-grained phonetic details (i.e. duration of nasalization on contrastive and coarticulatory nasalized vowels in Canadian French) influenced spoken word recognition in an L2, as compared to a group of native (L1) listeners. Results from L2 listeners (English-native speakers) indicated that fine-grained phonetics impacted the recognition of words, i.e. they were able to use nasalization duration variability in a way similar to L1-French listeners, providing evidence that lexical representations can be highly specified in an L2. Specifically, L2 listeners were able to distinguish minimal word pairs (differentiated by the presence of phonological vowel nasalization in French) and were able to use variability in a way approximating L1-French listeners. Furthermore, the robustness of the French “nasal vowel” category in L2 listeners depended on age of exposure. Early bilinguals displayed greater sensitivity to some ambiguity in the stimuli than late bilinguals, suggesting that early bilinguals had greater sensitivity to small variations in the signal and thus better knowledge of the phonetic cue associated with phonological vowel nasalization in French, similarly to L1 listeners.

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

Corresponding author:
Félix Desmeules-Trudel, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, CCT Building, Room 4019, Mississauga, ON L5L 1C6, Canada.
Email: felix.desmeules.trudel@utoronto.ca
Learning and mastering words is crucial to speaking and understanding a second language (L2), and spoken language processing is dynamic and time-dependent. Thus, research on how L2 listeners process words as the auditory signal unfolds is crucial to understand how bilinguals process language. Furthermore, the influence of fine-grained characteristics of speech sounds on L2 listeners’ word interpretation is still understudied, although the effect has been repeatedly demonstrated in the native language (L1), e.g. coarticulation (e.g. Beddor et al., 2013; Cross and Joanisse, 2018; Dahan et al., 2001; Desmeules-Trudel et al., 2020; Paquette-Smith et al., 2016; Zamuner et al., 2016) and within-category variability (Desmeules-Trudel and Zamuner, 2019; McMurray et al., 2002, 2008). The current article investigates the interplay between fine-grained speech processing and real-time word recognition in an L2. Specifically, in two eye-tracking experiments, we assessed native English listeners’ interpretation of minimal pairs of (Canadian) French words that differed in duration of vowel nasalization, a cue that correlates with the vowel nasality contrast in French, in order to better understand the extent of sensitivity to fine-grained phonetic details in an L2. The L2 listeners’ results were further compared with previously published data collected with L1 listeners of Canadian French (Desmeules-Trudel and Zamuner, 2019). The results suggest that variability in nasalization duration, which expresses a phonological distinction in the English listeners’ L2 (i.e. phonological contrast between nasal and non-nasal vowels in French) but not in the L1 (i.e. coarticulatory-only nasalized vowels in English), impacts online recognition of L2 words. We also evaluated how age of exposure to the L2 influences lexical activation of L2 words, as a first step towards more detailed investigations of the impact of L2 listeners’ language background on real-time processing and L2 learning (Flege and Bohn, 2021). Our results suggest that early exposure to the L2 enables listeners to process L2 words in a more native-like manner.

L2 sound perception and word recognition

A significant body of research has investigated how foreign and L2 speech sounds can be categorized according to pre-existing L1 phonological categories (Desmeules-Trudel and Joanisse, 2020; Escudero, 2005; Escudero and Chládková, 2010; Escudero and Vasiliev, 2011; Escudero and Williams, 2010; Escudero et al., 2014; Flege and MacKay, 2004; Faris et al, 2016; Frieda and Nozawa, 2007; Kuhl and Iverson, 1995; Levy, 2009; Levy and Strange, 2008; Major, 2008; Melnik-Leroy et al., 2021; Tyler et al., 2014). This phenomenon, referred to as assimilation (Best, 1995; Best and Tyler, 2007) or equivalence qualification (Flege, 1995), has been described in several theoretical frameworks on L2 speech perception, as well as for word recognition (Best, 1995; Best et al., 2001; Best and Tyler, 2007; Escudero, 2005; Flege, 1995; Flege and Bohn, 2021; Tyler et al., 2014; Weber and Cutler, 2004). For instance, Best’s (1995) Perceptual Assimilation Model (PAM) predicts that performance on non-native discrimination of L2 sounds depends on structural similarities and differences between the L1 and L2.

Most L2 perception studies have used stimuli such as non-meaningful syllables or sounds, as opposed to words that carry meaning. There is, however, some research that
has focused on the interaction between L2 sound perception and word recognition. This body of work shows that L2 sound assimilation to L1 categories creates asymmetries in recognition, and that the coactivation of lexical items depends on L2 sounds that can be confused in the listeners’ L1 (Broersma, 2012; Broersma and Cutler, 2011; Cutler et al., 2006; Cutler and Weber, 2007; Darcy et al., 2012; Escudero et al., 2008; Tuinman et al., 2007; Weber and Cutler, 2004). External factors such as age of first exposure, which are linked to the amount of L2 input, have also been shown to impact speech perception and lexical access (Flege and Bohn, 2021; Flege and Mackay, 2004; Schmidtke, 2014), suggesting that bilingual processing depends on more than just the structural similarities and differences between the L1 and the L2.

One of the examples that illustrate the interaction between speech perception and word recognition difficulties is the /ɹ–l/ English contrast that is misperceived by Japanese listeners (Goto, 1971). For instance, Cutler et al. (2006) suggested that the Japanese /r/ phoneme is activated for both L2-English /l/ and /ɹ/ (i.e. single-category assimilation) when L1-Japanese listeners are presented with English words, which in turn explains why both locker and rocket are activated for an extended period of time (i.e. until the second vowel is heard). In the tonal domain, Qin et al. (2019) found that L1-Mandarin listeners were highly sensitive to within-category variations in F0, due to the importance of F0 variations in Mandarin. L2-Mandarin (English-native) listeners were sensitive as well to variable F0, but had difficulties correctly categorizing the acoustic input compared to L1-Mandarin listeners. Thus, even though word recognition and sound-based tasks are different in nature and can be thought of as relatively independent (Darcy et al., 2012; in L1 acquisition, see Curtin et al., 2011; Werker and Curtin, 2005), research has shown that word recognition is hindered or disrupted when phonetic structures (e.g. sounds, prosody) are absent from the L2 listeners’ L1 phonological systems (Broersma, 2012; Cutler et al., 2006; Escudero et al., 2008; Qin et al., 2019; Tremblay, 2008; Tremblay et al., 2012; 2016; Ullman and Lovelett, 2018; Weber and Cutler, 2004).

2 Vowel nasalization in Canadian French and English

To investigate how word recognition and fine-grained phonetic details interact in L2 listeners, we used (Canadian) French words that contain phonological nasal vowels as well as potentially-nasalized vowels, i.e. oral vowels followed by a nasal consonant. These were tested with L1 (Canadian) English listeners who spoke French as an L2. We used pairs of words differentiated by vowel nasalization because it is a cue that is used phonologically in French, but is only coarticulatory in English. Furthermore, vowel nasalization is multidimensional and can be highly variable (Carignan, 2013, 2014; Delvaux, 2006; Desmeules-Trudel and Brunelle, 2018), therefore enabling us to assess whether (L2) listeners are sensitive to phonetic variability in the context of word recognition.

In Canadian French, phonological nasal vowels are realized with substantial nasalization duration on the vowel (i.e. the velum is lowered for at least 50% of the vowel duration, which gives the vowel its nasal quality, and nasalization increases as the vowel unfolds) – e.g. the word pain [pɛ̃] “bread”. They are different from European French nasal vowels in terms of oral articulation (Carignan, 2013, 2014), e.g. Canadian French
[ê] is fronted and raised and can be optionally diphthongized compared to European French [ɛ̃]. Nasal vowels can also be followed by a nasal appendix in Canadian French, i.e. a consonant-like resonance that emerges due to the overlap of the articulation of the vowel and a following consonant, when it precedes a stop within a word or in a following word (Desmeules-Trudel and Brunelle, 2018) – e.g. pantalons bleus [pãtalɔ̃blo] “blue pants” in which appendices are represented by superscripts. According to Desmeules-Trudel and Brunelle’s (2018) measurements, appendices are 28 ms long on average when the nasal vowel is followed by a heterosyllabic coda (oral) consonant (e.g. paon bleu [pãmblo] “blue peacock”), and 50 ms on average when the nasal vowel is followed by a homosyllabic coda consonant (e.g. pente douce [pãtdus] “gentle slope”). This phenomenon is pervasive in Canadian French, and has been documented in European French as well, e.g. in Southern France speakers (Carignan, 2017). In perception, appendices are impressionistically characterized as short nasal consonants present between the vowel and oral consonant (Coquillon and Turcsan, 2012).

Coarticulatorily-nasalized vowels are also found in Canadian French (Desmeules-Trudel and Brunelle, 2018; in European French, see Dow, 2020): oral vowels that are followed by a full nasal consonant display nasalization on the vowel, albeit shorter than for phonological nasal vowels – e.g. the vowel in peigne /pɛɲ/ “comb” can be optionally nasalized for a short period of time. Based on nasal airflow data, Desmeules-Trudel and Brunelle (2018) observed that approximately 20%–25% of a vowel in that context is nasalized on average with a gradual increase in “quantity of” nasalization leading to the actual nasal consonant. Multidimensional cues to vowel nasalization in French, i.e. nasal vowels differ from their oral counterparts in height, backness and absolute duration (Carignan, 2014), can also have an impact on perceptual patterns (Delvaux, 2009), but nasalization duration is a constant cue that indicates vowel nasalization. Importantly, L1 Canadian French listeners’ word recognition of quasi-minimal pairs with phonological nasal vowels and coarticulated nasalized vowels is gradient: more extensive nasalization on the vowel causes stronger activation of words with phonological nasal vowels, and listeners are sensitive to fine-grained variability in nasalization duration (Desmeules-Trudel and Zamuner, 2019). In other words, L1-French listeners were better at recognizing words with a phonological nasal vowel when the vowel was extensively nasalized compared to when the vowel was only partly nasalized, and this ability was not strictly categorical. These results are consistent with previous studies using voice onset time (VOT) values (McMurray et al., 2002, 2008) in which listeners were able to gradiently detect fine variations in VOT and use those for recognizing minimal pairs of words containing voiced-voiceless consonants (e.g. [p]–[b]). Specifically, McMurray et al. (2002, 2008) measured eye movements as a proxy of lexical activation while English-native listeners were presented with auditory stimuli containing stop consonants and images of objects corresponding to the heard words (e.g. beach–peach). Crucially, the initial stop consonants varied in terms of VOT in 5-ms increments between 0 ms (which corresponds to a voiced consonant in English, e.g. beach) and 40 ms (a typical VOT value for voiceless consonants in English, e.g. peach). The authors found that proportions of fixations to images depicting words containing a voiceless consonant gradiently increased as VOT values increased, suggesting that within-category variability impacted lexical access.
In English, words are not functionally differentiated based on vowel nasalization (e.g. the word *bed* /bɛd/ does not have a counterpart with a nasal vowel */bɛ̃d/), and duration of nasalization (i.e. amount of time during which the velum is lowered) on coarticulated vowels in English is variable (Beddor, 2009; Cohn, 1990; Tamminga and Zellou, 2015). The phonetic realization of vowel nasalization in English falls within the boundaries of both phonological and coarticulatory nasalized vowel categories of Canadian French. This is one of the main reasons for investigating L1 English speakers that have French as an L2, since the status of nasalization is different across languages but is based on similar phonetic correlates (e.g. duration of nasalization). Previous eye-tracking research has found that English listeners are able to use coarticulatory nasalization to recognize spoken words in their L1 (American English, Beddor et al., 2013; Canadian English, Desmeules-Trudel et al., 2020; Zamuner et al., 2016). Some differences across American and Canadian English have been underlined in the literature on vowel nasalization (Rochet and Rochet, 1999), but are expected to be minor in the current study since all L2 listeners spoke Canadian English as an L1. What is yet unknown is whether English listeners are able to use fine-grained variability (e.g. vowel nasalization) in an L2 in a similar way as listeners in their L1 (Desmeules-Trudel and Zamuner, 2019; McMurray et al., 2002, 2008). Based on previous reports, L2 listeners would be expected to have difficulties with both long-nasalized (i.e. phonological nasal) and short-nasalized (i.e. coarticulatorily nasalized) vowels when listening to (Canadian) French variably-nasalized vowels, given that all nasalized vowels are categorized as the same “sound category” in L1 English.

More specifically, we expect that L2-French listeners will be fairly sensitive to variations in vowel nasalization, since previous investigations have demonstrated sensitivity to this phonetic cue in their L1 (i.e. English), and that L2 listeners have previously been shown to consider phonetic details for word recognition (Qin et al., 2019). However, it is also possible that L1-English listeners will not be sensitive to variations in vowel nasalization in L2-French, based on the observation that recognizing words in an L2 is often more difficult than in the L1 (Weber and Broersma, 2012) and that gradient recognition likely requires substantial effort to access fine-grained representations (Desmeules-Trudel and Zamuner, 2019). Thus, gradient word recognition patterns that consider very fine-grained phonetic information (e.g. à la McMurray et al., 2002, 2008) are not expected across the board for L2 listeners. We also expect a bias towards interpreting nasalized vowels as a sequence of an oral vowel followed by a nasal consonant, since it is the only possible configuration in the listeners’ L1.

3 Age of L2 acquisition of a phonetic system

Many additional external (i.e. listener-specific) factors have been shown to influence speech perception abilities in an L2 (Flege and Bohn, 2021). For instance, researchers have observed that L2 speech perception abilities can benefit from earlier age of acquisition (Flege, 1995; Flege and Mackay, 2004; Wode, 1994; Yamada, 1995). In their investigation, Flege and Mackay (2004) tested several groups of L2 learners of Canadian English (Italian natives) and found that early learners, who arrived young in Canada and had an early age of exposure (AoE) to English, performed better on an L2 vowel discrimination task than learners who arrived in Canada at a later age and started acquiring
L2 English later. This suggests that, among other factors, earlier L2 learning combined with a sufficient amount of good-quality input (Flege and Bohn, 2021) improves perceptual discrimination abilities in an L2.

In terms of learning, the seminal Speech Learning Model (SLM; Flege, 1995) and its revised version (SLM-r; Flege and Bohn, 2021) postulated that learners from a variety of backgrounds can learn and use fine-grained phonetic details in an L2 (Flege and Hammond, 1982). In the initial model, this capacity was attributed to L2 learners’ ability to exploit the same mechanisms as for L1 acquisition, such as the extraction of distributions of L2 sounds to create abstract representations, and that the ability to create new categories in the L2 decreased as learners’ age of first exposure increased. However, focusing on L2 learners who use their L2 in their environment and are learning it naturally, Flege and Bohn’s (2021) SLM-r refined this hypothesis, stating that input is crucial to the specification of L2 categories, more than age of acquisition, since the critical period hypothesis cannot explain age-related differences in speech perception. It also states that L2 phonetic categories could potentially include as much details as the L1’s. Therefore, age of first exposure alone cannot motivate the establishment and mastery of L2 phonetic categories, but rather it needs to be combined with the amount of L2 use and exposure. For instance, calculating a full-time equivalent score by considering both length of residence in an L2 setting (i.e. with immigrant participants) and percentage of daily L2 use, Flege and Bohn (2021) argue that such a combination might better explains L2 speech learning phenomena. Concretely, higher full-time equivalent scores likely better predict native-like attainment than AoE. In other words, the amount and quality (Sabourin et al., 2016) of input may be a more accurate factor to explain perceptual and productive abilities for L2 sounds.

In our study, as a starting point for further research on the impact of L2 listeners’ language background on real-time word processing, we explored the influence of AoE on the recognition of spoken words in our bilingual group. Assuming that early learners received a greater amount of input, because they were exposed to L2-French early in their lives, we expect that they will be able to better identify words that have phonologically-nasal and coarticulatorily-nasalized vowels, as opposed to late learners who are expected to identify more instances of the stimuli as coarticulatory-nasalized (i.e. more “English”-like because these are the only types of nasalized vowels in English), mirrored by word choices and eye movement patterns. In other words, on the one hand, early-AoE L2-French listeners are expected to have a more robust nasal vowel category in their L2 phonological system due to more extensive exposure. We predict that they will be able to better categorize stimuli that have long-nasalized vowels. On the other hand, late-AoE L2 listeners are expected to categorize stimuli as instances of oral vowels followed by nasal consonants more than early-AoE listeners, due to the fact that limited input has not enabled them to develop a robust nasal vowel category in French. It is important to note, however, that some of the used stimuli were ambiguous (e.g. long-nasalized vowel with a short consonantal appendix), and that L1 listeners did recognize them as ambiguous (Desmeules-Trudel and Zamuner, 2019). Therefore, it is possible that early-AoE L2 listeners will also be highly sensitive to fine-grained phonetic details and will have difficulties recognizing these ambiguous stimuli, as opposed to late-AoE L2 listeners who are not as sensitive and therefore will not encounter as many recognition difficulties.
We also present a reanalysis of the data in Desmeules-Trudel and Zamuner (2019), in which L1-French listeners completed the same experiment as the current L2 listeners, in order to provide baseline results. For this group, we assessed the impact of variations in nasalization duration on the vowel but not AoE since, by definition, L1 listeners are all exposed early to the target language. In Desmeules-Trudel and Zamuner (2019), ambiguous stimuli containing both a nasalized vowel and a nasal appendix yielded gradient word recognition patterns, suggesting that word representations contained fine-grained phonetic information. However, other stimuli that were unambiguous led to more categorical patterns of word recognition, suggesting that inconsistent phonetic cues within words motivate the use of fine-grained phonetic information during word recognition. Comparing L1 and L2 listeners of French provides a more comprehensive picture of how L2 listeners can use fine-grained phonetic details for word recognition, and enables us to evaluate if L2 listeners can also use variable phonetic information in a gradient manner like L1-French listeners.

II Experiment 1

Experiment 1 focused on listeners’ word recognition patterns with stimuli that contained ambiguous phonetic information. Stimuli in Experiment 1 had a word-final nasal appendix (e.g. pain [pɛ̃ɲ] “bread”), which was included to reflect one of the possible realizations of phonological nasal vowels and to create some ambiguity in the stimuli (Desmeules-Trudel and Zamuner, 2019). This manipulation is expected to bias L2 participants’ responses towards words that have a sequence of an oral vowel and a nasal consonant (e.g. peigne [pɛɲ] “comb”). Section III presents a second experiment with stimuli that do not have a nasal appendix.

I Method

a Participants. Listeners were 44 speakers (33 female, 10 male, 1 undisclosed) of Canadian English who spoke French as an L2 at various levels of (self-reported) proficiency, aged between 17 and 27 years old ($M = 20.2$, $SD = 2.9$). According to self-reports on a background questionnaire, participants were exposed to French on average (i.e. weekly in recent years) 18.5% of the time in a school context, 9.6% in a work context, 3.1% in the family, 7.3% with friends, and 1.8% of the time with an intimate partner. All were residents of the Ottawa (Ontario, Canada) area which neighbors the province of Québec, specifically the majority French-speaking city of Gatineau, therefore were mostly exposed to L2 Canadian French (as opposed to Hexagonal French). These participants did not speak another language that had phonological nasal vowels in its phonological inventory. Two additional listeners were tested, but could not be analysed for failing to calibrate ($n = 1$) and equipment error ($n = 1$). Data collection stopped after testing 44 participants based on the fact that previous papers investigating similar questions have used fewer or approximately the same number of participants (20 in Weber and Cutler, 2004; 50 in Escudero et al., 2008).

All 44 participants completed a language-background questionnaire, and were asked to fill out a table detailing the proportion of weekly exposure to L2 French between the ages
of 0 and 5 years old, during elementary school, during high school, and after high school. Based on the responses, listeners were divided into two groups: early bilinguals (30% or more exposure to L2 French before elementary school) and late bilinguals (less than 30% exposure to L2 French before elementary school or overall, or onset of significant exposure to L2 French after 5 years of age). It is also interesting to note that listeners who were significantly exposed to L2 French before 5 years old were also exposed to their L2 in a more naturalistic setting overall (e.g. with their family), while late bilinguals were exposed to French in a classroom setting after age five, after starting formal education (Sabourin et al., 2016). Seventeen were early bilinguals and 27 were late bilinguals.

Data from 23 L1-French listeners are also presented below (published data from Desmeules-Trudel and Zamuner, 2019). These participants were all native speakers of Canadian French, aged between 18 and 36 years old ($M = 23.3$, $SD = 5.3$). All French listeners were bilingual in English, and some also spoke other languages. None of them spoke another language with contrastive nasal vowels in its phonological inventory. One additional L1 listener completed the experiment, but failed to properly complete the consent form, therefore their data were rejected from the analyses.

b Stimuli and experimental conditions. We used nine triads of common monosyllabic and picturable French words (see Appendix 1) and corresponding sound clips of the words (see below). On each trial, participants saw four images on a display (Figure 1, but

![Figure 1. Illustration of a display, as seen by participants in the eye-tracking experiments, and detailing the words and phonetic transcriptions (not seen by participants).](image)
without the orthographic labels and phonetic transcriptions), corresponding to words with a phonological nasal vowel (CV; *pain* [pɛ̃] “bread”), words with an oral vowel followed by a nasal consonant (CVN; *peigne* [pɛɲ] “comb”), words with an oral vowel followed by an oral consonant (CVC; *pêche* [pɛʃ] “peach”), and additional words that were used as fillers (e.g. *pluie* [plɥi] “rain”). Fillers had the same onset consonant as the other words (except for *feu* associated with *vin–veine–verre* and for *nid* associated with *dent–dame–dattes*), but a different nucleus. Details about the images for individual trials as well as an example of trial orders can be found in the dedicated Open Science Framework (OSF) repository: https://osf.io/rsh3e.

We investigated nasal [ã, ë̃, ɔ̃] vowels and their oral counterparts, but did not include the nasal [œ̃] due to its low functional load (Martin et al., 2001). Since we were interested in how listeners process nasalized vowels, we focused our analyses on CṼ and CVN words only. Visual referents of the critical sets (*n* = 27 words) and unrelated or filler words (*n* = 37, for a total of 64 images) were extracted from the International Picture Naming Project (Székely et al., 2004) when available in the database, or hand-drawn by a professional artist in the same style as the other images and scanned. Images were 220 × 220 pixels (approximately 56 × 56 mm on the screen).

Auditory stimuli were recorded by five native speakers of Canadian French. Speakers included two females and three males, were aged between 23 and 27 years old, and originated from Southwestern Ontario (1), Eastern Ontario (1), Québec (1), New Brunswick (1), and British Columbia (1). For recording, the words were embedded into meaningful carrier sentences at the end of an intonational phrase for consistency across targets. Each sentence was read twice by each speaker. The recordings were hand-segmented in Praat (Boersma and Weenink, 2015) and the resulting target words were normalized to a 70 dB amplitude value. The CṼ and CVN tokens were compared for each speaker, and the most similar pairs were selected based on the observation of vowel spectrograms (with the idea of avoiding abrupt formantic or pitch transitions at cross-splicing points within the stimuli during the cross-splicing procedure, which could have resulted in clicks in the signal) and auditory confirmation by the experimenter. We made sure that the vowels in the CVN tokens were not auditorily nasalized within the first 80% of their duration based on the observation of the spectrum, specifically the absence of nasal antiformants and nasal peaks in the 800–1,500 Hz region, as well as auditory confirmation by the experimenter (a native speaker of Canadian French). Each speaker contributed one or two word pairs in the experimental stimulus set.

For each experimental stimulus, the consonant frame of the CVN token was used as the basis for cross-splicing. For each cross-spliced stimulus, we removed a part of the oral vowel from the CVN word at vowel offset, and replaced it with a part of a CṼ vowel token towards vowel offset. We considered zero-crossings on the waveform to avoid clicks or noises in the signal and reduplicated amplitude peaks to adjust the vowel duration when necessary. The section of the Ṽ vowel that was pasted into the matrix was the same duration as the part of the V(N) vowel that was removed. For example, if the last 50 ms of the vowel from *peigne* [pɛɲ] “comb” were removed, they were replaced by the last 50 ms of the vowel from *pain* [pɛ̃] “bread”.

The cross-splicing procedure led to four experimental conditions (i.e. proportion of the CVN vowel that was replaced with a part of a phonological nasal Ṽ – %NasDur):
partially nasalized vowels (20N, 50N, 80N) and fully nasal vowels (100N). For the 20N, 50N and 80N conditions, the duration of the oral (CVN) vowel was calculated. Then, 20%, 50% or 80% of the vowel duration was cut from the original vowel (CVN) and replaced by a part of the nasal vowel of the matching CV \tilde{V} token. For 100N %NasDur, the oral vowel of the original CVN token was entirely replaced by the vowel of the matching CV \tilde{V} token. The final nasal consonant from the CVN matrix was shortened to 50 ms in order to mimic the realization of nasal vowels’ appendices (Desmeules-Trudel, 2015; Desmeules-Trudel and Brunelle, 2018). An illustration of a spectrogram of the dame “lady” – dent “tooth” stimulus in the 50N condition is presented in Figure 2. A total of 36 stimuli (9 word sets \times 4 %NasDur conditions) were analysed for the current article. Durational details about the stimuli are found in Appendix 2, and sound files and target images are found online in the OSF repository.

c Design, procedure and data extraction. The experiment was programmed and presented in Experiment Builder 1.10.63 (SR Research), and used an EyeLink 1000 (SR Research), arm mount with chin rest, monocular recording, and sampling at 500 Hz. Eye-tracking methodologies have the advantage of providing indications on real-time processing of linguistic information, e.g. assuming that listeners will direct their gaze towards pictures depicted on a display as they hear utterances with references to the visual targets. By calculating the amount of time spent fixating to the images on the display (e.g. target, competitor and filler images) through time, one can infer real-time lexical access, a principle known as the linking hypothesis (Allopenna et al., 1998), in addition to standard measures of language processing such as accuracy (e.g. mouse clicks) and reaction times.
The experiment started with a five-point calibration, then validation, keeping the maximum and average errors below 1° of visual angle for all participants. Listeners first completed a familiarization phase, in which they heard an unspliced version of each experimental and filler words once each (n = 70 trials) while they saw the associated images. For each word, they had to rate their degree of familiarity (“unknown”, “fairly familiar”, “familiar”, “very familiar”) with the meaning of this word. If listeners did not know one of the words within a pair (“unknown” response), trials where the unknown word was a possible image choice were removed from the analyses post-hoc (e.g. if participants did not know the word *peigne* [pɛɲ] “comb”, all trials where *pain* [pɛ̃] “bread” or *peigne* were a possible choice were removed).

For the word recognition task, images corresponding to each word triad (one CṼ, one CVN and one CVC) as well as an additional distractor image were arranged together on a display and presented in a Visual World Paradigm experiment (Allopenna et al., 1998; Huettig et al., 2011). We assume that the classical linking hypothesis on the relationship between eye fixation patterns and lexical activation holds. For a critique of the linking hypothesis in the Visual World Paradigm, see Teruya and Kapatsinski (2019). The four images were embedded within approximately 60 × 60 mm (250 × 250 pixels) interest areas for data collection. Image position was randomized across trials (e.g. CVN images were not always found in the top-right corner, but could rather be found in any position across trials). For each trial, drift check was first performed, and participants then saw the four images on the screen for 500 ms. The auditory stimulus in one of the experimental conditions (or filler word) was then played, and participants were asked to click on the image that corresponded to the heard word. Each experimental auditory stimulus was presented twice to ensure that enough trials per listener were included in the analyses. Participants completed 72 experimental trials, 74 filler trials (with four unrelated images on the display, with no CṼ or CVN choices), and 18 trials from an experimental condition that was rejected from the analysis post-hoc (i.e. a condition for which nasalization was absent from the vowel, but in which the place of articulation of the final consonant in CVN and CVC words did not always match and induced other unwanted coarticulatory effects on word recognition) for a total of 164 trials. Trial order was randomized, and block order presentation was counterbalanced over four different lists across participants. A document with a sample trial order (for the familiarization and word recognition tasks) is found in the OSF repository.

A trial-rejection procedure was first applied based on participants’ judgments during the familiarization phase. Out of a possibility of 3,168 experimental trials (9 pairs × 4 %NASDUR conditions × 44 participants × 2 repetitions), 656 trials were removed (20.7% of the trials; 82 pairs × 4 %NASDUR conditions × 2 repetitions). A total of 71 pairs were rejected for late-AoE bilinguals, and 11 pairs for early-AoE bilinguals. The rejection rate is partly due to the fact that we could not control for word frequency because our stimuli were restricted to monosyllabic CṼ-CVN minimal pairs of picturable words. Since L2 listeners likely have better knowledge of high-frequency words in general, some of the low-frequency used items were not as well known by the group of listeners and had to be rejected from the analyses. We were left with 2,512 trials in total for analysis in Experiment 1. In addition to removing trials involving unknown words to L2 listeners, we removed trials for which participants had reaction times (i.e. between vowel onset
and mouse click; RT) that were 3 standard deviations above or below the grand RT mean (176/2,512 trials, 7% of the L2 data), leaving 2,336 trials in total.

Experiment Builder automatically recorded eye movements, and fixations to the images were extracted in 50-ms time bins between vowel onset and mouse click using SR Research DataViewer and a Python script provided by the company. We also computed which image the participants fixated to the most in the last 1,000 ms of each trial in order to obtain an image choice measure. This was necessary because for some participants ($n = 16$ in Experiment 1, and $n = 4$ in Experiment 2), the mouse clicks did not compile properly due to an error in programming. The computed chosen image measure corresponded to the mouse click data in 95.5% of cases for the participants who completed the corrected version of the experiments ($n = 28$ in Experiment 1, with 96.5% matching responses-chosen images, and $n = 40$ in Experiment 2, with 94.8% matches). For L1 listeners, the chosen image measure corresponded to the mouse clicks in 98% of cases.

2 Results

a Chosen images. Statistical analysis of chosen images was conducted with generalized mixed-effects models (GLMERs), using lme4 1.1-26 (Bates et al., 2015) in R version 4.0.3 (R Core Team, 2017). We investigated the effects of $\%NasDur$ (20N, 50N, 80N and 100N) for both L2 and L1 listeners, AoE (early and late) for L2 listeners only, and Vowel Quality ([a], [ɛ] and [ɔ]) on the probability of choosing CV ̃ images. Models were run separately for L2 and L1 listeners. A backwards step-wise model comparison procedure was used: starting with a fully specified model (e.g. $\%NasDur*AoE + Vowel Quality$), a second model was run after removing the factor with highest p-values (e.g. removing vowel, leaving $\%NasDur*AoE$ only) and compared to the first model using a chi-square test (anova() function). Lack of significance ($p > 0.05$) warranted the exclusion of the factor under investigation (e.g. Vowel Quality) for further model comparisons. All possible comparisons were performed. We illustrate below only the differences between 50N (baseline of the $\%NasDur$ factor) and the other levels of $\%NasDur$, since this condition constitutes the boundary in production between phonologically-nasal and coarticulatorily-nasalized vowels in Canadian French. Detailed analysis code and data can be found in the OSF repository.

In both groups, sensitivity to fine-grained phonetic variability is expected to yield gradient patterns of word recognition, i.e. multiple significant differences between the category boundary (50N $\%NasDur$) and other levels of the $\%NasDur$ factor would indicate gradient sensitivity to phonetic variability. However, if the proportions of CV ̃ choices are significantly different across 50N and 20N stimuli but not with the other levels of $\%NasDur$, this would suggest a more categorical pattern of recognition, i.e. that listeners do not use fine-grained phonetic variability to recognize spoken words past the category boundary. Note that these possibilities on how gradient word recognition will be expressed in the data are the same for both the L1-French and L2-French groups, although we expect smaller proportions of CV ̃ choices in L2 listeners overall.

Figure 3 shows the proportions of CV ̃ image choices in L2 (panel A) and L1 listeners (panel B) for each $\%NasDur$ condition. For L2 listeners (Figure 3A), $\%NasDur$ was the only significant factor according to our model fitting procedure. AoE and Vowel Quality
were not significant and are thus not discussed further in this section. Model estimates revealed that only 50N was significantly different from 20N for L2 listeners (full model tables can be found in the supplementary materials; baseline $\beta = -0.29$, 20N $\beta = -0.78$, $p < 0.001$), which yielded significantly lower probability of choosing the CṼ image. This suggests that L2 listeners are aware of the difference between short-nasalized (20N) and long-nasalized vowels (50N or more) in Canadian French, but indicates a more categorical way of using nasalization duration for discriminating words in L2 listeners past the category boundary.

For L1 listeners (Figure 3B), $%\text{NASDUR}$ also had a significant effect on their proportions of CṼ choices. Model estimates showed that they were able to gradiently use variations in nasalization duration to recognize spoken words, as shown by significant differences between the 50N baseline and 20N (baseline $\beta = -0.83$, 20N $\beta = -1.12$, $p < 0.001$) as well as between 50N and 80N (80N $\beta = 0.35$, $p = 0.04$) and between 50N and 100N (100N $\beta = 0.69$, $p < 0.001$). Thus, the probability of choosing an image representing a word with a phonological nasal vowel gradiently increased as the duration of nasalization on the vowel increased in L1 listeners, as opposed to L2 listeners who displayed a more categorical pattern.

b Eye movements. Eye movements were analysed using generalized additive mixed-effects models (GAMMs; Wood, 2017). GAMMs enable one to analyse and model non-linear data, such as most eye movement datasets. Both continuous and categorical factors can be included in the models, in addition to random effects. Auto-correlation of the data is also considered, which refers to the fact that one given data point in time is necessarily correlated to the preceding data point in the time series (Baayen et al., 2018), and can be empirically determined based on the raw data. GAMMs also handle missing data points, which are common in eye-tracking research since listeners are free to look away from the images on the display at any point. Recent accounts that use GAMMs for psycholinguistic research include Desmeules-Trudel et al. (2020), Desmeules-Trudel and Zamuner (2019), Porretta et al. (2016, 2017), Tremblay and Newman (2015), van Rij et al. (2016), amongst others. We refer the reader to Porretta et al., (2017) for a comprehensive review and tutorial to fit GAMMs on eye-tracking data. For model fitting, comparison and visualization, we used mgcv 1.8-33 (Wood, 2017) and itsadug 2.3 (van Rij et al., 2017) in R.

We separately modeled empirical-logit-transformed proportions of fixations to the CṼ and CVN images, a transformation that is necessary to make the data unbounded for non-logistic models (Porretta et al., 2017), as the dependent variable. (Fixation patterns to the CVC and Filler images are presented in Figure S5 in the supplemental materials found in the OSF repository for Experiment 1, and Figure S10 for Experiment 2.) The model fitting procedure started with a baseline model containing $%\text{NASDUR}$ conditions and $\text{TIME WINDOW}$ of analysis (1,000ms, starting 200ms after target vowel onset to account for eye movement programming delay; Fischer, 1992), as fixed factors, in addition to AoE as a parametric factor and individual smooth terms through $\text{TIME}$ for each level of AoE, as well as $\text{VOWEL QUALITY}$ as a parametric factor and individual smooth terms through $\text{TIME}$ for each level (see code in OSF repository). We also empirically determined autocorrelation of the data and included it in the model formula as $\text{ARI}$. For model comparison, other models were fitted in the same way as for the GLMERs above, removing one factor at a
Figure 3. Effect of %NasDur on the probability of choosing the CṼ word for L2 (A) and L1 (B) listeners in Experiment 1. Notes. Bars represent the overall average of CṼ image choice, with the standard error to the mean as whiskers. **** p < 0.0001. * p < 0.05. n.s. = nonsignificant.
levels of the %NasDur factor). Note that we present different image choices separately since the target image varied across trials depending on the image choice (CVN images in Figure 5 below). Recall that overall, participants chose the CV image less than the CVN images, reflected in the proportion of image choices in Figure 3. This is important to keep in mind when looking at the eye-tracking data in Figures 4 and 5, which also include the number of analysed trials.

GAMMs determined that when CṼ was the chosen image, L2 listeners looked significantly more to the CṼ target in the 50N %NasDur condition than the 20N condition towards the end of the time window (i.e. after 998 ms), less to the CṼ target in the 50N condition than 80N condition (i.e. between 453 and 776 ms) as well as the 100N condition (i.e. 372 ms to 998 ms). This suggests that L2 listeners are gradiently sensitive to variations in nasalization duration in French, even though these sensitivities did not appear in the image choice data. Further differences also emerged regarding the vowel quality factor, suggesting that listeners fixated significantly more to the CṼ target for [a] and [ɛ] words than for [ɔ] words at the beginning of the time window of analysis (0–343 ms and 0–434 ms intervals respectively), suggesting that L2 listeners identified non-back vowels as nasals more quickly than back vowels (see Figure S1 in supplementary materials). Analysis of the L1 data reveals a slightly different pattern from L2 listeners, demonstrating that the only significant difference is between 50N–100N, thus that L1 listeners did not display as much gradience for word recognition in their eye movement patterns, although they did for their image choices, contrary to L2 listeners. Regarding the vowel quality factor, the GAMM analysis revealed that L1 listeners fixated significantly more to the [ɔ] words than [a] words between 424 and 737 ms within the time window of analysis, than [ɛ] during short time intervals (172–272 ms, 455–727 ms, and
within the time window, and more to [a] words than [ɛ] words at the end of the time window (909–1,000 ms).

Observing the fixations to the CVN image (Figure 5), early in the trial, fixations to the CVN targets are slightly higher in the 50N condition than the other %NasDur condition for L2 listeners. GAMM analysis revealed that early within the time window, there were significantly more fixations to the CVN target in the 50N %NasDur than the 20N (up to 493 ms) and 100N %NasDur (up to 655 ms) conditions for L2 listeners. However, it is unlikely that they would have recognized the target at that point in time (i.e. maximum 655 ms after word onset) and given the direction of the effect in the 50N–20N comparison, these differences might not be meaningful with regards to gradient vs. categorical patterns of word recognition, but rather due to (normal) noise in the eye movement data. Additional differences emerged for vowel quality, overall suggesting that [ɛ] words were fixated to less than [a] (483–1,200 ms) and [ɔ] words (210–1,200 ms). For L1 listeners, however, the significant difference between 50N and 80N (321–523 ms) and 100N %NasDur (351–1,018 ms) conditions suggests that listeners fixated significantly less to the CVN image when the vowel was nasalized for a longer period of time, supporting the idea that they are highly sensitive to fine-grained variations in nasalization duration. Differences across vowel qualities were also found: L1 listeners fixated more to the [a] (665–1,200 ms) and [ɛ] (564–1,200 ms) words than to [ɔ] words, and more to [ɛ] words than [a] words between 836–1,160 ms. More details about the effects of vowel quality in both L1 and L2 listeners can be found in the supplementary materials.

Model comparisons revealed that AoE also significantly contributed to better model fit in L2 listeners, for both CV choices (NasDur + vowel VS. NasDur + vowel + AoE;
Δ in ML $\chi^2$ score ($\Delta df = 8$) = 13.13, $p < 0.001$) and CVN choices ($\text{NASDur + vowel vs. NASDur + vowel + AoE}$; Δ in ML $\chi^2$ score ($\Delta df = 8$) = 36.11, $p < 0.001$). Raw fixation curves and results of the GAMM, shown in Figure 6, suggest that late-AoE L2 listeners fixated more to the $\text{CV} \tilde{V}$ target overall than early-AoE L2 listeners (523–1,200 ms). This was unexpected, since early learners would be expected to have more robust L2 lexical representations and thus fixate more to $\text{CV} \tilde{V}$ words. However, stimuli were ambiguous in Experiment 1 (i.e. combination of a nasalized vowel and a nasal appendix that can be perceived as a nasal consonant). It is thus possible that early-AoE listeners were more sensitive to the ambiguity and did not fixate to the target as much as late-AoE listeners, causing the unexpected pattern to emerge for both $\text{CV} \tilde{V}$ and CVN images. In other words, the presence of an appendix following a nasalized vowel was confusing if the word was interpreted as $\text{CV} \tilde{V}$, the presence of a nasalized vowel was confusing if the word was interpreted as CVN, and early learners were more sensitive to these phonetic cues than late learners.

### III Experiment 2

In Experiment 2, a second group of L2 listeners completed a similar experiment, but the word-final nasal appendix was removed from the stimuli. This was done to eliminate the bias towards CVN images, but also led to another set of ambiguous stimuli. Indeed, the short-nasalized 20N stimuli does not correspond to a canonical realization of a $\tilde{V}$, since it is only nasalized for a short period of time towards vowel offset. Therefore, image choices and fixations to the target were expected to be lower in the 20N %$\text{NASDur}$ condition than the other %$\text{NASDur}$ conditions in L2 listeners. As for AoE, we expect a greater probability of choosing the $\tilde{V}$ words in early-AoE L2 listeners, as displayed by both image
choices and eye movements. We also present below the L1 data from Desmeules-Trudel and Zamuner (2019), whose participants completed the same experiment.

1 Method

a Participants. Participants were 44 native speakers (38 female, 6 male) of Canadian English who spoke French as an L2. They were aged between 17 and 29 years old (\(M = 18.9, SD = 2\)) and did not speak another language that had phonological nasal vowels in its phonological inventory. According to self-reports on the questionnaire, participants were exposed to French on average (i.e. over a typical week) 26.8% of the time in a school context, 8.4% in a work context, 4.7% in the family, 9.1% with friends, and 6.4% of the time with an intimate partner. Sixteen were early bilinguals and 28 were late bilinguals.

Twenty-four L1 listeners, all native speakers of Canadian French, aged between 17 and 32 years old (\(M = 21.5, SD = 4.4\)) also completed the study. They were all bilingual in English, and some also spoke other languages. None of them spoke another language with contrastive nasal vowels in its phonological inventory.

b Stimuli and experimental conditions. Stimuli and conditions were the same as in Experiment 1, except for the fact that the word-final nasal appendix was removed for all tokens.

c Procedure, analyses and variables. The procedure used in Experiment 2 was the same as in Experiment 1, including the familiarization phase and the testing phase. Out of 3,168 experimental trials, 608 trials were removed based on the familiarization results (i.e. when participants reported not knowing one member of the target pairs; 19.2% of the trials; 76 pairs \(\times\) 4 %\(\text{NasDur}\) conditions \(\times\) 2 repetitions). Of these, 65 pairs were rejected for late bilinguals, and 11 pairs for early bilinguals. Trials for which participants gave RTs that were 3 standard deviations above or below the grand mean were also removed (261/2,560 trials, 10.2%).

d Analysis and variables. The analysis procedure was also the same as Experiment 1, using GLMERs and GAMMs to assess the influence of %\(\text{NasDur}\) for image choices and eye movements in L1 and L2 listeners, time (for eye movements only), AoE (early-AoE vs. late-AoE for L2 listeners’ data only), and vowel quality ([a], [e] and [ɔ]).

2 Results

a Chosen images. Let us recall that Desmeules-Trudel and Zamuner (2019) did not find a gradient pattern in L1 listeners with the current stimuli. Instead, L1 listeners recognized words categorically. Therefore, we did not expect L2 listeners to display differences between the baseline 50N condition and words with vowels that are nasalized for a longer proportion of their duration.

The probability of choosing CṼ images in L2 (panel A) and L1 listeners (panel B) are presented in Figure 7. Results of the model comparison procedure revealed that %\(\text{NasDur}\) and AoE both contributed to significantly better model fit for L2 listeners. The only level of the %\(\text{NasDur}\) factor that was significantly different from 50N was 20N for L2
listeners (baseline $\beta = 1.43$, 20N $\beta = -0.64$, $p < 0.001$), which yielded significantly lower probability of choosing the CṼ image. Early-AoE L2 listeners (Figure 7C) also had a higher probability of choosing the CṼ image than late-AoE L2 listeners (Early AoE $\beta = 1$, $p = 0.02$).

L1 listeners' image choices were also significantly impacted by %NasDur, and the model results revealed a significant difference between the 50N and 20N (baseline $\beta = 3.38$, 20N $\beta = -1.24$, $p < 0.001$). The direction of the effect was similar to L2 listeners. Thus, both groups of listeners were aware of the difference between short-nasalized
Second Language Research 00(0)

20

and long-nasalized vowels (50N or more) in Canadian French with these stimuli. This suggests a more categorical way of using nasalization duration for discriminating words when nasalized vowels are not followed by a nasal appendix (i.e. a short nasal consonant), i.e. when stimuli are not ambiguous.

b Eye movements. Proportions of fixations to each image for Experiment 2 L1 and L2 listeners are presented in Figure 8. Panel A shows the fixations to the CṼ image for CṼ choices, and the results of the difference curve analyses show that early in the trial, both the 80N (200–503 ms) and 100N %NasDur (200–311 ms) conditions yielded significantly higher proportions of fixations to the CṼ targets than the 50N condition. The lack of significant difference between 50N and the 80N–100N condition after 503 and 311 ms post-word onset, respectively, along with observation of the raw fixation curves suggest similar processing patterns. The word recognition patterns observed in Figure 8A are consistent with the hypothesis that L2 listeners are sensitive to fine-grained phonetic information (i.e. significant differences early in the trial, followed by similar recognition patterns). Furthermore, starting 553 ms after word onset, stimuli from the 20N %NasDur condition yielded significantly lower proportions of fixations to the CṼ target until the end of the trial (also between 200 and 250 ms, not displayed on Figure 8A; see supplementary materials). This again is consistent with the hypothesis that L2 listeners are sensitive to timing information in the stimuli. For L1 listeners, word recognition patterns are similar to L2 listeners, perhaps with a slight delay in stimulus recognition for some %NasDur conditions which might be due to slight differences in baseline fixations to the target before the 200 ms eye movement programming delay. However, note that differences in L1 listeners’ fixations to CṼ words in the 20N %NasDur condition vs.

Figure 8. Proportions of fixations to the CṼ image for CṼ choices for L2 (A) and L1 (B) participants in Experiment 1. Notes. Vertical broken lines indicate the window of analysis for the GAMMs, and horizontal line represent time intervals for which one level (see corresponding color and annotation above lines) of the %NasDur factor was significantly different from 50N (orange circles).
50N words is greater (i.e. longer time window) than for L2 listeners, suggesting greater sensitivity to the phonetic differences. For both groups, differences emerged across vowel qualities (see Figures S6 and S7 in the supplementary materials), suggesting that early within trials, L2 listeners fixated more to [a] and [ɛ] words than to [ɔ] words. However, later within trials (approximately 900 ms post-onset for L2 listeners) and for L1 listeners (for a short period, between approximately 450–600 ms for [a] words and 500–800 ms for [ɛ] words), participants fixated significantly more to [ɔ] words. Furthermore, for L2 listeners, they fixated significantly more to the [ɛ] words than to the [a] words between approximately 500 ms and 700 ms post-onset.

The GAMM analysis also revealed that AoE significantly contributed to the model for fixations to CṼ targets, thus that it had an impact on L2 listeners’ fixation patterns ($\text{NasDur + vowel VS. NasDur + vowel + AoE}$; $\Delta$ in ML $\chi^2$ score ($\Delta df = 8$) = 77.09, $p < 0.001$). In the early part of the time window of analysis (between 432 ms and 715 ms), late-AoE L2 listeners fixated significantly more to the CṼ target in a way similar to Experiment 1 late-AoE L2 listeners (Figure 9). This could reflect uncertainty in word interpretation for the early-AoE group early in the trial. However, at the end of the time window of analysis, early-AoE L2 listeners fixated significantly more to the CṼ target than late-AoE L2 listeners (1,028–1,200 ms). This suggests better recognition later in the trials for early-AoE L2 listeners, consistently with the word choice data (Figure 7C) and previous research on the topic.

![Figure 9. Proportions of fixations to the CṼ image for CṼ choices in L2 participants for Experiment 2 by AoE.](image)

Notes. Vertical broken lines indicate the window of analysis for the GAMMs, and horizontal line represent time intervals for which early and late AoE were significantly different in the time course of fixations.
Analyses of fixations to CVN images for CVN choices revealed that there were some time-dependent significant differences across stimuli corresponding to different %NasDur values (Figure 10). Note, however, that the number of analysed trials was lower than for other images analysed above (i.e. 486 trials for L2 listeners and 124 trials for L1 listeners). The difference curves showed that both the 20N and 80N %NasDur stimuli yielded significantly higher proportions of fixations to the CVN target in L2 listeners, for a short period of time (453–675 ms for 50N–20N and 382–655 ms for 50N–80N). For L1 listeners, only the 80N stimuli yielded significantly different proportions of fixations to the CVN target, starting 1,069 ms after word onset. Considering the low number of analysed trials and the fact that none of the auditory stimuli contained a nasal consonant, which is necessary to correctly interpret CVN words, it is difficult to extract generalizations from the current data. These image choices reflect an incorrect interpretation of the stimuli.

**IV Discussion and conclusions**

Experiments 1 and 2 demonstrated that L2 listeners of Canadian French (L1 English speakers) were able to use the variability in vowel nasalization duration to recognize words. This was expected, following Beddor et al.’s (2013) results that English listeners can use vowel nasalization timing in their L1 to recognize words. That is, since English listeners are sensitive to vowel nasalization in their L1, we predicted that they would be sensitive to vowel nasalization in their L2 (French). This was demonstrated by word choices and eye movement data in both experiments, and suggests that phonemes and words are activated based on the characteristics of the acoustic signal in an L1 as well as in an L2.
The expected effect of L2 listeners conflating vowels that were nasalized for a short duration (i.e. corresponding to coarticulatorily nasalized vowels in French) and those that were nasalized for a long duration (i.e. corresponding to phonological nasal vowels in French) to the same category was not as strong as expected in our data. This seems inconsistent with research on L2 lexical activation asymmetries and phonetic assimilation (Best, 1995; Best and Tyler, 2007; Weber and Cutler, 2004). Indeed, L2 listeners were overall good at identifying CV and CVN words and were highly sensitive to fine-grained variations in nasalization duration in a way similar to L1 listeners, as shown through eye movement patterns. The structure of the stimuli, which were tightly controlled for variations in nasalization duration, enabled us to demonstrate specification of fine-grained phonetic information in our group of L2 listeners. This suggests that an important amount of phonetic information can be specified in the L2 lexicon, and that exposure to an L2 contributes to phonetic specification. Furthermore, we found some differences across vowel qualities ([a–ã] vs. [ɛ–ɛ̃] vs. [ɔ–ɔ̃]), which are difficult to tease apart from duration of nasalization because they co-occur in production, but it seems like the dynamics of vowel quality also contributed to the perception of vowel nasality and nasalization in the context of spoken words.

Another aim of the study was to explore how age of first exposure to an L2 can impact word recognition patterns, both in terms of accuracy and real-time processing. In Experiment 1, we initially predicted that more robust lexical representations, which are likely in early learners, would yield to higher proportions of fixations to the target. However, we found the opposite: early learners fixated on the CṼ and CVN targets significantly less than late learners. One possible interpretation for the pattern is that early learners interpreted ambiguous stimuli (i.e. partially or fully nasalized vowel followed by a nasal appendix) with more caution, in a way similar to L1 listeners (Desmeules-Trudel and Zamuner, 2019), and resulting in fewer fixations to the targets. Late-AoE listeners could have attributed the Ṽ category to long-nasalized vowel without considering the presence of a nasal appendix. This would suggest greater sensitivity to phonetic details and stimulus ambiguity in early learners than late learners. Thus, L2 participants, especially those who had been exposed to French before 5 years old, were biased to interpreting nasalized vowels followed by a nasal appendix as instances of VN sequences, similarly to L1 listeners. This is consistent with the SLM-r hypothesis that more extensive exposure to L2 input leads to more fine-grained phonetic categories in the L2. Since we can be confident that early-AoE L2 listeners have heard more French tokens of nasal vowels than late-AoE L2 listeners, it is expected that their nasal vowel category be more refined, thus that they interpreted ambiguous stimuli with more caution and ended up fixating less to the target in Experiment 1.

When the stimuli were unambiguous in Experiment 2, the data was consistent with the initial predictions based on age of exposure (Flege, 1995). Early bilinguals tended to choose and fixate to the CṼ image more than late bilinguals, which suggest less confusion in the early-AoE L2 listeners as well as better learning of the French Ṽ category. The eye-tracking data revealed that early within the trials, late learners were slightly faster at directing their gaze towards the CṼ targets, but early learners caught up towards the end of the time window of analysis and ended up fixating more to the CṼ targets than late learners. This is also consistent with our prediction that late bilinguals would be biased
towards CVN images. If we assume that all listeners had at least basic knowledge of the nasality contrast in Canadian French, we can say that strength of the phonetic confusion effect and L2 word recognition abilities also depend on AoE as well as structural differences across languages within the L2 listeners’ systems (i.e. English and French).

To our knowledge, this is one of the first studies to evaluate L2 listeners’ sensitivity to within-category variability during word recognition. It would be interesting to extend this line of research to measures of voice onset time (VOT) as VOT is used differently in languages to express voicing contrasts (Lisker and Abramson, 1964). Moreover, VOT has been shown to gradiently impact recognition in the L1 (McMurray et al., 2002, 2008), i.e. listeners are able to detect slight variations in VOT for word recognition. Using VOT would also be a good way to investigate phonetic confusion in L2 listeners, since late learners would be expected to categorize stop consonants according to the VOT norms in their L1 and thus provide further evidence on the way L2 listeners are impact by fine-grained phonetic variability. In addition, using a set of words that are more frequent would enable us to better assess the questions of interest, notably because many of the words were unknown to the L2 listeners which led to a relatively high rate of trial rejection in the current experiments (approximately 26% of trials in Experiment 1, and 27% of trials in Experiment 2). In order to minimize the potential impacts of such high rejection rates on the results, it will be important to ensure that L2 listeners know the word meanings and forms in further research on this topic.

In conclusion, we found that spoken word recognition is dynamically mediated by fine-grained phonetic details in an L2, in addition to the significant impact of age of exposure on L2 word recognition abilities. Given the high prevalence of bilinguals in our communities, we hope to pursue this line of research by investigating further L2 contrasts (e.g. VOT contrasts across languages) or with listeners from different backgrounds (e.g. testing native speakers of Portuguese with the French nasality contrast, two languages for which the realization of nasalized vowels is different; Desmeules-Trudel and Brunelle, 2018), and further verify our predictions for L2 processing. The current proposal also offers an interesting line of research for modeling external factors (e.g. AoE) directly into the online word processing system. It will also be important to investigate other components such as the interactions between a multilingual’s languages at the phonemic and word levels, and shed light onto activation and inhibition mechanisms in an L2. This research significantly contributes to the knowledge surrounding L2 processing, especially since multilingualism and second-language learning are common and very often necessary for human communication, thus bringing models of cognition another step forward towards representing the actual operations of the human mind.

Acknowledgements

The research presented in this article benefited from numerous conversations and comments from Bob McMurray, Marc Brunelle, Laura Sabourin, Kevin McMullin and Marc Joanisse. We would also like to thank audiences at the 2017 Annual Meeting on Phonology and the 2020 edition of the Berkeley Linguistic Society, as well as Émilie Piché for her precious help with data collection and members of the Centre for Child Language Research and the Sound Patterns Laboratory at the University of Ottawa, Canada. All remaining errors are our own.
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: F.D.-T. benefited from doctoral scholarships from the Fonds de recherche du Québec – Société et Culture and the Joseph-Armand Bombardier program of the Social Sciences and Humanities Research Council of Canada for completing this research, and postdoctoral fellowships from the Fonds de recherche du Québec and the University of Toronto Mississauga for preparing the manuscript. T.S.Z. benefited from financial support from the Natural Sciences and Engineering Research Council of Canada.

ORCID iD
Félix Desmeules-Trudel https://orcid.org/0000-0002-8611-2569

References


Carignan C (2013) When nasal is more than nasal: The oral articulation of nasal vowels in two dialects of French. Unpublished PhD thesis, University of Illinois at Urbana-Champaign, IL, USA.


**Appendix 1.** List of word sets that were used as auditory and visual referents.

<table>
<thead>
<tr>
<th>Contrastive (\tilde{V}) word</th>
<th>IPA</th>
<th>Coarticulated VN word</th>
<th>IPA</th>
<th>Fully oral vowel word</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>bain</em> “bathtub”</td>
<td>[bɛ̃]</td>
<td>beigne “donut”</td>
<td>[bɛɲ]</td>
<td>bec “beak”</td>
<td>[bɛk]</td>
</tr>
<tr>
<td><em>pain</em> “bread”</td>
<td>[pɛ̃]</td>
<td>peigne “comb”</td>
<td>[pɛɲ]</td>
<td>pêche “peach”</td>
<td>[pɛʃ]</td>
</tr>
<tr>
<td><em>vin</em> “wine”</td>
<td>[vɛ̃]</td>
<td>veine “vein”</td>
<td>[vɛɲ]</td>
<td>verre “glass”</td>
<td>[vɛʁ]</td>
</tr>
<tr>
<td><em>camp</em> “campsite”</td>
<td>[kã]</td>
<td>canne “cane”</td>
<td>[kãn]</td>
<td>cave “basement”</td>
<td>[kav]</td>
</tr>
<tr>
<td><em>dent</em> “tooth”</td>
<td>[dã]</td>
<td>dame “lady”</td>
<td>[dam]</td>
<td>datte “date (fruit)”</td>
<td>[dat]</td>
</tr>
<tr>
<td><em>gant</em> “glove”</td>
<td>[ɡã]</td>
<td>gamme “musical scale”</td>
<td>[ɡam]</td>
<td>gare “bus/train station”</td>
<td>[ɡaʁ]</td>
</tr>
<tr>
<td><em>pont</em> “bridge”</td>
<td>[pɔ̃]</td>
<td>pomme “apple”</td>
<td>[pɔ̃m]</td>
<td>porc “pork”</td>
<td>[pɔʁ]</td>
</tr>
<tr>
<td><em>rond</em> “circle”</td>
<td>[rɔ̃]</td>
<td>thum “rum”</td>
<td>[rɔ̃m]</td>
<td>robe “dress”</td>
<td>[ʁɔb]</td>
</tr>
</tbody>
</table>
### Appendix 2. Durational details of auditory stimuli.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Original CVN duration (ms)</th>
<th>Final duration (ms)</th>
<th>Duration spliced / nasalized (ms)</th>
<th>Actual % spliced / nasalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bain_20N</td>
<td>194.3</td>
<td>184.9</td>
<td>42.1</td>
<td>22.77</td>
</tr>
<tr>
<td>Bain_50N</td>
<td>194.3</td>
<td>201</td>
<td>100.9</td>
<td>50.2</td>
</tr>
<tr>
<td>Bain_80N</td>
<td>194.3</td>
<td>201.2</td>
<td>159.3</td>
<td>79.17</td>
</tr>
<tr>
<td>Bain_100N</td>
<td>194.3</td>
<td>188.4</td>
<td>188.4</td>
<td>100</td>
</tr>
<tr>
<td>Camp_20N</td>
<td>189.4</td>
<td>188.8</td>
<td>41.1</td>
<td>21.77</td>
</tr>
<tr>
<td>Camp_50N</td>
<td>189.4</td>
<td>180.5</td>
<td>96.5</td>
<td>53.46</td>
</tr>
<tr>
<td>Camp_80N</td>
<td>189.4</td>
<td>204.1</td>
<td>165.3</td>
<td>80.99</td>
</tr>
<tr>
<td>Camp_100N</td>
<td>189.4</td>
<td>193.4</td>
<td>193.4</td>
<td>100</td>
</tr>
<tr>
<td>Dent_20N</td>
<td>177.8</td>
<td>182.2</td>
<td>36.3</td>
<td>19.92</td>
</tr>
<tr>
<td>Dent_50N</td>
<td>177.8</td>
<td>172.6</td>
<td>84.7</td>
<td>49.07</td>
</tr>
<tr>
<td>Dent_80N</td>
<td>177.8</td>
<td>183.6</td>
<td>144</td>
<td>78.43</td>
</tr>
<tr>
<td>Dent_100N</td>
<td>177.8</td>
<td>168.2</td>
<td>168.2</td>
<td>100</td>
</tr>
<tr>
<td>Gant_20N</td>
<td>199.8</td>
<td>197.6</td>
<td>42</td>
<td>21.26</td>
</tr>
<tr>
<td>Gant_50N</td>
<td>199.8</td>
<td>197</td>
<td>97.2</td>
<td>49.34</td>
</tr>
<tr>
<td>Gant_80N</td>
<td>199.8</td>
<td>207.5</td>
<td>163.9</td>
<td>78.99</td>
</tr>
<tr>
<td>Gant_100N</td>
<td>199.8</td>
<td>194.7</td>
<td>194.7</td>
<td>100</td>
</tr>
<tr>
<td>Pain_20N</td>
<td>192.5</td>
<td>196.4</td>
<td>41</td>
<td>20.88</td>
</tr>
<tr>
<td>Pain_50N</td>
<td>192.5</td>
<td>192.6</td>
<td>96.9</td>
<td>50.31</td>
</tr>
<tr>
<td>Pain_80N</td>
<td>192.5</td>
<td>188.5</td>
<td>153.3</td>
<td>81.33</td>
</tr>
<tr>
<td>Pain_100N</td>
<td>192.5</td>
<td>192.5</td>
<td>192.5</td>
<td>100</td>
</tr>
<tr>
<td>Pont_20N</td>
<td>159.6</td>
<td>159.5</td>
<td>35.7</td>
<td>22.38</td>
</tr>
<tr>
<td>Pont_50N</td>
<td>159.6</td>
<td>158.8</td>
<td>85.1</td>
<td>53.59</td>
</tr>
<tr>
<td>Pont_80N</td>
<td>159.6</td>
<td>159</td>
<td>122</td>
<td>76.73</td>
</tr>
<tr>
<td>Pont_100N</td>
<td>159.6</td>
<td>156.7</td>
<td>156.7</td>
<td>100</td>
</tr>
<tr>
<td>Thon_20N</td>
<td>176.6</td>
<td>185.4</td>
<td>33.8</td>
<td>18.23</td>
</tr>
<tr>
<td>Thon_50N</td>
<td>176.6</td>
<td>171.2</td>
<td>93.7</td>
<td>54.73</td>
</tr>
<tr>
<td>Thon_80N</td>
<td>176.6</td>
<td>182.4</td>
<td>144.8</td>
<td>79.39</td>
</tr>
<tr>
<td>Thon_100N</td>
<td>176.6</td>
<td>144.2</td>
<td>144.2</td>
<td>100</td>
</tr>
<tr>
<td>Vin_20N</td>
<td>156.4</td>
<td>157.9</td>
<td>31.1</td>
<td>19.7</td>
</tr>
<tr>
<td>Vin_50N</td>
<td>156.4</td>
<td>156.4</td>
<td>77.4</td>
<td>49.49</td>
</tr>
<tr>
<td>Vin_80N</td>
<td>156.4</td>
<td>156.1</td>
<td>130.6</td>
<td>83.66</td>
</tr>
<tr>
<td>Vin_100N</td>
<td>156.4</td>
<td>147.1</td>
<td>147.1</td>
<td>100</td>
</tr>
</tbody>
</table>