



# Gradient and categorical patterns of spoken-word recognition and processing of phonetic details

Félix Desmeules-Trudel<sup>1,2</sup> · Tania S. Zamuner<sup>1</sup>

Published online: 27 February 2019

© The Psychonomic Society, Inc. 2019, corrected publication 2019

## Abstract

The speech signal is inherently rich, and this reflects complexities of speech articulation. During spoken-word recognition, listeners must process time-dependent perceptual cues, and the role that these cues play varies depending on the phonological status of the sounds across languages. For example, Canadian French has both phonologically nasal vowels (i.e., contrastive) and coarticulatorily nasalized vowels, as opposed to English, which only has coarticulatorily nasalized vowels. We investigated how vowel nasalization duration, a time-dependent phonetic cue to the French nasal contrast, affects spoken-word recognition. Using eye tracking in two visual world paradigm experiments, the results show that fine-grained phonetic information is important for lexical recognition, and that lexical access is dependent on small variations in the signal. The results also show gradient interpretation of ambiguous vowel nasalization despite the phonemic distinction between phonological nasal vowels and coarticulatorily nasalized vowels in Canadian French. Gradience was found when words were ambiguous, and interpretation was more categorical when words were unambiguous. These results support the hypothesis of gradient interpretation of phonetic cues for ambiguously produced stimuli and the storage of coarticulatory information in phono-lexical representations for a language that has a phonological contrast for nasality (i.e., French).

**Keywords** Eye tracking · Coarticulation processing · Vowel nasalization · Gradience · Spoken-word recognition

During speech processing, listeners are presented with time-dependent and variable, fine-grained acoustic cues for phoneme and word recognition. These cues include within-category variability and coarticulation, which is the result of the overlap of adjacent sounds' articulatory movements (Fowler, 1980). These cues were traditionally considered redundant for formal theories

of phonetic, phonological, and lexical representations (Keating, 1988). In these frameworks (Archangeli, 1988; Keating, 1988; Steriade, 1995), within-category variability is idiosyncratic, and coarticulatory effects derive from rules (Cohn, 1990), and are not specified in word representations. Consequently, in perception, listeners are not expected to use these fine-grained phonetic details for word recognition. However, experimental evidence in favor of the importance of such fine-grained details has been repeatedly found in psycholinguistic studies (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Cross & Joanisse, 2018; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Desmeules-Trudel, Moore, & Zamuner, 2019; Gow, 2003; McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray, Tanenhaus, & Aslin, 2002; Paquette-Smith, Fecher, & Johnson, 2016; Zamuner, Moore, & Desmeules-Trudel, 2016).

For example, in English, speakers begin lowering their velum (an articulatory movement associated to nasalization) early during the production of vowels that are followed by a nasal consonant. This velum-lowering movement has an influence on the acoustic output of vowels that precede nasal consonants, yielding a partially or entirely nasalized vowel in production (Beddor, 2009). However, coarticulatory vowel nasalization has traditionally not been included as a part of phonological or lexical representations in English (Lahiri &

---

The original version of this article was revised: The Publisher regrets that, due to a typesetting mistake, it has now become necessary to make the following corrections: All phonetic transcriptions (between square brackets and slashes) should be corrected and displayed as follows, in the same font: [see line 22 to line 64 from the Correction article <https://doi.org/10.3758/s13414-019-01693-9>].

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.3758/s13414-019-01693-9>) contains supplementary material, which is available to authorized users.

---

✉ Félix Desmeules-Trudel  
fdesmeul@uwo.ca

<sup>1</sup> Department of Linguistics, University of Ottawa, Ottawa, ON, Canada

<sup>2</sup> Present address: Department of Psychology (Brain and Mind Institute), Western University, Western Interdisciplinary Research Building 5182, London, ON N6A 3K7, Canada

Marslen-Wilson, 1991), even though listeners use nasal coarticulation for word recognition (Beddor et al., 2013).

In addition to the actual use of fine-grained phonetic information, researchers have shown that coarticulation and within-category variability is not necessarily perceived categorically (McMurray et al., 2002), as would be expected based on seminal work on voice onset time (VOT) perception (Lieberman, Harris, Hoffman, & Griffith, 1957). For example, it has been shown that English adults gradiently pay attention to vowel nasalization (Beddor et al., 2013) when processing spoken words. Specifically, English listeners are faster at recognizing words that contain a nasal consonant (e.g., *scent*) when the preceding vowel is nasalized early than when the vowel is not (or nasalized later; Beddor et al., 2013). However, while there has been considerable research on listeners' sensitivity to coarticulatory cues (see references above), most of these studies have focused on English, yielding to an underrepresentation of data from a variety of language systems and the influence of a variety of variable phonetic cues, which would lead to a better understanding of the general word recognition capacity. Investigating a variety of languages is thus crucial to a better understanding of word processing.

Furthermore, the influence of variability in the realization of phonological contrasts on word recognition has not yet been thoroughly investigated, even less for vowel contrasts (for reports that include consonant contrasts, see Dahan et al., 2001; McMurray et al., 2008; McMurray et al., 2002). It is thus important to investigate participants from a variety of language backgrounds and on a variety of phonetic-detail types to gain a better understanding of the general process of word recognition, and of the influence of fine-grained phonetic cues on higher order processing. Mainly, this is because different languages make use of phonetic cues in different manners, and thus listeners can be expected to use those cues differently.

In this paper, we examine gradient versus categorical processing of vowel nasalization in Canadian French, a language that has contrastive nasal vowels (as in *pain* [pɛ̃] “bread”) and coarticulatorily nasalized vowels (Desmeules-Trudel & Brunelle, 2018; Léon, 1983)—that is, when oral vowels are followed by a nasal consonant (as in *peigne* realized [pej̃] or [pɛ̃j̃] “comb,” which can have a partially and optionally nasalized vowel). Nasal vowels can be followed by nasal consonants in French as well (e.g., *grand-mère* [gʁɑ̃mɛʁ] “grandmother,” as opposed to *grammaire* [gʁɑmɛʁ] “grammar”), which clearly demonstrates their contrastive status. Examining a single cue that varies in phonological status (contrastive vs. coarticulatory) within Canadian French enables us to formulate a new set of predictions on how listeners use fine-grained phonetic details during spoken-word recognition, as compared with coarticulatory (i.e., noncontrastive) processing in English. Note that another account of the influence of vowel nasalization on word recognition focused on Bengali, a language that has both phonologically and coarticulatorily nasalized vowels

(Lahiri & Marslen-Wilson, 1991), and found evidence for the early use of coarticulatory nasalization in this language. However, Lahiri and Marslen-Wilson (1991) did not directly manipulate duration of nasalization on the vowels, which is crucial to investigate the actual impact of fine-grained variations in nasalization duration on word recognition and if phonological contrasts are categorically or gradiently perceived.

For the current study, we manipulated duration of nasalization, similarly to Beddor et al. (2013). This manipulation enables us to formulate different predictions regarding perceptual patterns (i.e., gradient or categorical) of vowel nasalization in Canadian French compared with English, which is gradient. For example, categorical patterns of perception could be expected in French because listeners have to phonologically discriminate coarticulatorily nasalized vowels (e.g., *peigne* realized [pɛ̃j̃]) and contrastive nasal vowels (e.g., *pain* realized [pɛ̃]). On the other hand, it is possible that French listeners will still use variations in vowel nasalization in a gradient manner similarly to Beddor et al.'s (2013) English participants, since evidence for gradient perception has been found using phonological consonant contrasts (e.g., voicing in English; McMurray et al., 2002), in addition to evidence that listeners are able to pay attention to fine-grained phonetic details (see references above). Our main aim is thus to determine if variation in fine-grained phonetic details pertaining to a vocalic contrast is perceived in a gradient or categorical manner.

## Phonetic integration and gradient perception

As mentioned above, one challenge faced by the word-recognition system is the time-dependent aspect of spoken-word recognition, since phonetic cue duration can be the main indicator of a sound category. For example, voice onset time (VOT), a cue that varies in duration, is linked to word-initial consonant voicing in English (Lieberman et al., 1957). In English, long-lag VOT duration is associated with voiceless stop consonants, and short-lag VOT duration is associated with voiced consonants. Consequently, in perception, listeners must consider these dynamics before making a decision on the phonological identity of a stop consonant.

Previous research has found support for gradient sensitivity to VOT variations during word recognition (McMurray et al., 2008; McMurray et al., 2002), despite the categorical character of VOT in production (long-lag vs. short-lag). For example, McMurray et al. (2002) showed that native English listeners were sensitive to small variations in VOT values on a continuum between voiced (e.g., /b/, VOT = 0 ms) and voiceless (e.g., /p/, VOT = 40 ms) consonants. In their eye-tracking experiment, listeners were instructed to listen to auditory stimuli composed of minimal pairs (e.g., *bear-pear*) that had been modified to form nine-step continua between 0 ms and 40 ms in VOT in 5-ms increments. Participants were asked to click

on an image corresponding to the word they heard while their eye movements were recorded. McMurray et al. (2002) found that stimuli at the end points of the continuum (0 ms and 40 ms) led to high proportions of fixations to the *correct* target (high proportions of fixations to the *beach* for 0 ms VOT, and high proportions of fixations to the *peach* for 40 ms VOT), but that listeners displayed intermediate proportions of fixations for VOT values nearer to the category boundary. This is an example of gradient sensitivity to a contrast (see also McMurray et al., 2008).

## Vowel nasalization in (Canadian) French

As mentioned above, Canadian French vowel nasality is a phonologically contrastive property, meaning that words such as *pain* (/pɛ̃/ “bread”) and *paix* (/pɛ/ “peace”) constitute minimally different lexical items that are phonologically distinguished based on the nasality of the vowel. There are four contrastive nasal vowels in Canadian French: /ɛ̃, ɑ̃, ɔ̃, œ̃/ (Côté, 2012; Martin, 2002; Martin, Beaudoin-Bégin, Goulet, & Roy, 2001), which can be variably realized depending on a number of factors (Carignan, 2013; Delvaux, 2006; Desmeules-Trudel & Brunelle, 2018; Léon, 1983; Martin, Beaudoin-Bégin, Goulet, & Roy 2001). For example, Delvaux (2006) and Desmeules-Trudel and Brunelle (2018) found that nasalization of phonological nasal vowels can start at vowel onset or can be delayed (up to 50% of its duration) in Canadian French. In other words, phonological nasal vowels can be nasalized for their entire duration, or they can be only partly nasalized; this varies based on syllable structure, prosodic context, and individual speakers, among other factors. To our knowledge, however, no study has directly investigated how common are vowels that are nasalized for the entirety of their duration, but research suggested that they are variable overall and *can* be fully nasalized (Desmeules-Trudel & Brunelle, 2018; Léon, 1983). Furthermore, oral vowels followed by a nasal consonant can be (optionally) coarticulatorily nasalized in Canadian French (Desmeules-Trudel & Brunelle, 2018; Léon, 1983).

In addition, Carignan (2013) and Desmeules-Trudel and Brunelle (2018) found that short excrescent nasal consonants, also referred to as nasal appendices, are often present in the realization of phonological nasal vowels. This can lead to the ambiguous realization of a phonological nasal or coarticulatory nasalized vowel, when controlling for other differences in their realization (e.g., oral resonances; Carignan, 2013, 2014). For example, if the vowel in *pain* [pɛ̃] “bread” is produced with an excrescent nasal appendix (i.e., [pɛ̃ɲ]), it could be confused with the word *peigne* ([peɲ] “comb”), especially given that the vowel in *peigne* can be slightly nasalized as well. Consequently, it is likely that listeners consider the nasalization timing information (i.e., the moment at which

nasalization starts within the vowel) when interpreting nasal vowels in Canadian French.

In the experiments below, we varied the duration of nasalization on vowels (i.e., timing information) in real French words, analogously to McMurray et al.’s (2002) VOT continuum, and we varied the presence or absence of nasal appendices, to investigate how Canadian French listeners interpret words that contain vowels that are early-nasalized (i.e., vowels that are nasalized from onset or shortly after onset, and onwards) and late-nasalized (i.e., vowels that are nasalized late in their duration, until vowel offset). By cross-splicing portions of phonological nasal vowels (i.e., a part or the entirety of the vowel in *pain* [pɛ̃] “bread”) onto vowels that can be coarticulatorily nasalized (i.e., *peigne* [peɲ] “comb”), we were able to control for small variations in the duration of nasalization. To our knowledge, the possibility of gradient versus categorical perception has never been directly assessed in vowels during word recognition, even though it has been shown that consonants are generally perceived more categorically than vowels in syllables (Fry, Abramson, Eimas, & Liberman, 1962).

## Study predictions

Two eye-tracking experiments were designed to test how variations in fine-grained phonetic information on vowels is interpreted during spoken-word recognition (i.e., categorically vs. gradiently). Listeners had to identify spoken words in a Visual World Paradigm (Huetting, Rommers, & Meyer, 2011). Similarly to Beddor et al. (2013) and given the realization of the nasality contrast in Canadian French (Desmeules-Trudel & Brunelle, 2018), it was expected that variation in the duration of nasalization would have an influence on the recognition of spoken words containing oral vowels followed by nasal consonants (CVN) and phonological nasal vowels (C $\bar{V}$ ). For example, vowels that are nasalized for 50% or more of their duration are expected to be recognized mostly as phonological nasal vowels, especially when not followed by a nasal appendix. On the other hand, vowels followed by a nasal consonant (VN) in a closed syllable can be nasalized (i.e., coarticulated) for approximately 20% to 25% of their duration towards vowel offset (Desmeules-Trudel & Brunelle, 2018). Therefore, vowels that are nasalized for 20% or less of their duration are expected to be mostly recognized as oral vowels (followed by nasal consonants).

The main goal of this paper was to verify if phonetic information related to vowel nasalization is perceived categorically or gradiently by listeners of Canadian French. On the one hand, it is expected that listeners will display categorical patterns of nasalization perception since this property is phonological in Canadian French. Listeners might not pay attention to within-category variability, as was shown in early experiments on categorical perception with VOT (Liberman et al., 1957), and in

studies on perceptual compensation (Beddor & Krakow, 1999; Fowler, 2006). For instance, Beddor and Krakow (1999) showed that in nasal contexts, English and Thai native listeners (partially) compensated for vowel nasalization and attributed the coarticulatory vocalic phonetic cues to a following nasal consonant. In the current paper, if listeners perceive a nasal consonant (see stimuli below) and compensate for vowel nasalization, categorical patterns are expected to emerge.

On the other hand, analogously to Beddor et al.'s (2013) study, Canadian French listeners may look earlier at words that contained a contrastive nasal vowel when the stimulus is nasalized early in its duration than when it is nasalized late, and gradiently fixate more to the CVN word as vowel nasalization starts later on the vowel. This pattern can be expected based on other gradient perception studies (McMurray et al., 2002)—listeners dynamically integrate phonetic cues as they unfold in the speech signal and “update” their word choice online (McClelland & Elman, 1986). Earlier nasalization on the vowel could enable listeners to anticipate the identity of the (upcoming) nasal vowel, yielding gradient/continuous patterns of recognition. We used eye tracking because the investigated questions involve the *timing* of nasalization onset, and eye movements have been shown to be sensitive to such timing information (Beddor et al., 2013; McMurray et al., 2008; McMurray et al., 2002; Salverda, Kleinschmidt, & Tanenhaus, 2014).

## Experiment 1

The stimuli in Experiment 1 contained a nasalized vowel and a following nasal exrescent consonant, that is, a short consonant-like nasal sound, in coda position of the word. The choice was made to include a short consonant after the vowel to reflect the phonetic realization of phonological nasal vowels ( $\tilde{V}$ ; Desmeules-Trudel & Brunelle, 2018) in addition to variations in nasalization duration. However, phonologically, nasal vowels followed by a full nasal consonant in coda position are banned in French. Consequently, if listeners interpret the exrescent consonant as a full nasal consonant, stimuli can be considered to contain conflicting cues, that is, cues that pertain to both a phonological nasal vowel and an oral vowel followed by a consonant.

## Method

### Participants

Twenty-three native speakers of Canadian French (17 female, six male), between 18 and 36 years of age ( $\bar{X}$  = 23.3 years,  $SD$  = 5.3), were paid or received partial course credit for their participation in the experiment. Thirteen participants were from Ontario, nine from Québec, and one from New Brunswick.

All listeners reported having normal hearing, normal or corrected-to-normal vision, and did not report any type or history of language, hearing, or speech impairment. All participants completed a background language questionnaire, and self-reported knowing English as a second language at a moderate to high level of proficiency, as well as other languages. Most speakers were late bilinguals in English ( $N$  = 16), and some were early bilinguals in English ( $N$  = 7). The early versus late bilingual status was determined based on a criterion of participants having on average 30% or more exposure to English as a second language in one speaking context before 5 years of age (e.g., in the family, in daycare, with peers). Since all participants were bilingual, we do not expect differences to emerge regarding the listeners' language backgrounds, especially given that nasalization is only coarticulatory in English (or other languages known to the listeners). No speaker had been exposed to another language containing contrastive nasal vowels in its phonological inventory.

### Stimuli and experimental conditions

Stimuli were nine triads of monosyllabic, picturable French words containing either a contrastive nasal vowel ( $C\tilde{V}$ ), an oral vowel in a nasalization context (i.e., followed by a nasal consonant; CVN), or an oral vowel followed by an oral consonant (CVC; see Table 1). Note that the place of articulation of the oral consonant in CVC words did not always match the place of articulation of the nasal consonant in CVN words. Three triads contained midfront [ $\epsilon$ – $\tilde{\epsilon}$ ], three contained low [ $a$ – $\tilde{a}$ ], and three contained midback [ $\sigma$ – $\tilde{\sigma}$ ]. Note that the vowels [ $\alpha$ – $\tilde{\alpha}$ ] were not included as the functional load of the nasal vowel is very low in the French lexicon (Martin et al., 2001). Furthermore, frequency of the nasal and coarticulated words could not be controlled due to constraints on word choice, and was therefore not analyzed in the current paper. The word list was recorded by five native speakers of Canadian French (two female, three male, between the ages of 23 and 27 years) in order to avoid listener habituation to one speaker during the test phase. The speakers were from different regions in Canada to ensure representation of a variety of Canadian French dialects. The words were embedded into different carrier sentences that matched the target word meaning in order to avoid focus effects and were placed at the end of an intonational phrase in the carrier sentences. Speakers were instructed to read the sentences twice in a natural, but careful way.

Each word was hand-segmented in Praat (Boersma & Weenink, 2015) and normalized for amplitude at 70 dB. The  $C\tilde{V}$  and CVN tokens were compared for each speaker, and the most similar tokens were selected. Observation of the acoustic spectrum (absence of nasal antiformants and nasal peaks in the 800–1500 Hz region, though these cues can be variable across speakers for the production of nasalization) and auditory confirmation enabled us to ensure that the first 80% of the CVN

**Table 1.** List of minimal pairs (and fully oral words) that were used as auditory and visual referents

Contrastive $\tilde{V}$ word	IPA	Coarticulated VN word	IPA	Fully oral vowel word	IPA
<i>bain</i> “bathtub”	[bɛ̃]	<i>beigne</i> “donut”	[bɛɲ]	<i>bec</i> “beak”	[bɛk]
<i>pain</i> “bread”	[pɛ̃]	<i>peigne</i> “comb”	[pɛɲ]	<i>pêche</i> “peach”	[pɛʃ]
<i>vin</i> “wine”	[vɛ̃]	<i>veine</i> “vein”	[vɛɲ]	<i>verre</i> “glass”	[vɛʁ]
<i>camp</i> “campsite”	[kɑ̃]	<i>canne</i> “cane”	[kan]	<i>cave</i> “basement”	[kav]
<i>dent</i> “tooth”	[dɑ̃]	<i>dame</i> “lady”	[dam]	<i>datte</i> “date (fruit)”	[dat]
<i>gant</i> “glove”	[gɑ̃]	<i>gamme</i> “musical scale”	[gam]	<i>gare</i> “bus/train station”	[gɑʁ]
<i>pont</i> “bridge”	[pɔ̃]	<i>pomme</i> “apple”	[pɔm]	<i>porc</i> “pork”	[pɔʁ]
<i>rond</i> “circle”	[ʁɔ̃]	<i>rhum</i> “rum”	[ʁɔm]	<i>robe</i> “dress”	[ʁɔb]
<i>thon</i> “tuna”	[tɔ̃]	<i>tonne</i> “ton”	[tɔn]	<i>torse</i> “torso”	[tɔʁs]

token vowels’ duration were not nasalized. This was important since nasalization of this portion of the CVN vowels could compromise the 20N %NASDUR condition (see below). Details about the duration of oral and nasalized portions of the stimuli vowels are found in Table 3 of the Appendix.

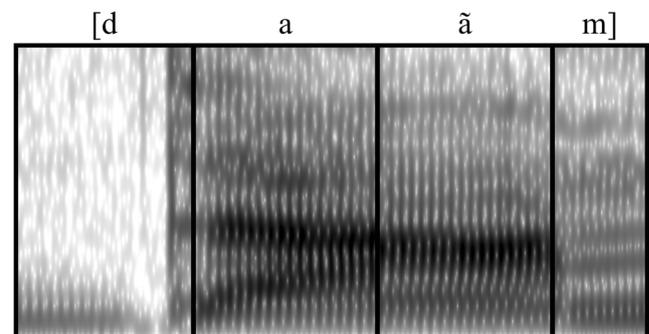
Each speaker contributed one or two word pairs. For each experimental stimulus, the consonant frame of a CVN token was kept by removing a part or the entirety of the oral vowel. This yielded five experimental conditions (i.e., proportion of duration of the vowel that is nasalized—%NASDUR): fully oral vowels (0N), partially nasalized vowels (20N, 50N, 80N), and fully nasal vowels (100N). The %NASDUR values are illustrated in Table 2. For 0N, the vowel from the CVN matrix token was removed entirely and replaced by an oral vowel of a CVC token. However, the 0N %NASDUR condition was not included in the current analyses, since place of articulation of the consonant following the vowel in the CVC word was not consistently the same as the place of articulation of the CVN word. This is further discussed in the Discussion. For 20N, 50N, and 80N, the duration of the oral vowel was calculated, and 20%, 50% or 80% of the vowel duration was cut from the original vowel (CVN). The portion of the vowel that was removed from the original token was replaced by a part of the nasal vowel of the matching  $\tilde{C}\tilde{V}$  token, considering zero passages in order to avoid unwanted acoustic artifacts from the splicing such as clicks or noise in the signal (see Fig. 1). Amplitude peak reduplication was performed when necessary to adjust the duration of the vowel. For 100N, the oral vowel of the original token was entirely replaced by the full vowel of the matching  $\tilde{C}\tilde{V}$  token. Note that the 50N, 80N,

and 100N contain potentially conflicting cues, as they are nasalized for a significant part of their duration (similar to phonological nasal vowels) and also contain an excrescent nasal coda. Final adjustments on fundamental frequency and amplitude were made, and the stimuli quality was verified by the first author (a native speaker of Canadian French). Participants from both experiments reported that the stimuli sounded natural, but sometimes ambiguous regarding their meaning. Fifty ms of the final nasal consonant following the vowel in the CVN word were kept in the stimuli in order to mimic the natural realization of nasal vowels’ appendices in Canadian French connected speech (Desmeules-Trudel, 2015; Desmeules-Trudel & Brunelle, 2018). A total of 45 experimental stimuli was used in the experiment (i.e., five stimuli for each word pair), and data from 36 items were analyzed for the current paper.

Visual stimuli corresponded to the words (Table 1) and were taken from the International Picture Naming Project (IPNP) database (Székely et al., 2004), which contains black-and-white pictures of a variety of (English) nouns. Pictures of words that could not be found in the database (12 critical  $\tilde{C}\tilde{V}$  or CVN items, seven noncritical CVC items,  $N = 19$ ) were hand drawn by a professional visual artist in the same style as the IPNP images, scanned and saved as JPEG files. All visual stimuli were 5.6 cm × 5.6 cm (220 × 220 pixels),

**Table 2.** Summary of the analyzed experimental conditions (%NASDUR)—Experiment 1

Vowel duration				
Condition	20N	Oral (80%)	Nasal (20%)	50 ms nasal cons.
	50N	Oral (50%)	Nasal (50%)	
	80N	Oral (20%)	Nasal (80%)	
	100N	Nasal (100%)		

**Fig. 1** Spectrogram of the dame/dent [dam]/[dã] “lady/tooth” stimulus in the 50N condition, showing the initial consonant ([d]), 50% of the vowel duration from the oral vowel in *dame* [dam], 50% of the vowel duration cross-spliced from the nasal vowel in *dent* [dã], and 50 ms from the final [m] consonant in *dame*

and images corresponding to the triads with one member containing a phonological nasal vowel (C $\tilde{V}$ ), a coarticulated nasalized vowel (CVN), and an oral vowel (CVC), as well as a distractor image were arranged together on a display. Words corresponding to the distractor image had the same initial consonant as the other stimuli of the triads. Images were embedded within approximately 6 cm  $\times$  6 cm (250  $\times$  250 pixels) interest areas for data collection. For example, C $\tilde{V}$  *pain* (“bread” [p $\tilde{e}$ ]), CVN *peigne* (“comb” [pej $\tilde{n}$ ]), CVC *pêche* (“peach” [peʃ]) and filler *pluie* (“rain” [pluj $\tilde{i}$ ]; see Fig. 2) were simultaneously presented on the screen.

## Procedure

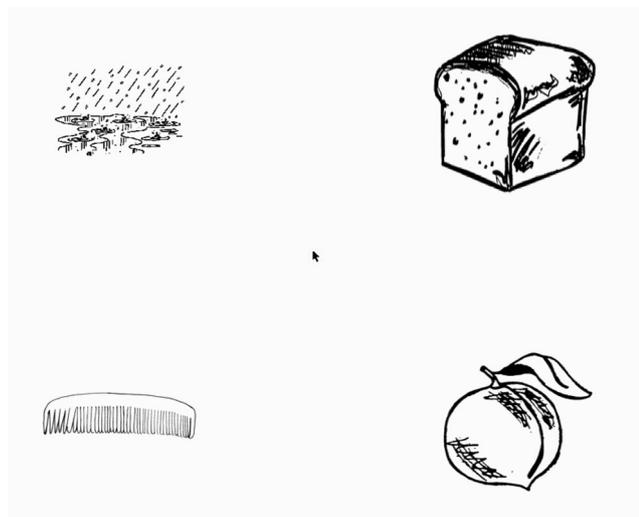
The experiment was programmed and presented in Experiment Builder (SR Research; Version 1.10.63) and presented with an EyeLink 1000 (SR Research), using a chin rest, monocular recording, and sampling at 500 Hz. The experiment started with a 5-point calibration then validation, keeping the maximum and average errors below 1° of visual angle for all participants. A familiarization phase followed, during which participants saw an image on the screen with the corresponding unambiguous auditory word. Unspliced tokens of the critical C $\tilde{V}$  ( $N = 9$ ) and CVN words ( $N = 9$ ), CVC words ( $N = 9$ ), and monosyllabic filler words ( $N = 43$ ) were presented (total items in familiarization phase,  $N = 70$ ). The odd number of filler words is due to the fact that some fillers were used simultaneously with the experimental triads (i.e., experimental trials), while other fillers were used on displays containing four filler images (i.e., filler trials). After familiarization, participants were tested on the word-picture association task. Drift correction was performed before each trial, which

was experimenter controlled. On each trial, participants saw four images corresponding to the C $\tilde{V}$ , CVN, CVC, and filler words for 500 ms, followed by the presentation of an audio token corresponding to an experimental (see Table 2) or filler stimulus. Participants were asked to click on the image that corresponded to the word they heard. This ended the trial and changed the display to the next drift correction. No feedback was provided on the accuracy of participants’ responses, as some stimuli were inherently ambiguous. Image position was pre-randomized across trials. There were 90 critical trials (i.e., each critical stimulus was presented twice), of which 72 critical trials were analyzed in the current paper since the 0N %NASDUR condition was removed from the analyses (see Discussion), and 70 filler trials per participant. The experiment lasted approximately 25 to 30 minutes. Participants were randomly assigned to one of four experimental lists to counterbalance block order presentation.

The collected data included the images that participants clicked on and the fixations to the various images on the display. Due to an error in programming, responses corresponding to the coarticulated image (CVN; e.g., *peigne*), the oral word (CVC; e.g., *pêche*), and the unrelated image (e.g., *pluie*) were all collapsed into the coarticulated (CVN) response category. To overcome this problem, we calculated an eye-tracking measure to infer which image participants chose in these cases. We based this on the image that was fixated to the most in the last 1,000 ms of each trial. Therefore, the CHOSEN IMAGE measure that will be presented below consists of the image that was fixated the longest in the last 1,000 ms of each trial.<sup>1</sup> Proportions of fixations to each image (i.e., fixations within the interest areas around the images) were calculated in 50 ms time bins using a Python script provided by SR Research. Analysis will focus on the proportions of fixations to the coarticulated-vowel image (CVN) when listeners chose the CVN item (969 trials), and on the proportions of fixations to the contrastive-nasal-vowel image (C $\tilde{V}$ ) when listeners chose the C $\tilde{V}$  word (526 trials) to uncover categorical or gradient patterns of perception depending on proportion of nasalization conditions.

## Analysis and variables

Both the chosen images and proportions of fixations were modeled using (separate) generalized additive mixed models (GAMMs; Wood, 2017). This statistical technique allows one to analyze factorial and/or gradient predictors on potentially



**Fig. 2** Example of an experimental display. Four images were displayed on the screen. On stimuli was from the C $\tilde{V}$  condition (*pain* “bread” [p $\tilde{e}$ ]), one from the CVN condition (*peigne* “comb” [pej $\tilde{n}$ ]), one from the CVC condition (*pêche* “peach” [peʃ]), and lastly there was one filler image (e.g., *pluie* “rain” [pluj $\tilde{i}$ ])

<sup>1</sup> The programming error was noted and corrected halfway through data collection in Experiment 2. There were 12 participants whose responses were correctly compiled. The CHOSEN IMAGE measure matched these 12 participants responses for more than 98% of cases. This supports our use of the measure CHOSEN IMAGE to infer the lexical choice that was made by the subset of participants for whom we do not have responses broken down for CVN, CVC, and filler responses.

nonlinear data, which is often the case for eye movements. Random effects structure (intercepts, slopes, and nonlinear smooth terms) can be added to the model. Also, autocorrelation of time series data is considered (Baayen, van Rij, de Cat, & Wood, 2018), which is necessary for time-dependent data such as eye tracking because one data point is correlated in time to the preceding point. Finally, GAMMs can handle unbalanced data (i.e., missing data points), which is common in eye tracking as participants are free to fixate outside of the predetermined interest areas during a trial. GAMMs have successfully been used in past research to analyze visual world eye movements (Porretta, Kyröläinen, van Rij, & Järviö, 2017; Porretta, Tucker, & Järviö, 2016; van Rij, Hollebrandse, & Hendriks, 2016). Also note that GAMMs require visual inspection of model estimates to interpret (non-)significance of factors, especially since “nonlinear trends are difficult to capture with a single parameter” (Porretta et al., 2017, p. 270). The *mgcv* (Version 1.8-16; Wood, 2017), *itsadug* (Version 2.2; van Rij, Wieling, Baayen, & van Rijn, 2016), and *ggplot2* (Version 2.2.1; Wickham, 2009) packages were used for analysis and visualization in R (R Core Team, 2017; Version 3.4.2).

For the GAMM analysis of chosen images, models included %NASDUR as the main factor of interest to assess if participants’ responses were gradient or categorical, outputting estimates of the probability (in log odds) of giving a C $\check{V}$  response for each condition. Random intercepts by participant and item were also included. The *p* values of parametric coefficients represent a significant difference of one level to the baseline. By using the 50N %NASDUR condition as a baseline, we were able to detect differences between this latter %NASDUR value and the other ones—significant differences between 50N and multiple other levels of the same factor (i.e., significant differences between 20N–50N, 50N–80N and 50N–100N) suggest gradient use of vowel nasalization variations, while nonsignificant differences between 50N–80N and 50N–100N would suggest more categorical patterns of perception. For example, if the probability of giving a nasal (C $\check{V}$ ) response is significantly lower in the 20N %NASDUR condition than in the 50N %NASDUR condition, and significantly higher in the 80N than the 50N %NASDUR condition, we can conclude that listeners gradually interpreted nasalization timing variations on vowels. On the other hand, if one or more levels are not significantly different from the 50N %NASDUR baseline condition, this would suggest more categorical patterns of interpretation overall. 50N %NASDUR was chosen as a baseline since the boundary between contrastive nasal vowels and coarticulatorily nasalized vowels in production is between 20% and 50% of the vowel duration (see airflow evidence in Desmeules-Trudel & Brunelle, 2018). Additional factors of interest (VOWEL QUALITY, TRIAL NUMBER), which do not directly pertain to processing of fine-grained phonetic detail, were also assessed by adding each factor to the baseline model (which only contained

%NASDUR and random intercepts), and performing chi-square tests between the baseline and more complex model using the *compareML()* function from the *itsadug* package. This procedure provided chi-square scores and *p* values, which enabled us to assess the contribution of each factor to model fit. The results of the second part of the analysis are presented in the [online Supplementary Materials](#) associated to the current paper.

For the analysis proportions of fixations to the C $\check{V}$  image (when participants chose the C $\check{V}$  image) and to the CVN image (when participants chose the CVN image), empirical logits (Barr, 2008) were used as the input variable to the model. We decided to analyze the fixations to the C $\check{V}$  and CVN images separately since our stimuli were inherently ambiguous and a constant target could not be chosen across the entire experiment. Thus, we separated the data into two separate data frames—when participants chose the C $\check{V}$  image and when participants chose the CVN image. This had the effect to provide a proper “target” for analysis, and therefore enabled us to carefully assess the gradient versus categorical patterns of fixations to the chosen image.

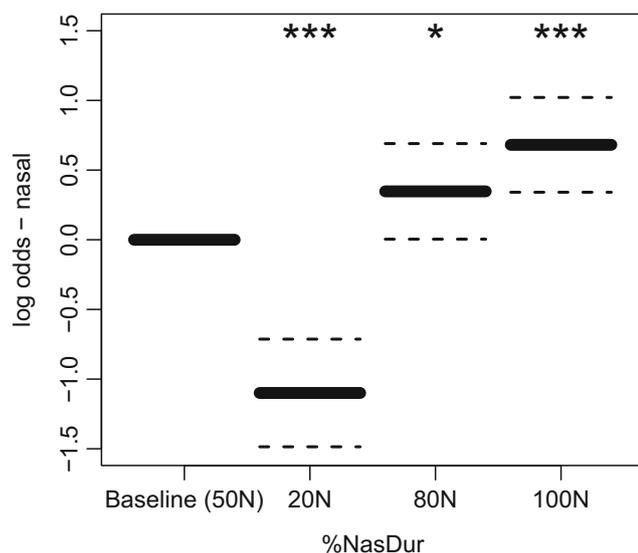
Similar to chosen images, the fixation analysis was twofold. First, to assess the gradient versus categorical perception of vowel nasalization timing variability, a “simple” model was fitted to the eye-tracking data, which included parametric %NASDUR CONDITION, TIME window of analysis (1,000 ms, starting 200 ms after vowel onset to account for delay in eye-movement planning; Fischer, 1992), an interaction between %NASDUR and TIME window, and random intercepts for each individual TRIAL. Also note that the time window of analysis was selected semi-arbitrarily based on the observation of raw data across experiments—participants seem to have made their choice on the word identity at around 1,200 ms based on the observation of higher proportions of fixations at that point in time. Random effects structure included random intercepts for each TRIAL. The *p* values of parametric coefficients represent a significant difference of one level to the baseline, and the *p* values of smooth terms indicate if it is different from zero. In the body of the text, we report the model estimates as well as a graphical representation of the parametric estimates of this model.

Second, to assess the impact of additional factors on the participants eye movements, each additional factor of interest (VOWEL QUALITY, TRIAL NUMBER, and interactions through TIME) was tested against the baseline model one at a time using the *compareML()* function from the *itsadug* package, similarly to the response analysis. Note that for testing the interactions (e.g., TIME  $\times$  TRIAL), the single factor (e.g., TRIAL) was also included in the model to offer a baseline for the interaction term. In the [Supplementary Materials](#), we report chi-square scores, difference in chi-square scores between the baseline and tested models, as well as *p* values that indicate significance and Akaike information criterion difference values, an indication of quality of model fit.

## Results

### Chosen images

Figure 3 depicts the model predictions, outputting the probability (in log odds) of choosing the  $C\tilde{V}$  image in Experiment 1, with 50N %NASDUR (134/380  $C\tilde{V}$  image choices in this %NASDUR condition, 35.3%) as a baseline against which the other levels of the factor were tested. The actual output of the GAMM model, as generated by the *summary()* function in the *mgcv* (Wood, 2017) package, is presented in Table 4 of the Appendix. In the figure, we observe that 20N %NASDUR (61/371  $C\tilde{V}$  image choices in this %NASDUR condition, 16.4%) yielded significantly lower probability of choosing the  $C\tilde{V}$  image than 50N. On the other hand, both the 80N (152/368  $C\tilde{V}$  image choices in this %NASDUR condition, 41.3%) and 100N %NASDUR (180/376  $C\tilde{V}$  image choices in this %NASDUR condition, 47.9%) values yielded significantly higher probability of choosing the  $C\tilde{V}$  image than in the 50N %NASDUR condition. This is consistent with the gradient perception hypothesis, which predicts that listeners will display gradiently increasing response patterns as the phonetic properties of the segment get closer to the canonical realization of the phoneme. For instance, contrastive nasal vowels in Canadian French are expected to be nasalized for a longer period (Desmeules-Trudel & Brunelle, 2018); therefore, listeners are expected to give more  $C\tilde{V}$  responses if the vowel is actually nasalized for a longer period. This is also consistent with Beddor et al.'s (2013) results in English, which also support gradient perception of (coarticulatory) nasalization.



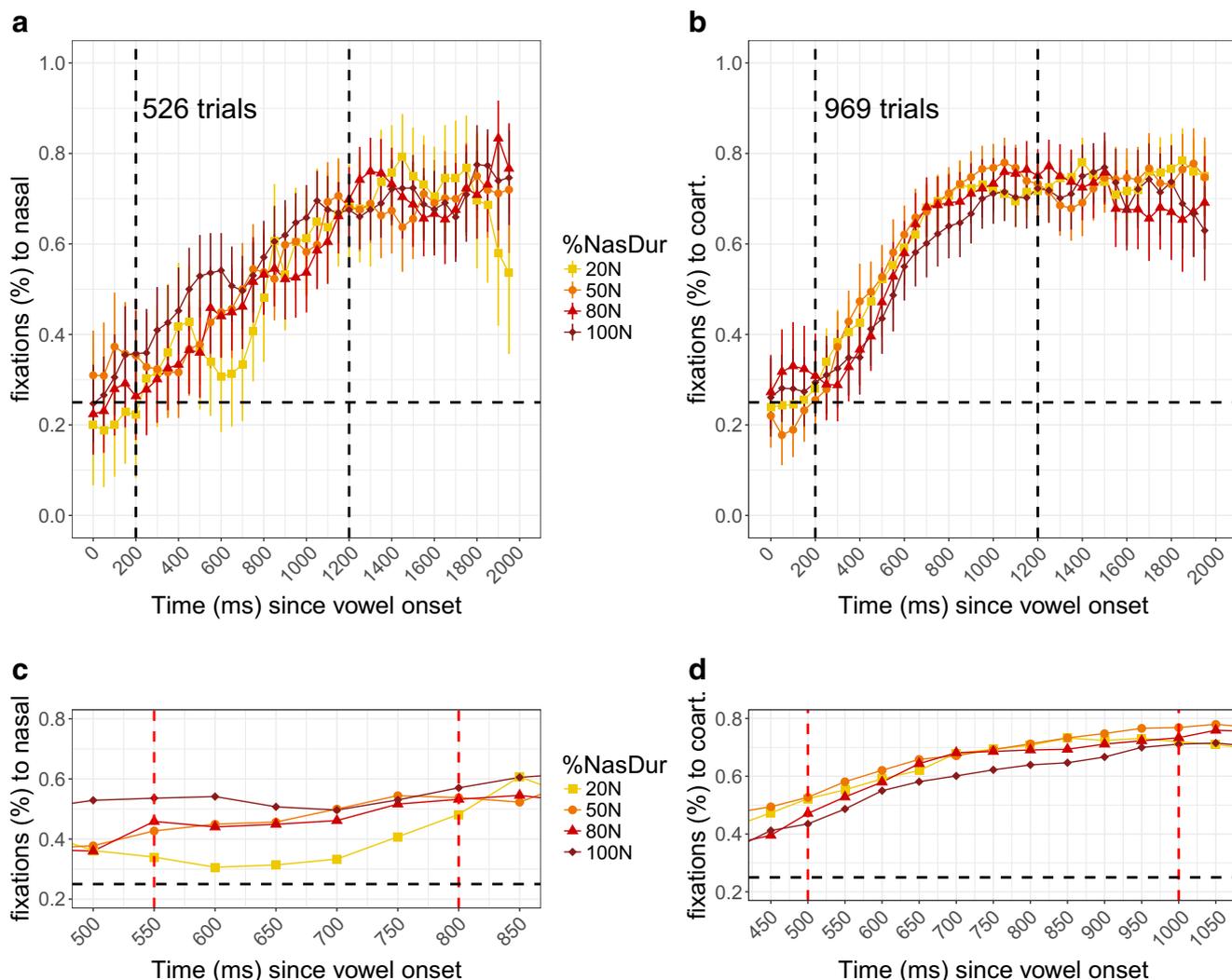
**Fig. 3** Experiment 1 GAMM predictions of choosing the  $C\tilde{V}$  image, significance indicators against the 50N %NASDUR baseline (\* $p < .05$ , \*\*\* $p < .001$ ). Note that the intercept is centered at zero

### Eye movements

This subsection presents the eye-tracking results of Experiment 1, with the proportions of fixations to the nasal ( $C\tilde{V}$ ) image when listeners chose the nasal image, and proportions of fixations to the coarticulated (CVN) image when they chose the coarticulated image between 200 (i.e., delay for programming an eye movement; Fischer, 1992) and 1,200 ms after vowel onset (between the dashed lines in Fig. 4a–b).

Observing the proportions of fixations in Fig. 4a–b, both fixations to the  $C\tilde{V}$  image for  $C\tilde{V}$  image choices and fixations to CVN for CVN image choices are higher than chance (dotted line at 25%; four images on the screen) shortly after the eye-movement programming delay. This suggests that listeners chose which image the auditory stimulus corresponded to shortly after the stimulus vowel onset. In Fig. 4a, although the error bars for the various conditions overlap across the whole window of analysis (and after thereof), the 20N %NASDUR condition yields slightly lower proportions of fixations than the other conditions of Fig. 4a between 550 ms and 800 ms (see Fig. 4c), similarly to the 100N in Fig. 4b between 500 ms and 1,000 ms (see Fig. 4d). This suggests that when the vowel was nasalized for a short period of time (20N %NASDUR), participants fixated slightly less to the nasal image during a short period of time even though they chose the  $C\tilde{V}$  image as their final choice, similar to the 100N %NASDUR–CVN image combination. On the other hand, on Fig. 4a, we see that the 100N %NASDUR value yielded to higher proportions of fixations to the  $C\tilde{V}$  image early on (200 ms to 650 ms) when participants chose the  $C\tilde{V}$  image. This is expected, since a vowel that is nasalized early corresponds to this image choice (except for when listeners interpreted the appendix as a realized nasal consonant).

Results of the GAMM analysis of fixations to the  $C\tilde{V}$  image ( $C\tilde{V}$  image choices) and to the CVN image (CVN image choices) are presented in Fig. 5a and b, respectively (parametric %NASDUR factor), and Table 5 and Table 6, respectively, in the Appendix. Let us recall that like the image choice analysis, the first part of the analysis focused on the parametric %NASDUR factor to assess gradient or categorical patterns of fixations, presented here in Fig. 5. The  $p$  values of parametric factors lower than .05 suggest a significant to the (50N %NASDUR) baseline. The  $p$  values of smooth factors lower than .05 in the tables suggest that the smooth term is significantly different from zero, suggesting statistical difference between two or more time bins through time (e.g., if proportions of fixations are significantly lower at 200 ms than at 1,000 ms, the term will be significant). The TIME factor only includes the window of analysis, taking into account the eye-movement planning delay (0 ms represents the beginning of the analysis window, 200 ms after vowel onset), and the auto-correlation AR1 value was set to 0.745 based on the data. This value corresponds to the average correlation between one data point and the preceding one across the data on the TIME dimension.



**Fig. 4** Experiment 1 proportions of fixations to the CṼ image when participants chose the CṼ image (a) and to the CVN image when they chose the CVN image (b). Panels c and d illustrate the lower proportions of fixations to the CṼ image in 20N %NasDur between 550 ms and 800 ms,

and to the CVN image in the 100N %NasDur between 500 ms and 1000 ms, respectively. The interval between dashed vertical lines represent the time window of the statistical analysis below, and the horizontal dashed line represents chance at 25% (four images on the screen). (Color figure online)

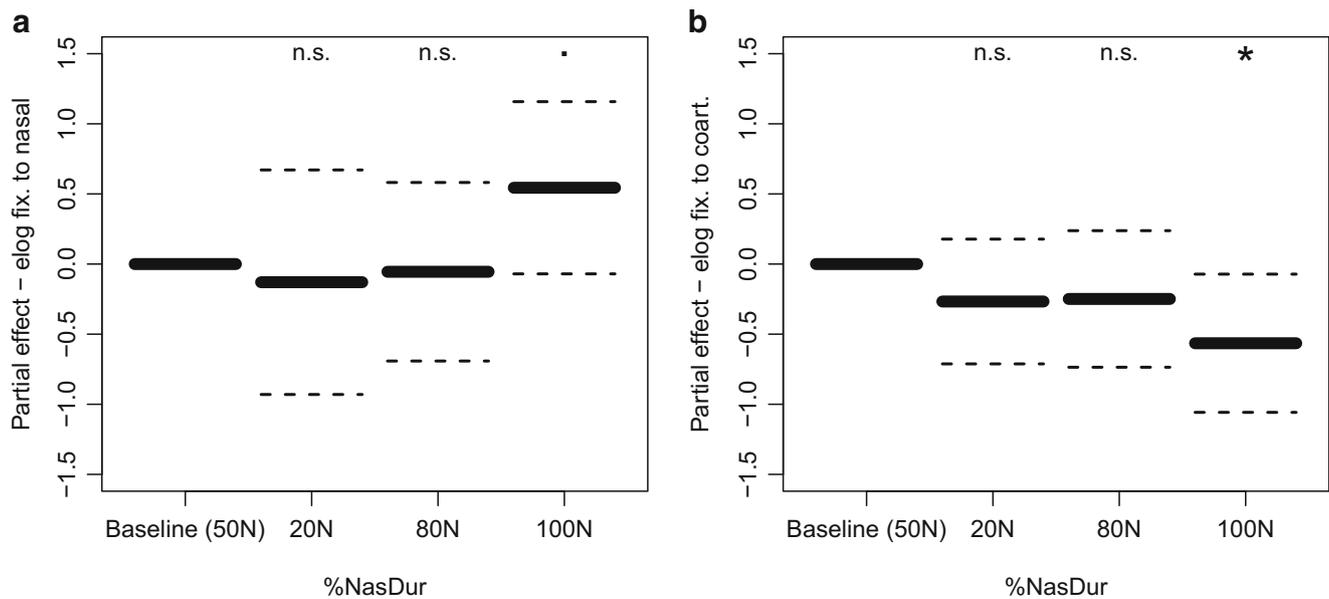
Figure 5a shows that no level of the %NasDur factor significantly differed from the 50N baseline overall (i.e., without considering the TIME factor), except for a trend towards significance for the 50N–100N %NasDur ( $p = .0764$ ), for the fixations to the nasal (CṼ) image when participants chose the nasal (CṼ) image. In this case, fixations to the nasal (CṼ) image are slightly higher in the 100N than in the 50N %NasDur. This result is expected, but does not support gradience in fixations to the nasal image for nasal image choices. However, this result does not convincingly support categorical use of variations in %NasDur for the recognition of nasal (CṼ) words either. Rather, it seems that when listeners chose the nasal (CṼ) image, they did not pay close attention to fine-grained variations in %NasDur.

On the other hand, in Fig. 5b, the difference between overall fixations to the coarticulated (CVN) image (CVN image

choices) in the (baseline) 50N %NasDur is significantly different from overall fixations in the 100N %NasDur. Overall fixations in the 100N %NasDur are lower than in the 50N %NasDur, which is expected based on the realization of vowels in CVN words in production (Desmeules-Trudel & Brunelle, 2018). The other %NasDur conditions (20N and 80N) did not yield significant differences with the baseline 50N condition. In the case of fixations to CVN words for CVN image choices, listeners categorically interpreted the words, as expected based on the predictions of categorical perception.

## Discussion

Our results indicate that variations in the duration of nasalization (%NasDur) have a significant influence on the recognition of phonological nasal (CṼ) and coarticulated nasalized



**Fig. 5** Experiment 1 GAMM predictions of fixations to the CV image (CV image choices) in Panel a, fixations to the CVN image (CVN image choices) in Panel b, and significance codes against the 50N %NASDUR baseline (n.s. = nonsignificant,  $p < .10$ ,  $*p < .05$ ). Note that the intercept is centered at zero

(CVN) vowels by listeners of Canadian French. This is consistent with Beddor et al.'s (2013) finding that (English) listeners are sensitive to the timing of nasalization onset on a vowel. For example, they found that listeners are slower at recognizing words that contain a nasal consonant (e.g., *scant*) when vowels are nasalized late in their duration than when they are nasalized early. This suggests that they are highly sensitive to nasalization timing information on the vowel. Here, participants' probability of choosing the nasal (CV) image significantly varied across %NASDUR values, which was expected based on acoustic analyses of the production of these vowels (Desmeules-Trudel & Brunelle, 2018). Importantly, the significant differences between the (baseline) vowels that were nasalized for 50% of their durations and all the other levels of the %NASDUR factor suggest that listeners gradually used variations in nasalization duration to identify the spoken words—both the observation of the patterns (i.e., constant increase in probability of choosing the CV image as %NASDUR values increase) and results of the statistical analysis support this gradient hypothesis. This pattern was found for vowels that were nasalized for a relatively long portion of their duration (i.e., 50% or more) and contained an excrescent (short) nasal coda, two cues that can be considered conflicting within the same word.

On the other hand, the analysis of proportions of fixations did not provide firm evidence in favor of gradient nor categorical patterns of perception. In the analysis of fixations to the CV image, we found a trend towards significance between the baseline 50N and 100N %NASDUR. Fixations to the nasal CV image are higher in the 100N condition than in the 50N condition. However, when participants chose the CVN image, they fixated to the coarticulated image significantly less in the 100N condition than in the 50N condition. This suggests

that, depending on the interpretation of the spoken word, the actual duration of nasalization yielded to different patterns of fixations. Taken individually, results of the individual statistical models support more categorical patterns of perception. However, more generally, the proportions of fixations to each image depend on fine-grained phonetic details. Specifically, when the duration of nasalization canonically corresponds to a nasal vowel (i.e., 100N %NASDUR) and participants interpret it as such, they fixate to the target more than when the vowel is more ambiguous (i.e., partly nasalized). However, when they interpret the same fully nasalized (100N) vowel as a CVN token, they fixate to the target significantly less than when the vowel was only partly nasalized. Note that in the latter 100N–CVN pairing, the stimulus contains “mismatching” cues (i.e., a vowel that corresponds to a contrastive nasal vowel and a short nasal consonant in an isolated word).

In summary, image choice data provide clear support in favor of gradient interpretation of spoken words based on our %NASDUR continuum for stimuli that contain conflicting phonetic cues (i.e., long nasalization on the vowel and excrescent nasal coda). On the other hand, proportions of fixations did not reveal a clear pattern. Taken individually, nasal (CV) image choices and coarticulated (CVN) image choices suggested more categorical patterns of interpretation. However, the 100N %NASDUR condition behaved differently depending on the image choice, which in turn suggests variability in how phonetic detail is interpreted. Although still not a decisive gradient pattern, the data do not support a strict categorical perceptual pattern either. The key of the current finding, however, could be that listeners had difficulties interpreting the stimuli that were (sometimes) ambiguous (see below).

Further discussion of the link between word ambiguity and gradience will be provided in the General Discussion.

It is also important to remind that (oral) vowels followed by a nasal consonant are only optionally nasalized in Canadian French. The stimuli that were analyzed here were all nasalized, therefore not reflecting the entire spectrum of possible phonetic realizations of these vowels. Participants in the experiment were also presented with vowels that were not nasalized at all (i.e., 0N %NASDUR condition), but the stimuli were rejected from the analysis for two main reasons. Firstly, all these stimuli were expected to be categorized as coarticulated (CVN) words, which could create a ceiling effect and “artificially” increase the number of CVN responses. This would not be a problem per se, but would also not contribute to the analysis of nasal (C $\tilde{V}$ ) image choices and to the influence of vowel nasalization on coarticulatorily nasalized and contrastive nasal vowels. Secondly, the splicing procedure implied pasting the vowel from a CVC word (i.e., a word that did not contain a nasal consonant) into a CVN word matrix. However, due to limitations for word choice in the French lexicon, the final consonants of the CVC words did not all have the same place of articulation as the CVN words. Including these stimuli in the analysis could create some additional interference in the participants’ responses and eye movements. Exposure to these 0N stimuli could also yield to a “learning” problem, meaning that listeners’ responses could be influenced by the presence of misleading, and potentially unreliable, phonetic cues for responding to the analyzed stimuli over the course of the experiment, and extend to all heard stimuli. However, results concerning the unambiguous 20N %NASDUR condition, that is, a short-nasalized vowel and a short nasal consonant/appendix are not conflicting cues within a word, already are at floor performance (i.e., 16.4% C $\tilde{V}$  image choices in this 20N %NASDUR condition). This suggests that listeners did interpret 20N stimuli as CVN words, and that the presence of other ambiguous stimuli did not impact their performance in this specific %NASDUR condition. Furthermore, a subanalysis of the first four (out of 10) blocks of the experiment is presented in the [supplementary materials](#), evaluating the image choice patterns early in the experiment, before learning could have occurred, and the effect of %NASDUR. A discussion of the potential influence of stimulus habituation over the course of the experiment and its impact on recognition is also found in the [Supplementary Materials](#). Based on the results in the 20N %NASDUR and the early emergence of the %NASDUR effect during the experiment, we can thus reject the idea that listeners learned not to pay attention to the variations in nasalization duration over the course of the experiment, and reiterate the support for the gradient perception hypothesis when stimuli contained conflicting phonetic cues.

Finally, as mentioned above, the stimuli that were used in Experiment 1 were somewhat ambiguous, as some of the stimuli (e.g., 80N and 100N) had contradicting cues: a long-nasalized vowel and a short nasal consonant, which are not a

possible combination for isolated words in Canadian French if the consonant is interpreted as a full segment, even though the nasal appendix is pervasive in connected speech (Desmeules-Trudel, 2015; Desmeules-Trudel & Brunelle, 2018). However, in general, isolated words do not vary as much (Farnetani & Recasens, 2010), and it has never been shown, to our knowledge, that word-final nasal vowels in (Canadian) French have a nasal appendix. In order to investigate if the nasal appendix had a significant influence on perceptual patterns in addition to variations in nasalization duration, a second experiment was conducted with another group of L1 speakers of Canadian French. We used different stimuli to verify if the effects that were found in Experiment 1 also apply when stimuli do not have a word-final nasal appendix.

## Experiment 2

In Experiment 1, the words were presented in isolation for the recognition task. However, it is expected that nasal (C $\tilde{V}$ ) and coarticulated (CVN) vowels are realized differently and more constantly in isolation than when they are produced in connected speech (Farnetani & Recasens, 2010). For instance, in Experiment 1, it is likely that listeners expected phonological nasal ( $\tilde{V}$ ) vowels not to be followed by a nasal appendix in more careful speech. Stimuli from Experiment 1 were thus modified in Experiment 2 by removing the final nasal consonantal appendix. This allowed us to test whether “unambiguous” words (i.e., that do not have a nasal appendix) are processed differently than “ambiguous” ones (i.e., that contain both a nasalized vowel and a final nasal consonant) and allowed us to have a more thorough idea of how phonological nasal ( $\tilde{V}$ ) and coarticulated nasalized (VN) vowels are processed and eventually recognized by listeners of Canadian French, and how phonetic information is interpreted by the spoken-word recognition system.

## Method

### Participants

Twenty-four native speakers of Canadian French (18 female, six male), between 17 and 32 years of age ( $\bar{X}$  = 21.5 years,  $SD$  = 4.4), were paid or received partial course credit for their participation in the study. Twelve were from Ontario, and 12 from Québec. All listeners reported having normal hearing, normal or corrected-to-normal vision, and did not report any type or history of language, hearing, or speech impairment. All participants completed a background language questionnaire, and self-reported knowing English as a second language at a moderate to high level of proficiency, as well as other languages. Most participants were late bilinguals in English ( $N$  = 19), and some were early bilinguals in English ( $N$  = 5).

No speaker had been exposed to another language containing contrastive nasal vowels in its phonological inventory.

**Stimuli**

The words and images were the same word triads as in Experiment 1. The only modification that was performed on the prespliced stimuli was the removal of the final 50-ms nasal appendix. The same duration of nasalization conditions (20N, 50N, 80N, and 100N) were analyzed.

**Procedure and analysis**

The experimental procedure, including calibration, familiarization, stimuli presentation, and collected data was the same as in Experiment 1. The analysis (GAMMs) and investigated variables were the same as in Experiment 1.

**Results**

**Chosen images**

Figure 6 shows the model predictions for the image choice analysis, outputting the probability of choosing the C $\tilde{V}$  image in Experiment 2, with 50N %NASDUR as a baseline against which the other levels of the factor were tested. The actual output of the GAMM model is presented in Table 7 of the Appendix. In the figure, we observe that 20N %NASDUR yielded significantly lower probability of choosing the C $\tilde{V}$  image than 50N. On the other hand, the 80N and 100N %NASDUR values did not yield significantly different probabilities. This is consistent with the categorical perception

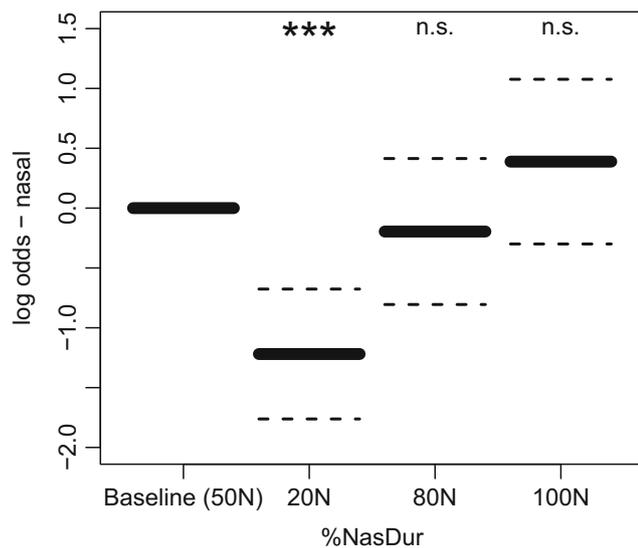
hypothesis, as opposed to the image choice pattern in Experiment 1. However, note that the absolute proportion of C $\tilde{V}$  image choices in the 20N %NASDUR condition (as opposed to choosing the CVN image) is still high, that is, 84.9%, and that the number of C $\tilde{V}$  choices in the other %NASDUR values is at ceiling, that is, 94.6% for 50N, 93.4% for 80N, and 96.1% for 100N. These proportions suggest that the significant difference that we found does not necessarily imply strict categorical perception of 20N %NASDUR stimuli à la Liberman et al. (1957), but rather that participants were not sensitive to (or did not use) variability in nasalization duration for vowels that were nasalized for 50% or more.

**Eye movements**

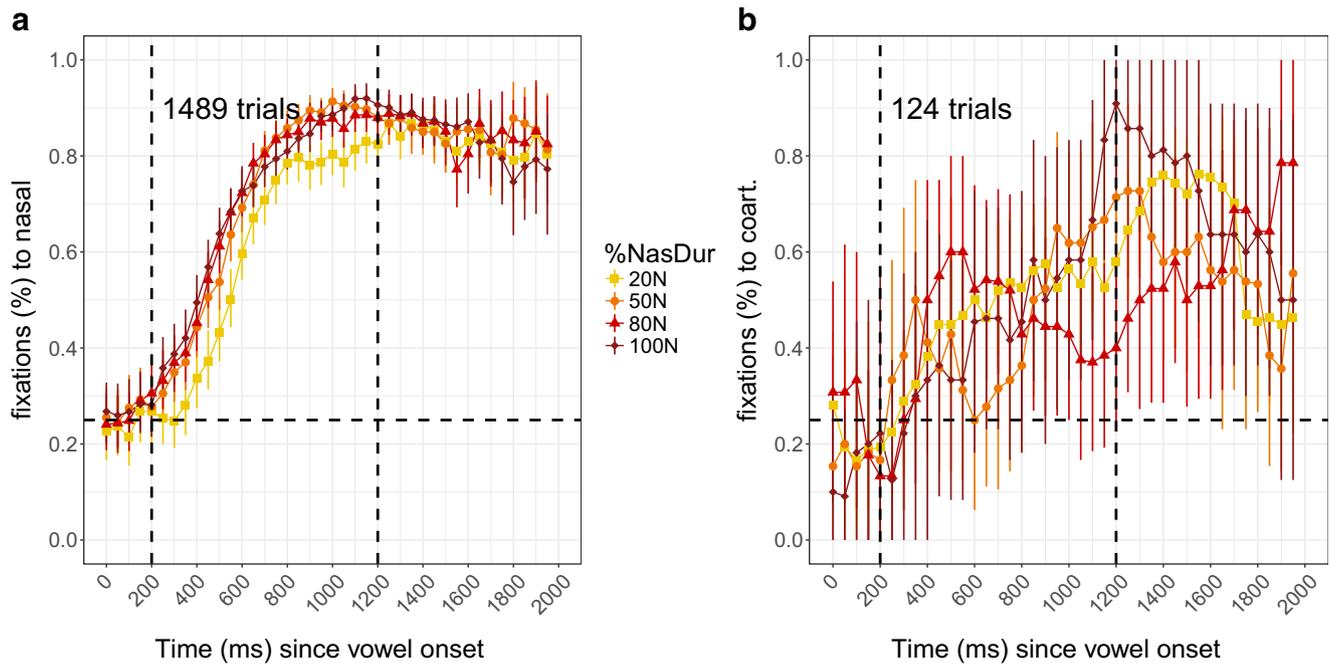
This subsection presents the eye-tracking results of Experiment 2, with the proportions of fixations to the nasal (C $\tilde{V}$ ) image when listeners chose the nasal image, and proportions of fixations to the coarticulated (CVN) image when they chose the coarticulated image between 200 (i.e., delay for programming an eye movement; Fischer, 1992) and 1,200 ms after vowel onset (between the dashed lines in Fig. 7).

Figure 7a shows that proportions of fixations in all %NASDUR conditions (with the exception of 20N early within the time window of analysis) are higher than chance (dotted line at 25%; four images on the screen) at the onset of the time window and increase quickly. This suggests that listeners quickly chose the C $\tilde{V}$  image in these trials, shortly after the stimulus vowel onset. In Fig. 7a, the 20N %NASDUR condition yields slightly lower proportions of fixations than the other conditions of Fig. 7a overall within the analysis window. This suggests that when the vowel was nasalized for a short period of time (20N %NASDUR), participants fixated slightly less to the nasal image than in the other %NASDUR conditions. This is expected, since a vowel that is nasalized early corresponds to this C $\tilde{V}$  image choice, while vowels that are nasalized for 20% of their duration do not canonically correspond to a contrastive nasal vowel (Desmeules-Trudel & Brunelle, 2018). In Fig. 7b, the small number of trials (i.e., 124 CVN choices out of 1,613 trials in Experiment 2, 7.7%) does not enable us to draw conclusions for fixations to CVN in CVN image choices. Indeed, the large error bars and overlap of several %NASDUR conditions prevents reliable interpretation of the results. More details about the patterns is provided in the statistical analysis below.

Results of the GAMM analysis of fixations to the C $\tilde{V}$  image (C $\tilde{V}$  image choices) and to the CVN image (CVN image choices) are presented in Fig. 8a and b, respectively (parametric %NASDUR factor), and Table 8 and Table 9, respectively, in the Appendix. The first part of the analysis focused on the %NASDUR factor (and interaction through TIME) to assess gradient or categorical patterns of fixations, presented here in Fig.



**Fig. 6** Experiment 2 GAMM predictions of choosing the C $\tilde{V}$  image, significance codes against the 50N %NASDUR baseline (n.s. = non-significant, \*\*\* $p < .001$ ). Note that the intercept is centered at zero



**Fig. 7** Experiment 2 proportions of fixations to the CṼ image when participants chose the CṼ image (a) and to the CVN image when they chose the CVN image (b). The interval between dashed vertical lines

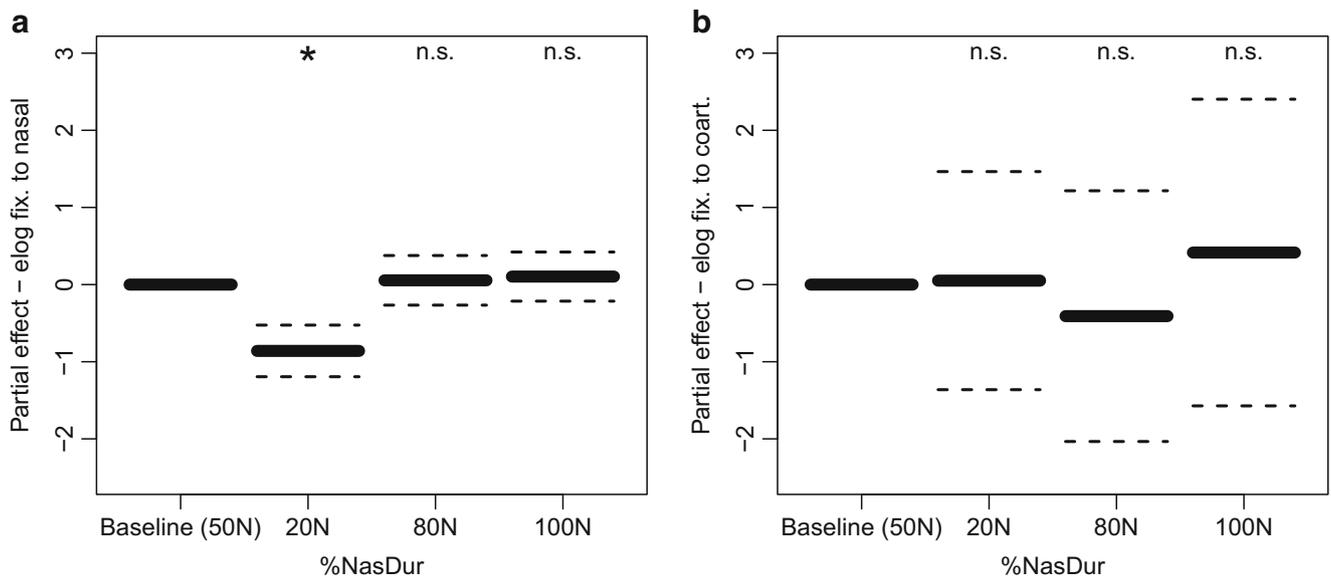
represent the time window of the statistical analysis below, and the horizontal dashed line represents chance at 25% (four images on the screen). (Color figure online)

8. The *p* values of parametric factors lower than .05 suggest a significant to the (50N %NASDUR) baseline.

Figure 8a shows that, as expected based on the observation of the raw data in Fig. 7a, only the 20N %NASDUR yielded significantly lower proportions of fixations to the CṼ image than the other %NASDUR values when participants chose the CṼ image. This supports a more categorical pattern of

fixations to the CṼ image, since listeners did not differentiate across nasalization duration values when the vowel was nasalized for 50% or more.

On the other hand, in Fig. 8b, the difference between overall fixations to the coarticulated (CVN) image (CVN image choices) in the (baseline) 50N %NASDUR is not significantly different from any of the other %NASDUR levels. Again, based



**Fig. 8** Experiment 2 GAMM predictions of fixations to the CṼ image (CṼ image choices) in Panel a, fixations to the CVN image (CVN image choices) in Panel b, and significance codes against the 50N %NASDUR baseline (n.s. = non-significant, \**p* < .05). Note that the intercept is centered at zero

on the observation of raw data in Fig. 7b, this is expected since there were very few trials to model and confidence intervals are larger (compare dashed lines in Fig. 8a and b). This suggests that participants did not find that the stimuli corresponded to CVN images in general, particularly given that there was no nasal consonant or excrescence at the end of the presented words.

## Discussion

The results of Experiment 2 show that nasalization duration had a significant impact on Canadian French listeners performance in the spoken-word recognition task. This result was expected based on previous results in English (Beddor et al., 2013) and Canadian French (Experiment 1). For example, results show that when a vowel is nasalized for a long part of its duration (i.e., in the 50N, 80N, and 100N %NASDUR), participants overwhelmingly identified the stimuli as C $\tilde{V}$  words (between 93% and 97% of the time). However, when participants heard a word with a vowel that was nasalized for a shorter period (i.e., 20N %NASDUR), the rate of identification as C $\tilde{V}$  was significantly lower (85%) than in the 50N %NASDUR.

A significant effect of %NASDUR was also observed in the statistical analysis of proportions of fixations to the nasal (C $\tilde{V}$ ) image. When the vowel was nasalized for a short period (20N), participants did not fixate the nasal (C $\tilde{V}$ ) word as much through time as in the other three conditions (50N, 80N, 100N) when they chose this latter image. Proportions of fixations to the nasal (C $\tilde{V}$ ) image was highly similar for the latter three conditions (see Fig. 7a). These results suggest that when a vowel is nasalized for 50% of its duration or more, processing is similar, and variability past this threshold does not seem to impact recognition for French listeners. This is consistent with the production of phonological nasal vowels ( $\tilde{V}$ ) in Canadian French (Desmeules-Trudel & Brunelle, 2018), for which duration of nasalization is variable but always starts between vowel onset and 50% of its duration. However, when a vowel is nasalized for a short period only (20N), processing seems different. In production (Desmeules-Trudel & Brunelle, 2018), vowels that are nasalized for 20% or less of their duration are always followed by a nasal consonant (CVN word). Therefore, a vowel that is nasalized for a short period and not followed by a nasal consonant, as in the current stimuli, does not correspond to either a coarticulated (CVN) vowel, nor a phonological nasal (C $\tilde{V}$ ) vowel. This likely makes recognition of the auditory word less consistent than when the physical characteristics correspond to the realization of one of the categories, and likely explains the slightly different behavior of the 20N %NASDUR condition in the word choices and eye-movement data.

Experiment 2 was also designed to verify if phonetic cues are gradually or categorically integrated when stimuli did not contain a final nasal appendix, which causes ambiguity in some cases (i.e., nasalization on the vowel and nasal

appendix, as in Experiment 1). As mentioned above, removing the nasal consonant at the end of the vowel for some stimuli disambiguated their phonological status for the 50N, 80N, and 100N %NASDUR. However, note that this modification made the 20N stimuli correspond to none of the presented word choices, therefore stimuli in this condition can be qualified as ambiguous. Keeping in mind the differences across %NASDUR regarding the “ambiguity” of the stimuli, the predictions of the categorical hypothesis are only partially confirmed. For instance, when nasalization duration is below the threshold of 50% nasalization (which corresponds to the minimal amount of nasalization necessary for recognizing the vowel as phonologically nasal; Desmeules-Trudel & Brunelle, 2018), fixations to the nasal (C $\tilde{V}$ ) word are still high. Thus, the interpretation of vowel nasalization could be gradient, especially if further %NASDUR values were tested with the current design and if stimulus ambiguity has an influence, as suggested in Experiment 1, but this hypothesis was not confirmed by the data in Experiment 2. It is thus a possibility that intermediate steps (e.g., additional %NASDUR values in 5% increments, such as 10N, 15N, 25N, 30N) could lead to more finely grained gradient patterns of word recognition and phonetic integration in Experiment 2—further investigation will be necessary to confirm this speculation.

Our main conclusion remains that when the identity of the vowel is disambiguated (once the 50% nasalization threshold is reached and the vowel is not followed by an ambiguous nasal appendix), the identification of the stimuli seems at ceiling, and by extension more categorical. Participants consistently identified these stimuli as phonological nasal (C $\tilde{V}$ ) words, and the time course of fixations seems similar across conditions. Note that in these conditions (50N, 80N, 100N), there seems to be a ceiling effect, which shows that the task was easier in these cases.

## General discussion

Two experiments showed that the duration of vowel nasalization has a major impact on the recognition of words containing a contrast between phonological nasal (C $\tilde{V}$ ) and coarticulatory nasalized (CVN) vowels in Canadian French. This influence was predicted based on the production of the contrast in Canadian French (Desmeules-Trudel & Brunelle, 2018), and on previous results obtained in English by Beddor et al. (2013) and Zamuner et al. (2016). For instance, Desmeules-Trudel and Brunelle found that phonological nasal vowels are produced significantly differently from coarticulated nasalized vowels in Canadian French throughout their duration. They also found that nasalization of phonological nasal vowels (C $\tilde{V}$ ) starts between the vowels onsets and 50% of their duration, and nasalization of coarticulated nasalized vowels (CVN) starts later. Participants recognized vowels for which

nasalization started between onset and 50% of their duration as  $C\tilde{V}$  more often than when it started later in the vowel, especially in Experiment 2 which contained unambiguous stimuli. However, due to conflicting cues (i.e., ambiguity) in the words in Experiment 1, the response patterns were strikingly different, but the general finding that an increase in duration of nasalization yielded more  $C\tilde{V}$  responses in both experiments still held. We draw a parallel between Beddor et al.'s (2013) participants who were more efficient at anticipating an upcoming nasal consonant when the vowel was nasalized early, and the participants to the current experiments who tended to recognize words that contained a phonological nasal vowel ( $C\tilde{V}$ ) in a greater proportion when the stimuli were nasalized early. It has also been repeatedly demonstrated that a language's coarticulatory patterns in production are tightly linked to the perceptual patterns of native listeners (Beddor, Harnsberger, & Lindemann, 2002; Beddor & Krakow, 1999), as convincingly shown by Beddor and Krakow (1999) who underlined systematic differences between English and Thai native speakers' perception of coarticulatory nasalization. Since English and Thai have different implementations of coarticulatory vowel nasalization, Beddor and Krakow (1999) attribute the group differences in perception to the production of these vowels. This is also the case in the current study, which shows that perceptual patterns of vowel nasalization in Canadian French mirror the production of vowel nasalization in this language.

In addition to the general influence of nasalization timing on recognition, by varying the duration of nasalization on the vowel, we found support in favor of gradient interpretation of variations in time-dependent phonetic information (McMurray et al., 2008; McMurray et al., 2002), especially when the stimuli contained mismatching phonetic information (i.e., a long-nasalized vowel followed by a short nasal consonant). Indeed, in Experiment 1, the gradient pattern of recognition emerged for word choice patterns when a nasalized vowel was followed by a nasal consonantal appendix. Furthermore, fully nasalized vowels were interpreted differently from vowels that were nasalized for 50% of their duration regardless of the image that the participants chose, but the direction of the effect was different. This also supports some kind of gradience (or *noncategoricity*) in the interpretation of nasalization, as reflected through eye movements. This gradience emerged early in the experiment, suggesting that participants' word recognition/speech perception system is sensitive to variability in nasalization duration, regardless of other task or potential learning effects. In Experiment 2, a more categorical pattern was observed, and conformed to our predictions. In the cases for which vowels are nasalized for a short period of time (20N), listeners expect the full realization of a nasal consonant to explain the short period of nasalization on the vowel, and therefore chose the  $C\tilde{V}$  image less (and fixated it less as well) than in the other conditions. However, observing the raw proportions of image choices and fixations, it seems like

the pattern is not strictly categorical. Vowels that were nasalized for a short period of time still yielded an important proportion of  $C\tilde{V}$  image choices (and fixations to  $C\tilde{V}$ ). Consequently, it seems like gradient-like recognition emerged for words that contained ambiguous cues or lacked sufficient cues. On the other hand, when stimuli did not contain conflicting cues, more categorical patterns of recognition emerged, though the influence of additional nasalization duration values should be investigated in future research. This is an important finding, that listeners are able to use a great amount of fine-grained phonetic information, but only when it is necessary to disambiguate or identify "incomplete" stimuli.

In addition to providing support for the gradient interpretation hypothesis, our results provide support for the storage of fine-grained phonetic cues in phonological representations (McMurray et al., 2002), as listeners were (sometimes) able to consider short variations in the signal in order to categorize the stimuli, even though the phonetic cue is involved in contrasting vowel phonemes, similarly to studies on VOT (McMurray et al., 2008; McMurray et al., 2002). The fact that listeners are able to use fine-grained phonetics for higher order language processing has been found in English using a variety of phonetic cues, such as VOT and length of vowel transitions (McMurray et al., 2008), direction of vowel transitions (Dahan et al., 2001), and vowel nasalization (Beddor et al., 2013; Zamuner et al., 2016), as well as in bilingual adults (Desmeules-Trudel, 2018). Furthermore, listeners are able to use these cues, but do not seem to do so when stimuli are simpler. On the other hand, when the task demands are greater or when the stimuli present conflicting cues, listeners can rely on more fine-grained strategies for interpreting spoken words.

Finally, we will be interested in addressing the question of phonetic integration in a variety of languages that have a phonological contrast between oral and nasal vowels, along with a process of coarticulatory nasalization. As it has been shown that coarticulatory nasalization can be speaker controlled (Beddor, 2009; Cho, Kim, & Kim, 2017), the realization of this phonetic cue can be variable across languages and individuals. Assessing the perception of this property in the context of word recognition across more languages will provide greater insights into the general architecture of the phonological recognition system, and enable us to have a better understanding of how phonological contrast based on one phonetic cue impacts phonetic integration. Furthermore, the general finding that word ambiguity demands access to more fine-grained phonetic information, as opposed to unambiguous words which do not require as much resources, has implications for connected speech processing. If words that are presented in the context of other words but are still ambiguous (both in terms of phonetic variability and phrasal context), it would be interesting to determine if listeners still have access to fine-grained, low-level phonetic information, or if they tend to recognize the words in a more categorical way.

## Appendix

**Table 3.** Auditory stimuli details

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

**Table 4.** Experiment 1 chosen image GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	-0.8108	0.3999	-2.0273	.0426*
%NASDUR (20N)	-1.0994	0.1931	-5.6924	<.001***
%NASDUR (80N)	0.3473	0.1715	2.0253	.0428*
%NASDUR (100N)	0.6812	0.1701	4.0057	.001**
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
Random (participant)	19.3899	22	212.0043	<.001***
Random (item)	7.7188	8	252.6659	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001

**Table 5.** Experiment 1 fixations to CV̄ (526 CV̄ image choices) GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	0.0384	0.234	0.1641	.8696
%NASDUR (20N)	-0.1296	0.4003	-0.3238	.7461
%NASDUR (80N)	-0.0551	0.3183	-0.173	.8627
%NASDUR (100N)	0.544	0.307	1.772	.0764
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
s(TIME)	1.0137	1.0197	5.0862	.024*
s(TIME):%NASDUR (50N)	3.4056	4.4313	1.3827	.2298
s(TIME):%NASDUR (20N)	1.5413	2.3093	1.1739	.4254
s(TIME):%NASDUR (80N)	3.2922	4.2877	1.3038	.2771
s(TIME):%NASDUR (100N)	6.0643	7.4329	1.9707	.0408*
Random (item)	291.0753	519	1.2892	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001**Table 6.** Experiment 1 fixations to CVN (969 CVN image choices) GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	1.2894	0.1656	7.7856	<.001***
%NASDUR (20N)	-0.2671	0.225	-1.2009	.2298
%NASDUR (80N)	-0.2490	0.2435	-1.0225	.3065
%NASDUR (100N)	-0.5649	0.2464	-2.2926	.0219*
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
s(TIME)	5.2872	6.6003	10.7944	<.001***
s(TIME):%NASDUR (50N)	2.5137	3.2652	1.0458	.3201
s(TIME):%NASDUR (20N)	1.0364	1.067	0.668	.4028
s(TIME):%NASDUR (80N)	5.0041	6.3353	3.3109	.0026**
s(TIME):%NASDUR (100N)	1.041	1.0777	0.8011	.3894
Random (item)	601.032	981	1.6047	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001**Table 7.** Experiment 2 chosen image GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	3.2758	0.4799	6.8262	<.001***
%NASDUR (20N)	-1.2192	0.2714	-4.4930	<.001***
%NASDUR (80N)	-0.1956	0.3049	-0.6416	.5211
%NASDUR (100N)	0.3884	0.3440	1.1290	.2589
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
Random (participant)	15.0763	23	46.4411	<.001***
Random (item)	7.2095	8	58.8859	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001**Table 8.** Experiment 2 fixations to CV̄ (1489 CV̄ image choices) GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	2.0357	0.1135	17.9385	<.001***
%NASDUR (20N)	-0.8599	0.1672	-5.1435	<.001***
%NASDUR (80N)	0.0551	0.1607	0.3427	.7318
%NASDUR (100N)	0.1037	0.1592	0.6512	.5149
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
s(TIME)	7.6009	8.4086	20.1675	<.001***
s(TIME):%NASDUR (50N)	4.1155	5.2565	1.5365	.1472
s(TIME):%NASDUR (20N)	4.4730	5.7735	4.598	<.001***
s(TIME):%NASDUR (80N)	1.0455	1.0642	0.6432	.4146
s(TIME):%NASDUR (100N)	5.5412	6.8375	2.8802	<.01**
Random (item)	874.0988	1,491	1.4349	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001

**Table 9.** Experiment 2 fixations to CVN (124 CVN image choices) GAMM summary

Parametric coefficients	Estimate	SE	<i>t</i> value	<i>p</i> value
Intercept (Baseline 50N)	−0.1691	0.6067	−0.2787	.7805
%NASDUR (20N)	0.0511	0.7068	0.0723	.9423
%NASDUR (80N)	−0.409	0.8125	−0.5034	.6148
%NASDUR (100N)	0.4153	0.9938	0.4179	.6761
Smooth terms	Edf	Ref.edf	<i>F</i> value	<i>p</i> value
s(TIME)	3.5382	4.5734	2.0183	.0895
s(TIME):%NASDUR (50N)	4.3518	5.5659	1.7771	.1145
s(TIME):%NASDUR (20N)	1.0041	1.0061	0.052	.819
s(TIME):%NASDUR (80N)	3.1818	4.1376	1.1675	.3216
s(TIME):%NASDUR (100N)	2.3267	3.3239	4.1477	<.01**
Random (item)	91.1194	130	2.3588	<.001***

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001

**Author Note** We are grateful to Dr. Bob McMurray, Dr. Marc Brunelle, Dr. Laura Sabourin, and Dr. Kevin McMullin for their comments on earlier versions of this work. We also want to thank Émilie Piché for her help with data collection, research teams at the University of Ottawa’s Centre for Child Language Research and Sound Patterns Laboratory, audiences at the LabPhon 16 Conference, 2017 CLA meeting, MOLT workshop 2016, and two anonymous reviewers. F.D.-T. benefited from doctoral scholarships from the Fonds de recherche du Québec - Société et Culture and Social Sciences and Humanities Research Council of Canada (Bombardier scholarship). All remaining errors are our own.

## References

- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.
- Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-effects regression models in linguistics* (pp. 49–69). Cham: Springer.
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821.
- Beddor, P. S., Hamsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4), 591–627.
- Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America*, 106(5), 2868–2887.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, 133(4), 2350–2366.
- Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer software]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Carignan, C. (2013). *When nasal is more than nasal: the oral articulation of nasal vowels in two dialects of French* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- Carignan, C. (2014). An acoustic and articulatory examination of the “oral” in “nasal”: The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics*, 46(1), 23–33.
- Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*, 64, 71–89.
- Cohn, A. C. (1990). *Phonetic and phonological rules of nasalisation* (Unpublished doctoral dissertation). University of California, Los Angeles.
- Côté, M.-H. (2012). Laurentian French (Québec): Extra vowels, missing schwas, and surprising liaison consonants. In R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological variation in French: Illustrations from three continents* (pp. 235–274). Amsterdam: John Benjamins.
- Cross, A. M., & Joanisse, M. F. (2018). Eyetracking of coarticulatory cue responses in children and adults. *Language, Cognition, and Neuroscience*, 33(10), 1315–1324.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5/6), 507–534.
- Delvaux, V. (2006). Production des voyelles nasales en français québécois [Production of nasal vowels in Quebec French]. In *Actes Des 26<sup>es</sup> Journées D’études Sur La Parole*, 383–386.
- Desmeules-Trudel, F. (2015). Propriétés aérodynamiques des voyelles nasales et potentiellement nasalisées en français québécois [Aerodynamic properties of Québécois French nasal and potentially nasalized vowels]. In S. Vinerte (Ed.), *Proceedings of the 2015 Annual Conference of the Canadian Linguistics Association*. Ottawa, ON: Canadian Linguistics Association.
- Desmeules-Trudel, F. (2018). *Spoken word recognition in native and second language Canadian French: Phonetic detail and representation of vowel nasalization* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.
- Desmeules-Trudel, F., & Brunelle, M. (2018). Phonotactic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Québécois French and Brazilian Portuguese. *Journal of Phonetics*, 69, 43–61.
- Desmeules-Trudel, F., Moore, C., & Zamuner, T. (2019). *Monolingual and bilingual children’s processing of coarticulation cues during spoken word recognition*. Manuscript submitted for review.
- Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd, pp. 316–352). New York: Wiley-Blackwell.

- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). New York: Springer.
- Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113–133.
- Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68(2), 161–177.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, 5, 275–292.
- Lahiri, A., & Marslen-wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.
- Léon, P. R. (1983). Les voyelles nasales et leurs réalisations dans les parlers français du Canada [Nasal vowels and their realizations in spoken French in Canada]. *Langue Française*, 60, 48–64.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
- Martin, P. (2002). Le système vocalique du français du Québec: De l'acoustique à la phonologie [The vowel system of Québec French. From acoustics to phonology]. *La Linguistique*, 38(2), 71–88.
- Martin, P., Beaudoin-Bégin, A.-M., Goulet, M.-J., & Roy, J.-P. (2001). Les voyelles nasales en français du Québec [Nasal vowels in Québec French]. *La Linguistique*, 37(2), 49–70.
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, 15(6), 1064–1071.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42.
- Paquette-Smith, M., Fecher, N., & Johnson, E. K. (2016). Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, & Psychophysics*, 78(8), 2329–2340.
- Porretta, V., Kyröläinen, A.-J., van Rij, J., & Järviö, J. (2017). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent decision technologies 2017* (pp. 268–277). Cham: Springer.
- Porretta, V., Tucker, B. V., & Järviö, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, 58, 1–21.
- R Core Team. (2017). The R Project for Statistical Computing [Computer software]. Retrieved from <https://www.r-project.org>
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71(1), 145–163.
- Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 114–175). Malden: Wiley-Blackwell.
- Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247–250.
- van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 267–293). Berlin: De Gruyter.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2016). *itsadug: Interpreting time series and autocorrelated data using GAMMs* [R package]. Retrieved from <https://rdrr.io/cran/itsadug/man/itsadug.html>
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton: Chapman and Hall/CRC.
- Zamuner, T. S., Moore, C., & Desmeules-Trudel, F. (2016). Toddlers' sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology*, 152, 136–148.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.