



Use of gradient anticipatory nasal coarticulatory cues for lexical perception in French

Georgia Zellou*, University of California, Davis, US, gzellou@ucdavis.edu

Anne Pycha, University of Wisconsin, Milwaukee, US, pycha@uwm.edu

Ioana Chitoran, Université Paris Cité, Laboratoire Clillac-ARP, FR, ioana.chitoran@u-paris.fr

*Corresponding author.

The current study examines anticipatory nasal coarticulation in French, a language which is known to exhibit a) quality differences between phonemically oral and nasal vowels and b) relatively low amounts of nasal coarticulation in CVN contexts. In a production study, thirty native Northern Metropolitan (Parisian) French speakers produced seven sets of C \tilde{V} -CVC-CVN words (e.g., [sed] *cède*, [sen] *scènes*, [sæ̃] *saint*). Consistent with previous studies, results indicated quality differences between the vowels in C \tilde{V} versus CVC and CVN words, and also that acoustic nasalization in CVN contexts was relatively small; nevertheless, it was still significantly greater than in CVC contexts, and variable across speakers. In a perception study, the CV portions of the production recordings were played to fifty French listeners, who identified the corresponding word in a forced-choice task. Results showed that stimuli from CVN contexts were highly confusable with CVC items, but not with C \tilde{V} items. Most importantly, increased degree of acoustic nasalization on individual CVN stimuli significantly correlated with accuracy. We conclude that, despite the overall relative weakness of coarticulatory cues to nasality, French listeners can nevertheless employ these cues when they are present.



1. Introduction

Spoken language exhibits a constant interplay between physical events and abstract categories. We can see this interplay in speech production, where vocal tract articulations are necessary in order to instantiate speech sounds. For example, lowering of the velum is required in order to realize phonemically nasal vowels (\tilde{V}). Less obviously, physical events may be constrained by the system of abstract categories in which they are embedded. For example, in sequences of a vowel followed by a nasal consonant (VN), coarticulation between adjacent sounds can give rise to velum-lowering during the vowel, such that the oral vowel becomes nasalized to some extent. When this type of nasalization occurs in a language whose inventory includes phonological vowel nasality, it potentially threatens the maintenance of phonemic contrasts. In such cases, it has been proposed that speakers may control their coarticulation (e.g., Manuel, 1990; for discussion, see Solé, 1992, 2007), such that the vowel in VN sequences is only lightly nasalized. As Manuel (1990) acknowledged, coarticulation may also be modulated by abstract factors that are less immediately evident. Kleurling Afrikaans, for example, lacks phonological vowel nasality, yet exhibits only light nasalization in VN (Coetzee et al., 2022); meanwhile, Brazilian Portuguese (Clumeck, 1976) and Lakota (Scarborough et al., 2015) possess phonological vowel nasality alongside extensive nasalization in VN. The point is that physical events are constrained, via either the inventory or other language-specific factors, by the phonological grammar in which they occur.

The interplay between abstract categories and physical events is also evident in speech perception. Numerous studies have demonstrated that fine-grained coarticulatory cues help listeners perceive the presence of particular speech sounds (e.g., Dahan et al., 2001; Martin & Bunnell, 1981; Salverda et al., 2014; Whalen, 1984; Scarborough & Zellou, 2013; Zellou et al., 2023). This applies to VN sequences, where cues to nasalization on the vowel can help listeners correctly identify the sequence as VN (e.g., Malécot, 1960; Beddor et al., 2013; Fowler & Brown, 2000). Here again, however, this use of acoustic cues may be modulated by the system of abstract categories in which they are embedded. For example, in languages with phonemically nasal vowels, cues to nasalization on a vowel may lead listeners to interpret the signal as \tilde{V} , rather than VN (Lahiri & Marslen-Wilson, 1991; see also Kotzor et al., 2022; Stevens, Andredde, & Viana, 1987). Thus, perception – even of fine-grained cues – may also be constrained by the grammar.

These issues come to a head in the French language, whose sound system employs phonemically oral vowels (*cède* [sɛd]), phonemically nasal vowels (*saint* [sɛ̃]), and sequences of oral vowels followed by nasal consonants (*scènes* [sɛn]). Previous production research has shown that French CVN sequences exhibit relatively low degrees of coarticulatory vowel nasalization (e.g., Cohn, 1990; Delvaux et al., 2008; Dow, 2020), consistent with the idea that coarticulation may be reduced for grammatical reasons, such as to avoid confusability with phonological vowel nasality (cf. Manuel, 1990). In parallel, previous perception research suggests that French

listeners behave accordingly. For example, they assign lower nasality ratings to vowels from CVN contexts, compared to C \tilde{V} contexts (Benguerel & Lafargue, 1981). Furthermore, at least for Canadian French listeners, when the temporal duration of nasality on a vowel decreases, they are more likely to select a picture depicting a CVN lexical item, compared to a picture of a C \tilde{V} lexical item (Desmeules-Trudel & Zamuner, 2019). These findings might lead us to conclude that, unlike what has been shown for other languages such as American English (Beddor et al., 2013), French listeners do not – and arguably, cannot – use coarticulatory nasalization to interpret the speech signal as VN.

And yet there are at least three reasons to suspect that this conclusion is incorrect. First, French nasal vowels differ systematically in quality from their oral counterparts (Delvaux, 2009). Although they are traditionally described using the phonemic representations / $\tilde{\epsilon}$, $\tilde{\jmath}$, $\tilde{\alpha}$ / (Dow, 2020, and references cited therein), nasal vowels in Parisian French have been shown to be produced as more retracted (lower F2), and with lower tongue position (higher F1) for / $\tilde{\epsilon}$ / or higher tongue position (lower F1) for / $\tilde{\alpha}$ / and / $\tilde{\jmath}$ / (Carignan, 2014). On the basis of this evidence, one researcher has suggested that the nasal counterparts of oral / ϵ , \jmath , α / are actually realized as [$\tilde{\alpha}$, $\tilde{\sigma}$, $\tilde{\jmath}$] (Dow, 2020 and references cited therein) while another has suggested the surface transcriptions [$\tilde{\jmath}$, $\tilde{\epsilon}$, $\tilde{\sigma}$] (Carignan, 2014). The vowel quality differences are evident in gender alternations across stems such as *pleine* ~ *plein* [plɛn ~ pl $\tilde{\alpha}$] ‘full’ and *pionne* ~ *pion* [pjɔn ~ pj $\tilde{\sigma}$] ‘hall monitor’ ~ ‘pawn’, as well as non-semantically related word pairs such as *clame* [klam] ‘proclaim-3s’ versus *clan* [kl $\tilde{\alpha}$] ‘clan’.

Not surprisingly, French listeners take advantage of these quality differences during perception: in order to identify V versus \tilde{V} , they rely not just on acoustic correlates of nasality, but equally heavily on F2 (Delvaux, 2009). This multifaceted situation opens up certain possibilities. In VN sequences, for example, the F2 cues on the vowel may be sufficient to establish that the vowel is phonemically oral, and not nasal. If that were the case, French listeners would be free to use any coarticulatory nasalization on the vowel to predict an upcoming nasal coda – just as listeners do in languages that lack phonemic \tilde{V} .

Second, even though French words with VN sequences exhibit low degrees of coarticulatory vowel nasalization overall, studies have reported that there is substantial variation from one speaker to the next (Styler, 2017). Thus, even if many or most VN utterances offer impoverished opportunities for interpreting the signal as nasalized, some portion of utterances may offer richer opportunities. If that is the case, we expect that listeners may take advantage of these opportunities, by using coarticulatory cues *when they are present*, as Zellou (2022) recently demonstrated for American English.

Finally, at least two previous studies of other languages have demonstrated that listeners can use coarticulatory cues, even when those cues are relatively weak. Beddor et al. (2013) demonstrated that although American English listeners responded more quickly to coarticulatory

nasalization when it began early in the vowel, they also used these cues even when it occurred late. In a similar vein, Coetzee et al. (2022) demonstrated that although Afrikaans listeners responded more quickly to stimuli from White Afrikaans, a dialect in which coarticulatory nasalization begins early, they also used these cues in stimuli from Kleurling Afrikaans, a dialect in which nasalization begins late.

Putting all of this together, we ask: in those cases where Parisian French listeners *could* potentially use acoustic nasalization cues to interpret the signal as a VN sequence, do they *actually* do so? That is, despite a situation that would seem to mitigate against the use of this physical cue, do French listeners nevertheless make use of coarticulatory nasalization? In the current paper, we address this question by leveraging the natural variability exhibited across speakers in production. In a production study, we recorded thirty native Northern Metropolitan French speakers producing seven sets of CVC-CVN-C \tilde{V} words (e.g., [sɛd] *cède* ‘give up-3s’, [sɛn] *scènes* ‘scenes’, [sɛ̃] *saint* ‘saint’). We conducted an acoustic analysis of the vowels, with emphasis on oral versus nasal quality differences, as well as individual variability among speakers in degree of produced nasal coarticulation. In a perception study, we took the recordings from all thirty speakers and excised the codas, where present. We played the CV portions to fifty French listeners in a forced-choice lexical identification task in which they selected which of the three lexical options was the originally intended word. For example, after hearing a stimulus [sɛ] from the context [sɛn], listeners indicated whether the intended word was *cède*, *scènes*, or *saint*.

The data allow us to pursue two research questions, crucially within the context of interspeaker variability. First, are CVN stimuli significantly confusable with C \tilde{V} stimuli? Given the situation of French, in which formant cues can help eliminate spurious perceptual interpretations of \tilde{V} , we predict that the answer will be no, a scenario which creates the potential for listeners to make use of nasality cues to distinguish between CVN and CVC. In contrast, since vowels in CVC and CVN words are actually the most acoustically similar in terms of formant values, and not substantially different in terms of vowel nasalization, we predict that coarticulated and oral vowels will be most confusable. Second, are the particular tokens of CVN stimuli that contain greater coarticulatory nasality more likely to be identified correctly, compared to the CVN tokens that contain less? We predict the answer will be yes, a result which would demonstrate that listeners use coarticulatory cues even in relatively unlikely circumstances.

Our data will also speak to theoretical issues within the literature. Previous work has investigated the nature of the mapping from the acoustic signal onto lexical items, and the potential role played by abstract features (Eulitz & Lahiri, 2004; Lahiri & Reetz, 2002, 2010). Important supporting evidence has come from studies that focused specifically on acoustic nasality and its potential interaction with features such as [Nasal] (Kotzor et al., 2022; Lahiri & Marslen-Wilson, 1991). Our experiments in French will contribute to this discussion by examining how listeners interpret nasality in a novel context, namely, in a language where oral and nasal vowels differ crucially in quality.

Note that we remain agnostic as to which particular phonetic transcription is correct for nasal vowels (cf. Carignan, 2014; Dow, 2020), and we have used [ã, õ, ɔ̃] solely for visual ease. When we refer to vowel types, for example in our statistical analyses, we use phonemic transcriptions in slash brackets (thus, e.g., the category /ɛ/ refers to both oral [ɛ] and nasal [ɛ̃]); meanwhile, when referring to individual words, we use phonetic transcriptions in square brackets.

2. Production study

2.1. Methods

2.1.1. Participants

Thirty native Northern French speakers (mean age = 21.2 years old; range 19–31; 20 female, 0 non-binary or other, 10 male) participated. Participants were students at Université Paris Cité, recruited via flyers and emails. All participants completed informed consent approved by the Université Paris Cité Comité d'Éthique de la Recherche (CER).

2.1.2. Word list

The word list for this study consisted of seven sets of CVC-CVN-C \tilde{V} French words containing non-high vowels ([ɛ, ɔ, a] and nasal counterparts [ɛ̃, õ, ɔ̃]), shown in **Table 1**. The coda place of articulation for the CVC and CVN items were matched as closely as possible. As can be seen in **Table 1**, the stimulus onsets could be either simple (e.g., [sɛd]) or complex (e.g., [plɛd]). For visual ease, and because we do not expect onset structure to play a role in our results, we refer to both types of onsets with a single C.

CVC Oral vowels	CVN Coarticulated vowels	C \tilde{V} Nasal vowels
[plɛd] <i>plaid</i> 'plead-3s'	[plɛn] <i>plaine</i> 'plain (noun)'	[plɛ̃] <i>plein</i> 'full-Masc'
[sɛd] <i>cède</i> 'give up-3s'	[sɛn] <i>scènes</i> 'scenes'	[sɛ̃] <i>saint</i> 'saint'
[pjɔʃ] <i>pioche</i> 'pickaxe'	[pjɔn] <i>pionne</i> 'hall monitor'	[pjõ] <i>pion</i> 'pawn'
[trɔk] <i>troque</i> 'barter-3s'	[trɔɲ] <i>trogne</i> 'face (comical)'	[trõ] <i>tronc</i> 'trunk'

(Contd.)

CVC Oral vowels	CVN Coarticulated vowels	C\tilde{V} Nasal vowels
[kɫap] <i>clappe</i> 'click-3s'	[kɫam] <i>clame</i> 'proclaim-3s'	[kɫ̃] <i>clan</i> 'clan'
[fad] <i>fade</i> 'tasteless'	[fan] <i>fane</i> 'fade-3s'	[f̃] <i>fend</i> 'split-3s'
[krab] <i>crabe</i> 'crab'	[kr̃am] <i>crame</i> 'burn-3s'	[kr̃ɔ̃] <i>cran</i> 'notch (noun)'

Table 1: Word list used in the study.

2.1.3. Procedure

Participants were seated in a sound-proof booth facing a computer running a Qualtrics survey. They were informed that they would be producing sentences, in the form of instructions about how to organize a set of words onto different lists. On each trial, participants saw one target sentence and read it aloud (e.g., *Écrivez le mot 'scènes' sur la première liste.* "Write the word 'scènes' on the first list."). The sentences were presented in random order for each speaker. Participants' productions were recorded using an Audio-Technica ATM33a microphone and USB audio mixer (Sound Devices, USB Pre 2) and digitized at a 44.1 kHz sampling rate using Audacity. Each target word was presented once for each speaker, yielding a total of 630 productions (3 word types \times 7 stimulus sets \times 30 participants).

2.1.4. Acoustic measures

Individual words and speech sounds were initially segmented using the Montreal Forced Aligner (McAuliffe et al., 2017). All of the boundaries were then hand-verified, and corrected where necessary by two phonetically-trained researchers (one annotator hand-corrected any boundaries using Praat, a second annotator verified that work). **Figure 1** displays a sample segmented waveform and spectrogram. Ten words which had been mispronounced were excluded, yielding 620 items for the final analysis.

Three acoustic measurements were obtained via Praat from each of the vowels.

First, we measured the duration of each vowel.

Second, F1 and F2 frequency was measured at three timepoints, 25%, 50%, and 75% of vowel duration, which we refer to as early, mid, and late. Formant values were measured based

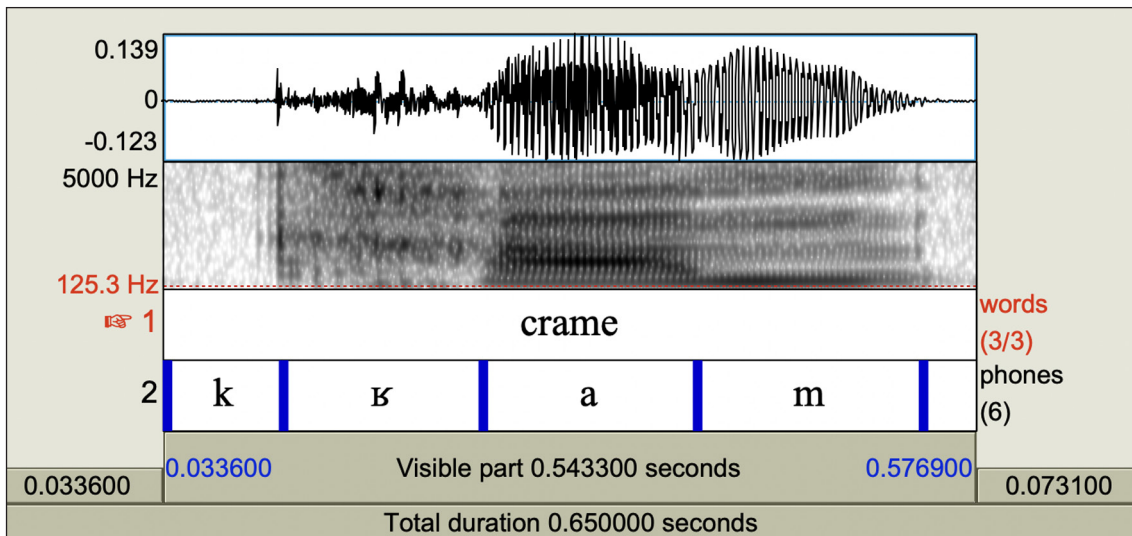


Figure 1: Segmented waveform and spectrogram of a sample word, *crame* [kram], ‘burn-3s’.

on default Burg formant tracking analyses (5 formants in 5500 Hz for female speakers and 5 formants in 5000 Hz for male speakers; 25 ms window) and verified by visual examination of wide band spectrograms.

Finally, acoustic vowel nasalization was measured at the same three timepoints using A1-P0, a measure derived from spectral characteristics of vowel nasalization (Chen, 1997). Nasalized vowels show the presence of an extra-low frequency spectral peak (P0), generally below the first formant, accompanied by a concomitant reduction in the amplitude of the first formant spectral peak (A1). The acoustic manifestation of vowel nasalization may be quantified, then, by examining the relative amplitudes of the nasal peak (in non-high vowels, where F1 and P0 would be separate) and the first formant in a measure A1-P0. As nasalization increases, P0 increases and A1 decreases. Therefore, a smaller A1-P0 value indicates greater acoustic vowel nasality. Although there are other methods for quantifying acoustic vowel nasalization (see Styler, 2017 for discussion), we used A1-P0 in order to facilitate comparison with prior work on nasalized vowels in French (e.g., Scarborough et al., 2018; Zellou & Chitoran, 2023).

2.2. Results

2.2.1. Aggregate results

A1-P0 values were analyzed with a mixed-effects linear regression model using the *lmer()* function in the *lme4* package (Bates et al., 2015) in the R program for statistical computing (R core team). Estimates for degrees of freedom, t-statistics, and p-values were computed using Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017). The model

was run on centered and scaled A1-P0 values and included fixed effects of Vowel Type (three levels: oral [reference level], coarticulated, nasal), and Timepoint (continuous, coded as 1, 2, 3; centered), and the two-way interaction between the two effects. The random effects structure of the model included by-speaker and by-word random intercepts and by-speaker random slopes for Vowel Type did not lead to convergence. All categorical variables were treatment coded. (lmer syntax: A1-P0~Vowel Type*Timepoint + (1|Speaker) + (1|Word)).

Figure 2 provides the average A1-P0 values over normalized time. As expected, nasal vowels contain greater acoustic vowel nasalization (lower A1-P0 values) than vowels overall [*coef.* = -1.4 , $t = -24.8$, $p < 0.001$]. Coarticulated vowels also contained greater acoustic nasality (lower A1-P0) than oral vowels [*coef.* = -0.3 , $t = -4.9$, $p < 0.001$]. There was also an interaction between Vowel Type and Timepoint, such that nasalization increases in nasal vowels over time [*coef.* = -0.2 , $t = -7.9$, $p < 0.001$], and also increases in coarticulated vowels over time [*coef.* = -0.1 , $t = -2.5$, $p < 0.05$]. No other effects were significant.

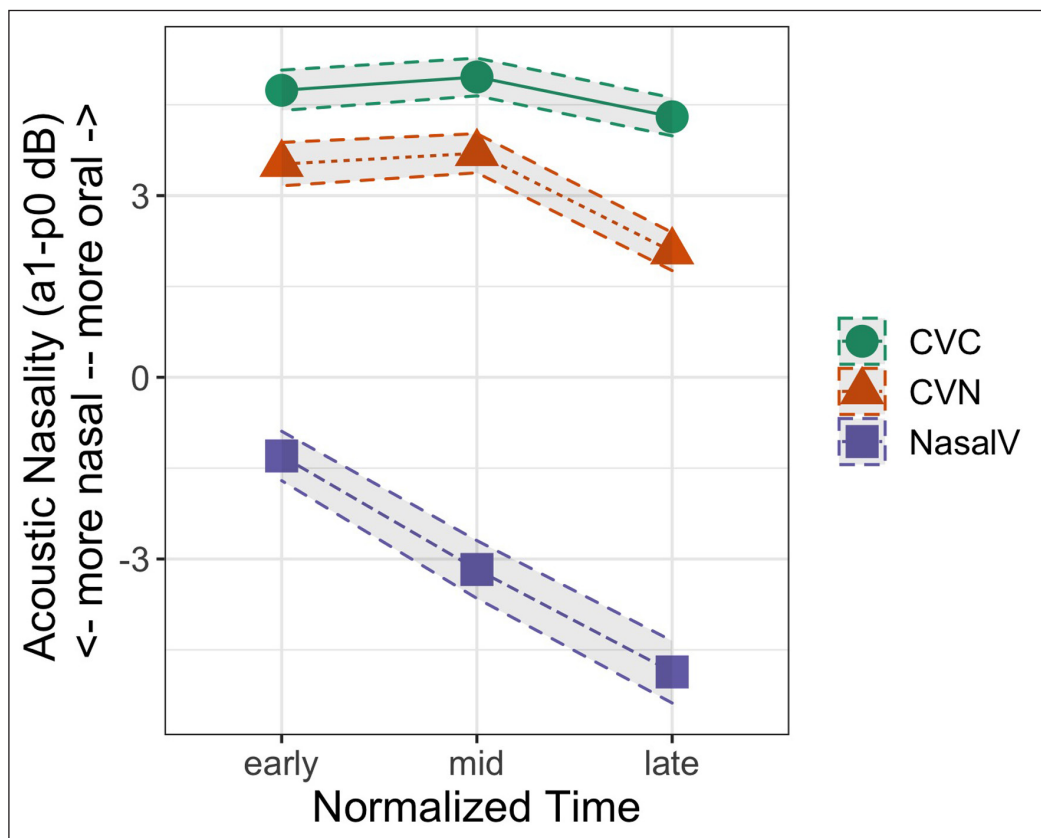


Figure 2: Mean acoustic vowel nasalization (A1-P0 dB) values across three equidistant timepoints for oral vowels (from CVC words), coarticulated vowels (from CVN words), and nasal vowels (from CṼ words).

To assess the relationship between nasality and vowel quality, two separate lmer models were run on log F2 and F1 values taken at the midpoint (centered and scaled). Each model contained fixed effects of Vowel Type (three levels: oral [reference level], coarticulated, nasal), and Phonemic Vowel (three levels: /a/ [reference level], /ɛ/, /ɔ/), and the two-way interaction between the two effects. The random effects structure of both models included by-speaker and by-word random intercepts and by-speaker random slopes for Vowel Type and Vowel. All categorical variables were treatment coded. (lmer syntax: $\text{Formant} \sim \text{Vowel Type} * \text{Vowel} + (1 + \text{Vowel Type} + \text{Vowel} | \text{Speaker}) + (1 | \text{Word})$).

Figure 3 displays mean log F1 and F2 values. In the model for F2, entirely as expected, there was an effect of Vowel: overall, /ɛ/ has a higher F2 than /a/ [$\text{coef.} = 0.9, t = -2.3, p < 0.05$], consistent with a fronter tongue position. While there was not a main effect of /ɔ/ on F2, there was a marginal interaction between Vowel Type and Vowel involving /ɔ/ [$\text{coef.} = 1.3, t = -2.1, p = 0.05$], such that the difference between oral and nasal /ɔ/ is smaller than that between oral and nasal /a/.

In addition, the model for F2 revealed a main effect of Vowel Type: nasal vowels contain lower F2 values than oral vowels [$\text{coef.} = -1.3, t = -3.3, p < 0.01$], consistent with a more retracted tongue position. Meanwhile, oral and coarticulated vowels do not differ significantly [$p = 0.9$]. No other main effects or interactions were significant.

In the model for F1, again as expected, there was a main effect of Vowel such that the mid vowels have lower F1 values relative to the low vowel (for the front mid vowel: [$\text{coef.} = -1.2, t = -4.9, p < 0.001$]; for the back mid vowel: [$\text{coef.} = -1.0, t = -4.4, p < 0.001$]), consistent with a higher tongue position.

In addition, the F1 model also revealed that nasal vowels contain a lower F1 than oral vowels [$\text{coef.} = -0.7, t = -3.5, p < 0.01$], consistent with a higher tongue position. Yet, this was mediated by an interaction between Vowel Type and Vowel: /ɛ/ has a higher F1 value when it is nasal, compared to when it is oral [$\text{coef.} = 1.5, t = 4.4, p < 0.001$], consistent with a lower tongue position. There were no main effects or interactions involving the comparison between the oral and coarticulated vowels.

Finally, with respect to vowel duration, vowels were longest in C \tilde{V} words (139 milliseconds), and shorter in CVC (110 ms) and CVN (113 ms) words. We ran a mixed effects linear regression model on vowel duration (logged, centered, and scaled) with a fixed effect of Vowel Type (three levels: oral [reference level], coarticulated, nasal), including by-speaker and by-item random intercepts and by-speaker random slopes for Vowel Type (lmer syntax: $\text{Vowel Duration} \sim \text{Vowel Type} + (1 + \text{Vowel Type} | \text{Speaker}) + (1 | \text{Word})$). The model confirmed that nasal vowels were longer than oral vowels [$\text{coef.} = 0.92, t = 3.9, p < 0.001$]. There was no difference in duration between oral and coarticulated vowels [$\text{coef.} = 0.22, t = 0.5, p = 0.6$].

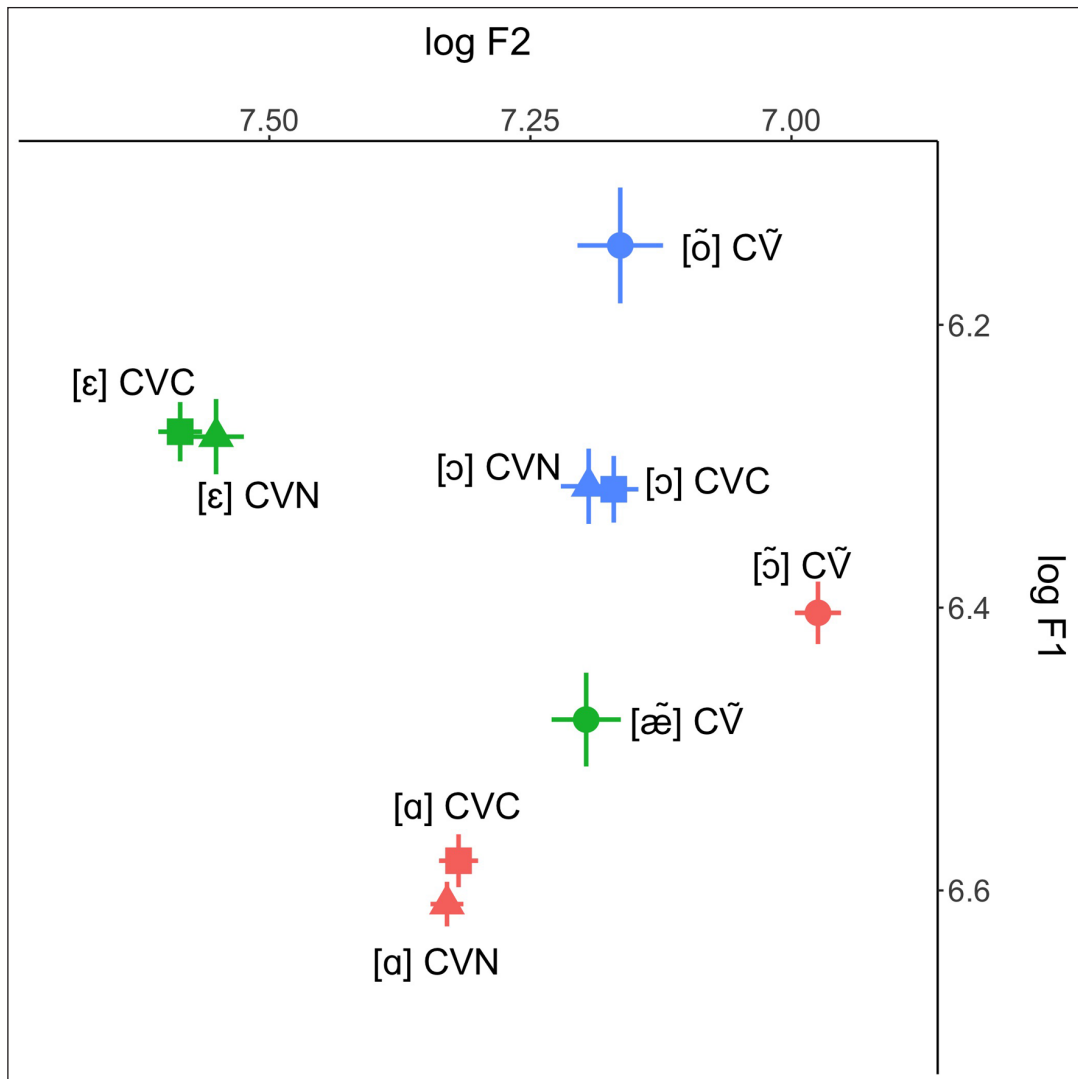


Figure 3: Mean and standard errors of log F1 and F2 values (at midpoint) for oral vowels (from CVC words, squares), coarticulated vowels (from CVN words, triangles), and nasal vowels (from CṼ words, circles). Colors reflect oral and nasal counterparts: green = [ε] and [æ̃], blue = [ɔ] and [õ], red = [ɑ] and [ã].

2.2.2. By-speaker results

Figure 4 provides the mean A1-P0 values across the thirty speakers. Significant variation is present. Some speakers produce comparable amounts of nasality in CVC versus CVN contexts (i.e., circles and triangles nearly overlap, such as for S009), suggesting that they coarticulate only to a very limited degree, if at all. Meanwhile, other speakers produce similar amounts of nasality in CṼ versus CVN contexts, suggesting that they coarticulate to a greater degree (i.e., triangle and squares nearly overlap, such as for S025).

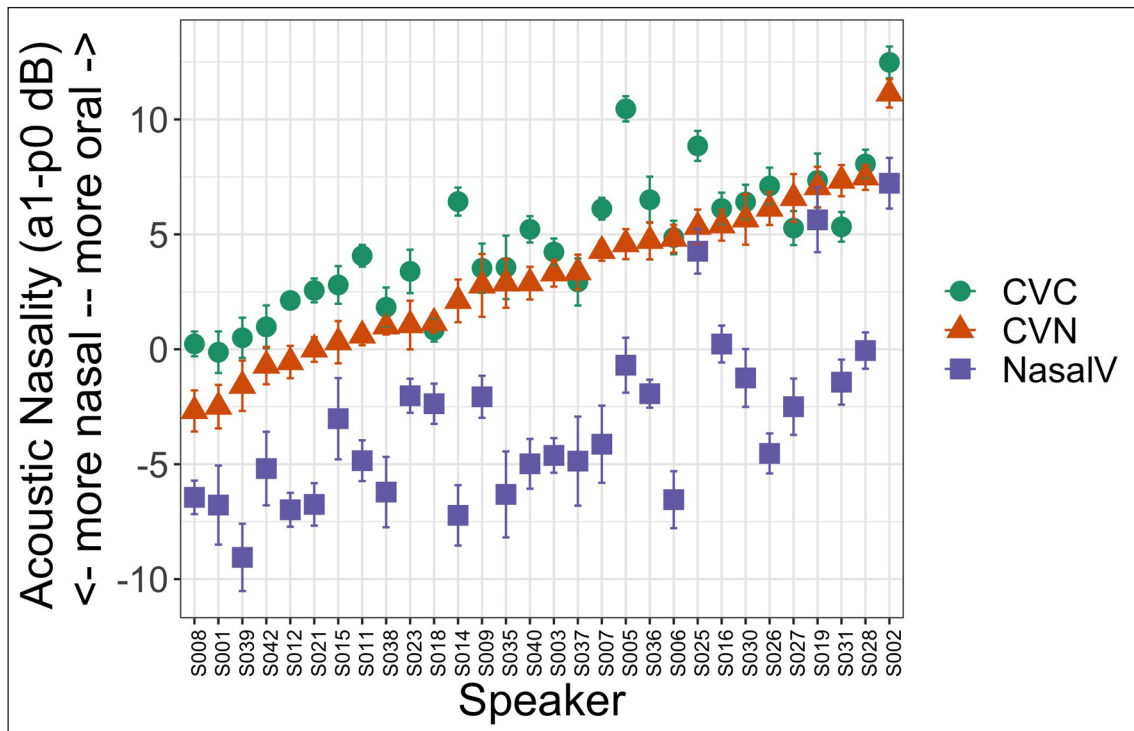


Figure 4: Mean acoustic vowel nasalization (A1-P0 dB) values (averaged across three vowel timepoints) for each speaker for oral vowels (from CVC words), coarticulated vowels (from CVN words), and nasal vowels (from $\tilde{C}\tilde{V}$ words).

2.3 Interim summary

Our production results replicate previous findings from the literature. As expected, acoustic nasality is greatest in $\tilde{C}\tilde{V}$ contexts, and relatively diminished in other contexts. While this pattern is broadly consistent with an account in which speakers deliberately limit nasal coarticulation (cf. Manuel, 1990), our findings also show that CVN contexts still exhibit significantly more nasality than CVC contexts, suggesting that some degree of coarticulation does occur. Also as expected, $\tilde{C}\tilde{V}$ contexts differ significantly in F2 and F1, compared to CVC and CVN contexts. This finding is consistent with previous accounts suggesting that French oral and nasal vowels exhibit differences not just in nasality, but also in quality. We also observe individual variation in coarticulatory patterns.

3. Perception study

The perception study consisted of a word identification task of the CV portion of each word produced by the thirty speakers in the production study. Using a paradigm similar to the classic study designed by Ali et al. (1971), we created truncated syllables spliced from $\tilde{C}\tilde{V}$, CVC, and CVN items. The CV syllables were then gated into noise, and presented to native French listeners who

performed a three-option forced-choice lexical categorization (either $C\tilde{V}$, CVC or CVN). Thus, the aim of the perception experiment is to determine how listeners map acoustic information from CV sequences onto lexical items.

3.1. Methods

3.1.1. Stimuli

Stimuli consisted of CV (in some cases, CCV) syllables truncated from the word productions by the 30 speakers from the production study. The syllables were then normalized to 60 dB (using the “Scale intensity...” function in Praat) and gated into wide-band noise, at a level 5 dB less than the peak intensity of the vowel. This was done in order to avoid a stop-bias that might occur if the syllables were to abruptly end in silence (cf. Ohala & Ohala, 1995). **Figure 5** displays the waveform and spectrogram of one of the stimulus items.

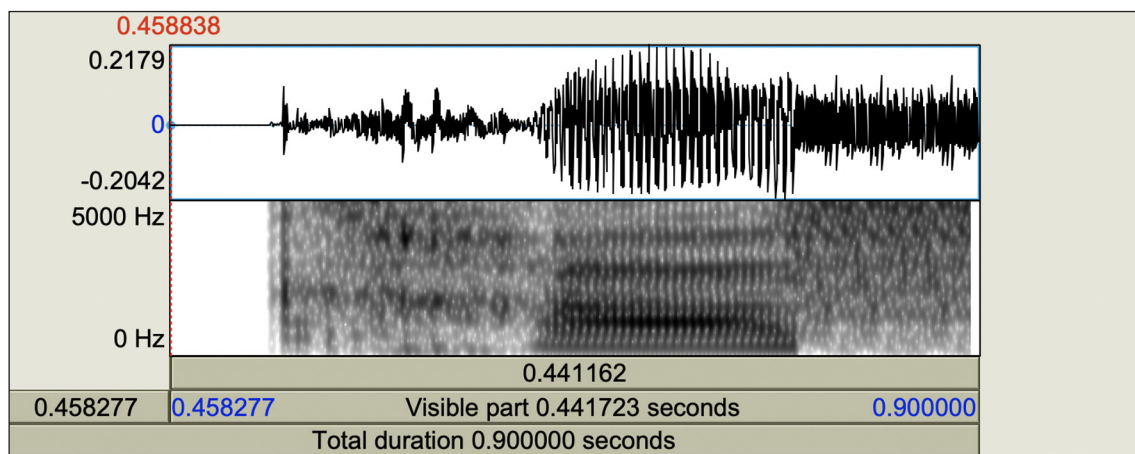


Figure 5: Waveform and spectrogram of a stimulus for the target word *crame* [kram] ‘burn-3s’ with the coda spliced away and gated into noise.

3.1.2. Participants and procedure

Fifty native French-speaking participants (16 female, 1 non-binary, 33 male; mean age = 32 years old) were recruited online via Prolific to complete a word identification task. Individuals could participate if they reported that: a) their first and primary language was French, and b) their nationality was French.

The experiment was conducted online using Qualtrics. Participants were instructed to take the test on their personal computers and to wear headphones. The experiment began with a sound calibration procedure: participants heard one sentence presented auditorily (*Je pense que j’ai compris, est-ce que j’écris le mot hyène dans le cercle vert?* “I think I understood, do I write the word hyena in the green circle?”), presented in silence at 60 dB, and were asked to identify the

final word of the sentence from four multiple choice options (*bleu, vert, jaune, blanc* “blue, green, yellow, white”). Afterwards, they were instructed to not adjust their sound levels again during the experiment.

Then, participants completed the word identification task. On a given trial, listeners heard one of the speakers produce either a $C\tilde{V}$ syllable, or a CV sequence truncated from a CVC or CVN word, gated into noise. Then, listeners selected one of three minimal pair choices, corresponding to the minimal triplet option for that syllable. For example, after hearing a stimulus [se] from the context [sen], listeners indicated whether the intended word was *cède*, *scènes*, or *saint*.

Five experimental lists were created, and six speakers from the production study were randomly assigned to each list. No speaker was assigned to more than one list. Each list contained all of the target items produced by each speaker, and therefore contained a target number of 126 words (21 stimulus words x 6 speakers). A few lists contained 124 or 125 items, depending upon whether the speaker had mispronounced an item or two (mispronounced items were excluded). Each participant was randomly assigned to one of the five lists, and heard each list item once. List items were presented in random order for each participant. Progression through trials was self-paced, there were no breaks or intervals between stimuli. The experiment took an average of 15 minutes for participants to complete.

The study was approved by the UC Davis Institutional Review Board (IRB) and subjects completed informed consent before participating.

3.2. Results

3.2.1. Aggregate performance

A listener response was considered accurate (1) if it corresponded to the original lexical item produced by the speaker, otherwise it was considered inaccurate (0). A mixed effects logistic regression model was run on the accuracy data. The model included a fixed effect of Vowel Type (oral [reference level], nasal, coarticulated) and by-speaker and by-listener random intercepts, as well as by-speaker and by-listener random slopes for vowel type. (glmer syntax: $Acc \sim Vowel\ Type + (1 + Vowel\ Type|Listener) + (1 + Vowel\ Type|Speaker)$).

Overall, mean response patterns for each vowel type were above chance (three options, chance-level = 33%). Yet, there were systematic differences across vowel types. Listeners were most accurate at identifying lexical items when the vowel was phonemically nasal (98% accuracy) [*coef.* = 4.5, *z* = 7.6, *p* < 0.001]. $C\tilde{V}$ items were likely not confused with either CVC or CVN; This indicates that there is little competition from CVN and CVC lexical competitors. In contrast, performance for lexical identifications from the oral (75%) and coarticulated (51%) vowels (from CVC and CVN contexts, respectively) was lower, with lowest performance for coarticulated vowels [*coef.* = -1.1, *z* = -6.0, *p* < 0.001].

Table 2 provides a confusion matrix for the different vowel types. Notably, coarticulated vowels are rarely misidentified as phonemically nasal vowels. Rather, CV stimuli from CVN contexts are either correctly identified, or identified as CVC lexical items.

		<i>Response</i>		
		CVC	CVN	C \tilde{V}
<i>Stimulus Type</i>	CV(C)	0.75	0.24	0.01
	CV(N)	0.47	0.51	0.02
	C \tilde{V}	0.01	0.01	0.98

Table 2: Confusion matrix with proportion responses to the word identification task.

3.2.2. Relating acoustic features of vowels and perception

In Section 2.2, we noted that speakers exhibited variability in the extent to which they coarticulated in CVN contexts. To explore the influence of this variability on perception, we ran a mixed-effects logistic regression model on responses to CVN contexts only. The model included fixed effects of A1-P0 (averaged over 3 timepoints of each vowel), log F1 and log F2 (taken at midpoint), and log vowel duration. All fixed effects were centered and scaled. The model included by-speaker and by-listener random intercepts, as well as by-listener random slopes for A1-P0, F1, and F2. (glmer syntax: $\text{Acc} \sim \text{A1-P0.std} + \text{F1.std} + \text{F2.std} + \log \text{Vowel duration} + (1 + \text{A1-P0.std} + \text{F1.std} + \text{F2.std} | \text{Listener}) + (1 | \text{Speaker})$).

The model computed a significant effect of acoustic nasality on accuracy [*coef.* = -0.4 , $z = -4.3$, $p < 0.001$]. The estimate for this effect is negative, indicating that listeners are more likely to correctly select a CVN lexical item when the vowel contains a lower A1-P0 value (i.e., greater acoustic vowel nasalization). Additionally, there was an effect of F1, such that vowels with smaller F1 values (i.e., higher tongue position) are more likely to be correctly identified as originating from CVN words [*coef.* = -0.6 , $z = -6.5$, $p < 0.001$]. There was also an effect of F2: vowels with smaller F2 values (indicating more retracted tongue position) are more likely to be correctly identified [*coef.* = -0.6 , $z = -7.0$, $p < 0.001$]. There was not an effect of vowel duration ($p = 0.2$).

3.2.3. Individual differences across speakers

Figure 6 provides mean accuracy rates for the thirty individual speakers who provided stimuli, ordered by mean identification rates for CVN contexts.

For phonemically nasal vowels from C \tilde{V} contexts, accuracy is very high and does not vary much across speakers, similar to what we see in the aggregate results. So, essentially all speakers produce nasal vowels that are unambiguous, in the sense that they are not confusable with either oral or coarticulated vowels.

For the oral vowels from CVC contexts, performance is overall lower, again similar to what was seen in the aggregate pattern. However, the range of values across speakers, from 60% to 90% accuracy, is larger than for the nasal vowels, suggesting that there is more variation in how speakers provide cues to orality.

Finally, for the coarticulated vowels from CVN contexts, we see the largest range of variation across speakers. In the lower accuracy ranges depicted on the left side of **Figure 6**, stimuli from CVN contexts are correctly identified as CVN lexical items around 35–40% of the time. Technically these results are at chance, since listeners choose from three options. However, the confusion matrix in **Table 2** makes clear that the nasal vowel is not really in competition here. Instead, for stimuli from these speakers, there is greater competition from the CVC category. Moving toward the middle of **Figure 6**, there are also many speakers whose stimuli from CVN contexts are identified in the 50% accuracy range. For these speakers, CVN vowels are truly ambiguous between CVN and CVC. Finally, toward the right side of **Figure 6**, we see several speakers whose stimuli are accurately identified as originating from CVN words over 60% of the time.

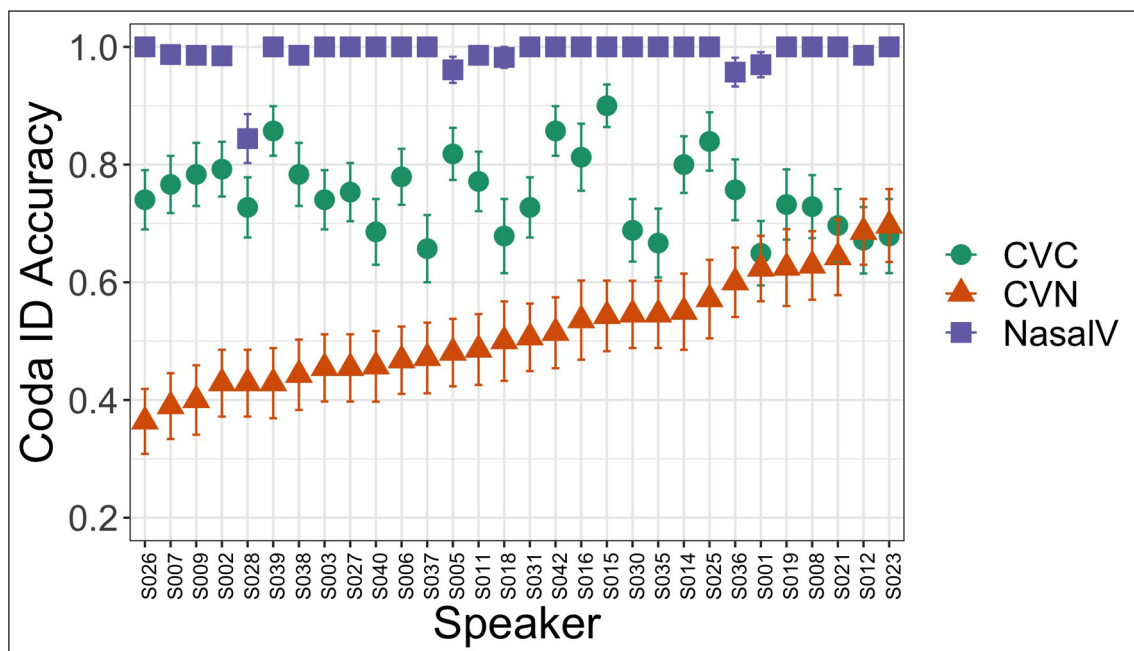


Figure 6: Proportion correct identification by stimulus type for each individual speaker.

To assess the relationship between acoustic properties of nasal coarticulation and the talker-specific identification patterns, we ran a linear regression model on mean accuracy from CVN contexts only. This revealed a significant effect, depicted in **Figure 7** [*coef.* = -0.01 , $t = -3.8$, $p < 0.001$].

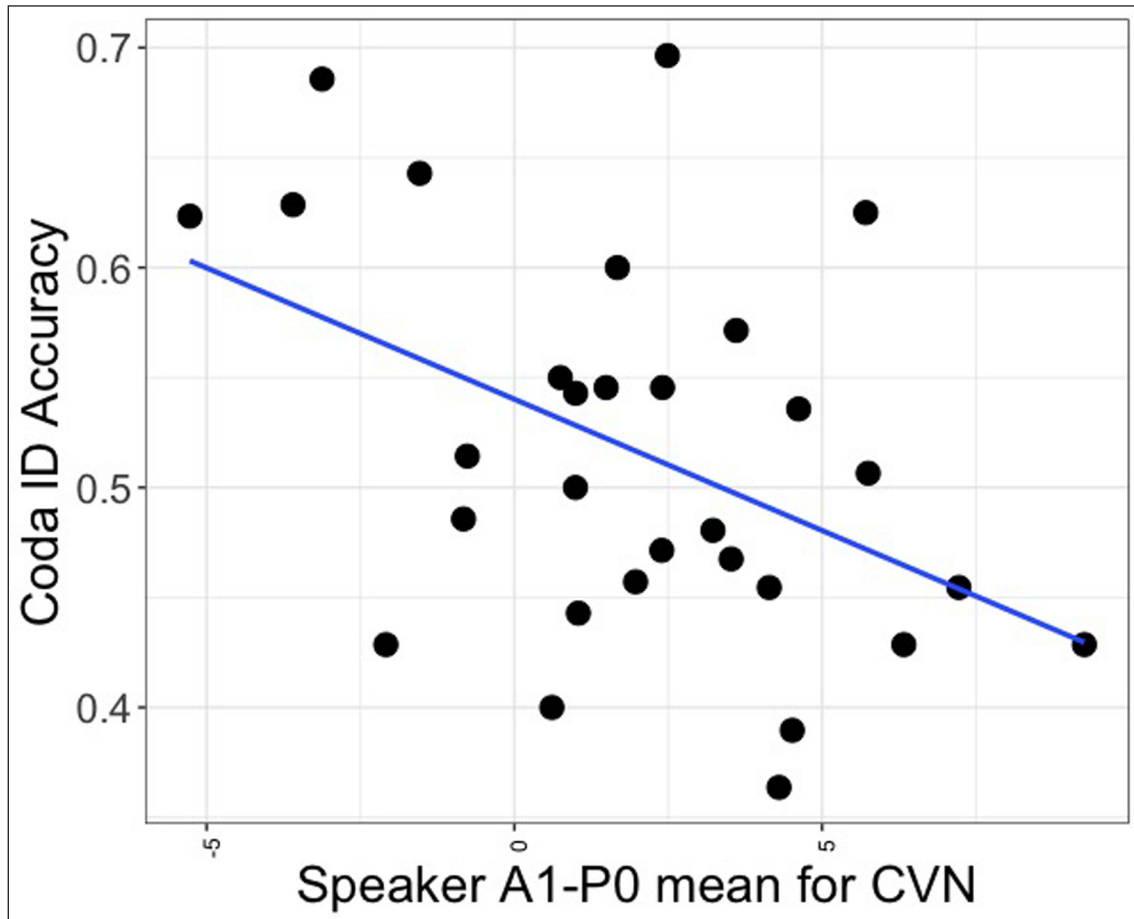


Figure 7: Correct identifications of coarticulated vowels by speakers' A1-P0 mean.

Finally, to assess the role of vowel quality in perception, we calculated the difference between mean log F2 values for coarticulated versus nasal vowels, for each individual speaker. The difference between mean log F1 values for coarticulated versus nasal vowels for each speaker was also calculated. We ran two separate linear regressions, with speaker mean accuracy for CVN items as the dependent variation and the speaker-specific formant difference as the predictor variable. The F2 model did not compute a significant effect [*coef.* = 0.04 , $t = 0.3$, $p = 0.8$]. There was also not a significant effect in the F1 model [*coef.* = 0.05 , $t = 0.5$, $p = 0.6$].

4. Discussion

The production and perception studies reported in this paper leveraged the natural variability of individual speakers to investigate how listeners respond to coarticulatory nasality on French vowels. A key question was whether, when nasality was present, listeners would interpret the signal as a VN sequence. The production study examined thirty Metropolitan French speakers' productions of CVC ~ CVN ~ C \tilde{V} triplets, and replicated previous findings in two important respects. First, although vowel nasality was present in both CVN and C \tilde{V} contexts, its magnitude was significantly diminished in CVN (as in Delvaux et al., 2008; Dow, 2020). Second, although F1 and F2 values for corresponding vowels were comparable across CVC and CVN contexts, they shifted notably in C \tilde{V} contexts (as in Carignan, 2014; Delvaux, 2009). Thus, CVN and C \tilde{V} contexts were distinguished not just by the degree of nasality on the vowel, but also by the quality of the vowel itself.

The recordings from the production study were used as stimuli in the perception study, in which fifty listeners heard the CV portion of a stimulus, and identified the originally-intended word. Perception results conformed to our predictions. Namely, while there was strong evidence that listeners confused CVN and CVC contexts, there was no evidence that they confused CVN and C \tilde{V} contexts. This is a scenario which creates the potential for listeners to make use of nasality cues, when they do occur, to help distinguish between CVN and CVC. Importantly, a correlation analysis suggests that listeners did just that: the greater the magnitude of nasality on a CVN stimulus, the more likely listeners were to correctly identify it as a CVN word. Thus, even though the phonological grammar of French would seem to prohibit listeners from using coarticulatory cues to interpret the signal as CVN, listeners can do it anyway.

4.1. Interplay between coarticulation and the grammar

We interpret our results as a first step in resolving an important conundrum in the interplay between physical events and abstract categories: namely, while coarticulatory cues are a rich source of information for speech perception, their effects can be modulated by the phonological grammar. As noted in the Introduction, the study by Beddor and colleagues (2013) provides a representative example of how coarticulatory cues help listeners interpret speech. In an experiment with American English stimuli in a visual world paradigm, they showed that listeners fixated more quickly on a visual image corresponding to a target such as spoken *bent* when coarticulatory nasalization occurred early in the vowel, rather than late. These findings are in concert with a long line of previous studies (starting e.g., with Malécot, 1960), and demonstrate that listeners respond to coarticulatory cues as soon as they become available.

Moreover, other studies have suggested that such conclusions must necessarily be language-specific. For example, in an experiment with CVC and CVN stimuli in a gating paradigm, Lahiri

and Marslen-Wilson (1991; see also Kotzor et al., 2022) showed that while English listeners interpreted nasality on the vowel as evidence for the presence of a nasal coda, Bengali listeners interpreted such cues as evidence for a phonemically nasal vowel. On this basis, Lahiri and Marslen-Wilson (1991) argued that when a language does not contain phonemically nasal vowels, as in English, listeners interpret nasality as evidence for coda N; but when a language does contain phonemically nasal vowels, as in Bengali, listeners interpret nasality as evidence for \tilde{V} .

Our findings introduce an important refinement to this scenario. Like Bengali, French contains phonemically nasal vowels. However, we have shown that French listeners *can* interpret nasality as evidence for VN. As we speculated earlier, the differences between French versus Bengali listeners most likely stem from an important fact about French, namely that oral and nasal vowels are distinguished not just by nasalization, but also by differences in quality. The same is not true of Bengali, where nasal vowels are equivalent in quality to their oral counterparts (e.g., Ferguson & Chowdhury, 1960). Thus, in mapping the speech signal onto a particular vowel phoneme, Bengali listeners must rely solely on nasality cues, whereas French listeners interpret nasality and quality in tandem (Delvaux, 2009). More broadly, both in the current work and in previous works (e.g., Beddor, Krakow, & Goldstein, 1986; Benguerel & Lafargue, 1981; Carignan, 2014; Delvaux et al., 2008; Delvaux, 2009; Dow, 2020; Krakow et al., 1988; Solé, 1992, 2007), French provides us with an opportunity to reconsider the characteristics of nasality as a phonological feature, and to highlight its multifaceted nature.

Recall that our analysis of perception results for CVN contexts did not show a significant correlation between F2 differences across CVN versus $C\tilde{V}$ contexts, on the one hand, and listener accuracy, on the other. In other words, greater differences between oral versus nasal vowel qualities did not necessarily lead to greater accuracy in distinguishing CVN and $C\tilde{V}$ stimuli. However, this correlation is not strictly necessary in order for our basic logic to hold: as long as the baseline quality difference between oral versus nasal vowels remains steady – and our production results show that this is indeed the case – listeners should be able to interpret nasal cues against the backdrop of quality cues. Future work could probe this issue further by systematically exploring how variation in vowel quality affects the perception of coarticulation in VN contexts, similar to the manner in which Delvaux (2009) approached $C\tilde{V}$ contexts.

A particularly interesting issue, which we were not able to address in the current study, concerns nasal coarticulation on high vowels (e.g., in contexts such as *fine* [fin] ‘thin, fine’ and *lune* [lyn] ‘moon’). Previous authors have reported that coarticulation is greater in these contexts, compared to non-high vowel contexts (Delvaux et al., 2008; Dow, 2020), possibly due to the fact that the inventory of French lacks phonemically nasal high vowels (e.g., *[fĩ], *[lɥ̃]). On this basis, we could hypothesize that, when listening to vowels that contain acoustic nasality, French listeners should make stronger predictions about the presence of N when the vowel is high, compared to when it is non-high.

4.2. Individual differences

In studying the perception of coarticulation, previous researchers have typically created stimuli using cross-splicing and/or acoustic manipulation. While useful, these techniques probe only a narrow range of human behavior. For example, if we use cross-splicing to create matched versus mismatched stimuli (i.e., CV_{NASAL}N versus CV_{ORAL}N), then we probe only how listeners respond to those two types of stimuli. By contrast, a distinctive aspect of the current study is that we presented listeners with stimuli “from the wild” – that is, with stimuli whose patterns of nasal coarticulation varied naturally among thirty different speakers of Metropolitan French. This allowed us to probe how listeners responded to a wide range of stimuli, which more closely resembles the challenges posed by everyday listening. Crucially, this approach also allowed us to show that certain speakers of French have stronger nasality cues in CVN contexts than others, and that listeners make use of these cues *when they are present*.

This approach closely resembles the methodology employed by Zellou (2022) for American English. In that study, sixty different speakers produced CVC and CVN words (e.g., *bed* and *ben*). Their recordings were used in a perception experiment, in which listeners heard CV stimuli with the coda removed, and identified the corresponding word (*bed* or *ben*?). Results showed that some speakers’ CVN tokens were (incorrectly) identified as CVC words, while other speakers’ CVN tokens were (correctly) identified as CVN words. American English does not have phonemically nasal vowels, and anticipatory coarticulation in CVN sequences is generally long and strong (e.g., Cohn, 1990). Nevertheless, there is substantial variability from one speaker to the next (Beddor et al., 2018; Styler, 2017; Zellou, 2017; Zellou & Brotherton, 2021). Zellou (2022) showed that this variability exerts direct consequences in listeners’ perceptual choices. That is, American English listeners make use of nasality cues *when they are present*. (We note that Beddor et al. (2013) and Coetzee et al. (2022) reached similar conclusions, using stimuli that varied categorically between early versus late onset of nasalization.)

Our current pattern of results for Metropolitan French is essentially identical to what Zellou (2022) found for American English: that is, some speakers’ CVN tokens were (incorrectly) identified as CVC words, while other speakers’ CVN tokens were (correctly) identified as CVN words. Unlike American English, however, French does have phonemically nasal vowels, and anticipatory coarticulation in CVN sequences is generally short and weak (Delvaux et al., 2008; Dow, 2020). Despite this rather profound difference, the basic behavior of French listeners is nevertheless equivalent to that of American English listeners. This is a novel finding which suggests the broad hypothesis that, regardless of phonological grammars, listeners always use coarticulatory cues, *when they are present*, to make perceptual choices. To test this hypothesis in subsequent studies with different languages and participant populations, we emphasize that it will be crucial to present a wide range of naturally-varying stimuli, as we have done here, in order to produce the broadest and most accurate characterization of listener behavior.

4.3. Theoretical implications

Our findings contribute new evidence to certain debates within the speech perception literature. In particular, different frameworks have made differing predictions about how listeners should interpret nasality when it occurs in the signal, particularly for those languages which contain both phonetically and phonemically nasalized vowels, such as French and Bengali. Under a surface-based account, listener behavior simply follows the distributional cues: if nasalization on the vowel is typically followed by a nasal consonant, then listeners will interpret nasalization as a cue to N (for discussion, see Chong & Garellek, 2018). By contrast, under a representational account as laid out in the Featurally-Underspecified Lexicon (FUL) model, listeners will interpret nasalization as a cue to \tilde{V} (Kotzor et al., 2022; Lahiri & Marslen-Wilson, 1991). This is because the nasalization in the signal is a “match” for $C\tilde{V}$ contexts, which underlyingly contain the feature [Nasal]. Meanwhile, nasalization in the signal is a “no-mismatch” for both CVN and CVC contexts, which are underlyingly unspecified and do not contain the feature [Nasal].

Our perceptual results are not entirely consistent with the representational account, at least as it currently stands. Listeners did interpret $C\tilde{V}$ stimuli as $C\tilde{V}$ words, as predicted by FUL. However, listeners did *not* interpret CVN stimuli as $C\tilde{V}$ words, a finding that would seem to contradict the predictions of FUL. Instead, listeners overwhelmingly interpreted CVN stimuli as either CVC or CVN words. That is, despite the presence of nasality in the signal, they did not map this to a lexical vowel with an underlying [Nasal] feature. To some extent, of course, this result may be due to the fact that the French CVN words contained relatively small amounts of nasalization as well as distinct quality differences across CVN and $C\tilde{V}$ vowels, as demonstrated by our production study. Nevertheless, even when the magnitude of nasalization did increase due to natural variation among speaker productions, listeners did not become more likely to select $C\tilde{V}$ words. Instead, as our correlation analysis showed, they became more likely to select CVN words. Note, meanwhile, that although our listeners did not map acoustic nasality onto a lexical [Nasal] vowel, they did correctly use nasality as evidence for the presence of a lexical [Nasal] consonant. This suggests that a refined FUL model could potentially make correct predictions for languages like French, particularly if it were to incorporate temporal information, as well as the presence of quality differences between oral and nasal vowels. For instance, if representations for [Nasal] vowels in French included both vowel nasality and vowel quality features, this could explain the lack of confusability by listeners between CVN and $C\tilde{V}$ vowels in the current study.

Whether our perceptual results support a surface-based account remains an open question, because the nature of the distribution itself has yet not been fully established. On the one hand, summary trends show that very little nasalization occurs on vowels in CVN contexts, both in our own data and in previous studies (Delvaux et al., 2008; Dow, 2020). On the other hand, some speakers produce greater amounts of nasalization than others do, again in our own data and in previous studies (Styler, 2017). To really pin down the distribution, however, we must ask

whether this variation among speakers is principled. For instance, is it consistently affected by sociolinguistic factors such as speaker gender and social class, and/or by stylistic factors such as clear versus casual speech? If the answer to any of these questions is yes, then we would generate an interesting set of predictions. For example, if female speakers of French are generally more likely to exhibit nasal coarticulation than male speakers of French, then a distributional account predicts that listeners will be more likely to make use of coarticulatory cues when listening to a female voice, compared to a male one. Coetzee et al. (2022) tested this hypothesis in a somewhat different context — namely, predictable differences in the amount of nasal coarticulation between two socio-ethnic varieties of Afrikaans — but did not find evidence that listeners adjusted their perceptual strategies based on dialect information alone. The same question could be asked for vowel quality, as well – are there differences across regional varieties in vowel quality across categories that listeners are able to make use of? Our F1 and F2 models confirm that our listeners did not rely on vowel quality (at least not exclusively) to identify the words. But, could listeners still have used the nasality in the signal against the backdrop of vowel quality differences to identify a phonemic nasal vowel? If yes, it would also predict individual perception/production differences whereby some speaker-listeners would rely more heavily on nasal coarticulation, and others more heavily on vowel quality differences, but never on just one of these alone. These issues can be further explored in future work.

4.4. Avenues for future work

Future research on French could employ time-sensitive measurements, such as eye-tracking or gating, in order to probe listeners' perception of nasality at various time points within the vowel. To our knowledge, only a couple of previous studies have employed these techniques. Desmeules-Trudel and Zamuner (2019) used a visual world paradigm to examine how Canadian French listeners responded to changes in the temporal duration of nasality on vowels. Their results showed that as the nasality duration decreased, listeners were more likely to select a picture depicting a CVN lexical item, compared to a picture of a C \tilde{V} lexical item; notably, however, the eye fixation results were not comparable, since proportion of fixations to CVN versus C \tilde{V} pictures was not affected. A similar experiment using Metropolitan French, whose vowel system differs in crucial ways from Canadian French, could shed further light on our current results and also allow for a more direct comparison with the findings of Beddor et al. (2013).

Meanwhile, Ingram and Mylne (1994) examined French listeners' responses to CVC ~ CVN ~ C \tilde{V} C words, such as *tête* ~ *tante* ~ *tanne*, using a gating paradigm. Their descriptive results, which were summed over all gates, showed a pattern of results that are different from what we report here. Specifically, they indicated greatest accuracy for CVC, then C \tilde{V} C, with lowest accuracy for CVN, whereas our own results showed greatest accuracy for C \tilde{V} C, then CVC, and lowest accuracy for CVN. Ingram and Mylne (1994) do not report any information about their

acoustic stimuli (such as who recorded them, or what acoustic characteristics they possessed), and their French listeners resided in Australia, rather than in a French-speaking country. These factors make it difficult to speculate about why our results differ from theirs. Suffice to say that future experiments with gating paradigms could nevertheless shed further light on our current results, and allow for a more direct comparison with the findings of Lahiri and Marslen-Wilson (1991) and Ohala and Ohala (1995), as well as Kotzor et al. (2022).

The use of eye-tracking and/or gating could also help to address certain limitations of the current study. In our perceptual experiment, for example, listeners heard stimuli of the shape CV, since they were presented with recordings in which the original coda, if present, had been excised. This means that while the CVC and CVN stimuli underwent an experimental manipulation, the C \tilde{V} stimuli did not, a fact which may have rendered them easier to identify. Note, in addition, that the vowels [ɑ] and [ɔ] rarely occur in open syllables in French (Carignan, 2014), raising the possibility that listeners may have had difficulty interpreting truncated CV sequences such as [fa] and [trɔ] (from [fan] ‘wilt-1-3PERSON’ and [trɔŋ] ‘face-FAMILIAR’, respectively). Furthermore, in many languages including French, nasal vowels are typically longer than oral vowels (Delvaux, 2009), and this was also the pattern that we found in our own stimuli, raising the possibility that our participants may have distinguished between VN and \tilde{V} not only on the basis of vowel quality, as we have speculated, but also on the basis of overall duration.

Another direction for future work is to explore individual differences in the relationship between the production and perception of nasality cues across French speakers. There is some work starting to look at this question. Rodriquez et al. (2023) found that French speakers who use nasalization to a lesser degree to signal \tilde{V} vs. VN tend to have longer and especially more variable anticipatory nasalization in CVN tokens than those speakers who use nasalization stronger to signal nasal vowels. Thus, this presents a ripe direction for future work.

Finally, future work could also address the question of compensation for coarticulation. Although we have cited previous evidence that listeners use coarticulation as a cue to the presence of particular segment types, there is also ample evidence that listeners “undo” coarticulation in the presence of the segment that caused it (e.g., Mann & Repp, 1980). For example, while listeners are highly likely to perceive a phonemically nasal vowel in the context C \tilde{V} C, they are less likely to do so in the context C \tilde{V} N, because they attribute some or all of the vowel’s nasality to the following consonant (e.g., Kawasaki, 1986). Note that in the stimuli for the current study, we excised the coda, which “hid” the source of coarticulatory effects from listeners and rendered it infeasible to assess any potential effects of compensation. Different stimuli and alternative experimental tasks could allow us to assess these effects, an important goal given the evidence that compensation – like the use of coarticulatory cues – may also be modulated by the phonological grammar of a language (Beddor & Krakow, 1999).

5. Conclusion

On the face of it, coarticulation can seem to be a minor detail in the broader landscape of speech production and perception – just the result of anatomical movements that happen to overlap in time. And yet we have long known that the situation is far more complex than that; indeed, Liberman and Studdert-Kennedy called coarticulation “the essence of the speech code” (1978: 163). With this backdrop, the current study adds one more element to our understanding. French is a language that is, in theory, not supposed to use coarticulation as a cue to the presence of a nasal consonant. But we have shown that it does, and we have done so by using a large sample of naturally-occurring coarticulations from real speakers. Of course, it is still the case that the French language observes a basic production constraint: coarticulatory nasality is relatively weak, potentially due to the presence of phonemically nasal vowels. But it is those same vowels which also create a perceptual freedom: since their quality differs so much from oral vowels, listeners can adjust their interpretation of nasality accordingly. The upshot of our study, along with many others before it, is that the interplay of fine-grained acoustic cues with higher-level categories manifests itself in manifold ways, and we are only beginning to investigate the true depth and complexity of human knowledge about coarticulatory patterns.

Acknowledgements

Thanks to Editor Eva Reinisch and two anonymous reviewers for their helpful feedback on this paper. This work was supported by a Fulbright Research Scholar Grant to GZ and Labex EFL (ANR-10-LABX-0083-LabEx EFL) to Université Paris Cité. Thanks to Ziqi Zhou for help with segmentation and hand-correction of the production data.

Competing Interests

The authors have no competing interests to declare.

References

- Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. (1971). Perception of coarticulated nasality. *The Journal of the Acoustical Society of America*, 49(2B), 538–540. DOI: <https://doi.org/10.1121/1.1912384>
- Bates, D. M. (2015). lme4: mixed-effects modeling with R <https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>.
- Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, 94(4), 931–968. DOI: <https://doi.org/10.1353/lan.2018.0051>
- 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. DOI: <https://doi.org/10.1121/1.428111>
- Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology*, 3, 197–217. DOI: <https://doi.org/10.1017/S0952675700000646>
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4), 2350–2366. DOI: <https://doi.org/10.1121/1.4794366>
- Benguerel, A.-P., & Lafargue, A. (1981). Perception of vowel nasalization in French. *Journal of Phonetics*, 9(3), 309–321. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30974-X](https://doi.org/10.1016/S0095-4470(19)30974-X)
- 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, 23–33. DOI: <https://doi.org/10.1016/j.wocn.2014.05.001>
- Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4), 2360–2370. DOI: <https://doi.org/10.1121/1.419620>
- Chong, A. J., & Garellek, M. (2018). Online perception of glottalized coda stops in American English. *Laboratory Phonology*, 9(1), 4. DOI: <https://doi.org/10.5334/labphon.70>
- Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, 4(4), 337–351. DOI: [https://doi.org/10.1016/S0095-4470\(19\)31260-4](https://doi.org/10.1016/S0095-4470(19)31260-4)

- Cohn, A. C. (1990). *Phonetic and phonological rules of nasalization*. Ph.D. Dissertation. University of California, Los Angeles.
- Coetzee, A. W., Beddor, P. S., Styler, W., Tobin, S., Bekker, I., & Wissing, D. (2022). Producing and perceiving socially structured coarticulation: Coarticulatory nasalization in Afrikaans. *Laboratory Phonology*, 13(1). DOI: <https://doi.org/10.16995/labphon.6450>
- 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. DOI: <https://doi.org/10.1080/01690960143000074>
- Delvaux, V. (2009). Perception du contraste de nasalité vocalique en français. *Journal of French Language Studies*, 19(1), 25–59. DOI: <https://doi.org/10.1017/S0959269508003566>
- Delvaux, V., Demolin, D., Harmegnies, B., & Soquet, A. (2008). The aerodynamics of nasalization in French. *Journal of Phonetics*, 36(4), 578–606. DOI: <https://doi.org/10.1016/j.wocn.2008.02.002>
- Desmeules-Trudel, F., & Zamuner, T. S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics*, 81(5), 1654–1672. DOI: <https://doi.org/10.3758/s13414-019-01693-9>
- Dow, M. (2020). A phonetic-phonological study of vowel height and nasal coarticulation in French. *Journal of French Language Studies*, 30(3), 239–274. DOI: <https://doi.org/10.1017/S0959269520000083>
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16(4), 577–583. DOI: <https://doi.org/10.1162/089892904323057308>
- Ferguson, C. A., & Chowdhury, M. (1960). The phonemes of Bengali. *Language*, 36(1), 22–59. DOI: <https://doi.org/10.2307/410622>
- Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, 62, 21–32. DOI: <https://doi.org/10.3758/BF03212058>
- Ingram, J., & Mylne, T. (1994). Perceptual parsing of nasal vowels. In *ICSLP*. DOI: <https://doi.org/10.21437/ICSLP.1994-114>
- Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. J. J. Ohala & J. J. Jaeger (Eds.), In *Experimental Phonology* (pp. 81–103). (MIT Press, Cambridge, MA).
- Kotzor, S., Wetterlin, A., Roberts, A. C., Reetz, H., & Lahiri, A. (2022). Bengali nasal vowels: lexical representation and listener perception. *Phonetica*, 79(2), 115–150. DOI: <https://doi.org/10.1515/phon-2022-2017>
- Krakov, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *The Journal of the Acoustical Society of America*, 83(3), 1146–1158. DOI: <https://doi.org/10.1121/1.396059>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of sStatistical Software*, 82, 1–26. DOI: <https://doi.org/10.18637/jss.v082.i13>

- Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245–294. DOI: [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, 7, 637–675. DOI: <https://doi.org/10.1515/9783110197105>
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44–59. DOI: <https://doi.org/10.1016/j.wocn.2010.01.002>
- Lieberman, A. M. & Studdert-Kennedy, M. (1978). Phonetic perception. In: Perception (Edited by Held et al.) Handbook of sensory physiology, Vol. VIII. Heidelberg: Springer Verlag. DOI: https://doi.org/10.1007/978-3-642-46354-9_5
- Malécot, A. (1960). Nasal syllabics in American English. *Journal of Speech and Hearing Research*, 3(3), 268–274. DOI: <https://doi.org/10.1044/jshr.0303.268>
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, 28(3), 213–228. DOI: <https://doi.org/10.3758/BF03204377>
- Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, 88(3), 1286–1298. DOI: <https://doi.org/10.1121/1.399705>
- Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *The Journal of the Acoustical Society of America*, 69(2), 559–567. DOI: <https://doi.org/10.1121/1.385484>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using kald. In *Interspeech* (Vol. 2017, pp. 498–502). DOI: <https://doi.org/10.21437/Interspeech.2017-1386>
- Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. *Phonology and phonetic evidence. Papers in Laboratory Phonology IV*, 41–60. DOI: <https://doi.org/10.1017/CBO9780511554315.004>
- Rodriquez, F., Pouplier, M., Alderton, R., Lo, J. J. H., Evans, B. G., Reinisch, E., & Carignan, C. (2023). What French speakers' nasal vowels tell us about anticipatory nasal coarticulation. In: Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 848–852). Guarant International.
- 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. DOI: <https://doi.org/10.1016/j.jml.2013.11.002>
- Scarborough, R., Fougeron, C., & Marques, L. (2018). Neighborhood-conditioned coarticulation effects in French listener-directed speech. *The Journal of the Acoustical Society of America*, 144(3), 1900–1900. DOI: <https://doi.org/10.1121/1.5068323>
- Scarborough, R., & Zellou, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5), 3793–3807. DOI: <https://doi.org/10.1121/1.4824120>

- Scarborough, R., Zellou, G., Mirzayan, A., & Rood, D. S. (2015). Phonetic and phonological patterns of nasality in Lakota vowels. *Journal of the International Phonetic Association*, 45(3), 289–309. DOI: <https://doi.org/10.1017/S0025100315000171>
- Solé, M. J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech*, 35(1–2), 29–43. DOI: <https://doi.org/10.1177/002383099203500204>
- Solé, M. J. (2007). Controlled and mechanical properties in speech. *Experimental Approaches to Phonology* (pp. 302–321). DOI: <https://doi.org/10.1093/oso/9780199296675.003.0018>
- Stevens, K. N., Andredde, A., & Viana, M. C. (1987). Perception of vowel nasalization in VC contexts: A cross-language study. *The Journal of the Acoustical Society of America*, 82(S1), S119–S119. DOI: <https://doi.org/10.1121/1.2024621>
- Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America*, 142(4), 2469–2482. DOI: <https://doi.org/10.1121/1.5008854>
- Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, 35(1), 49–64. DOI: <https://doi.org/10.3758/BF03205924>
- Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61, 13–29. DOI: <https://doi.org/10.1016/j.wocn.2016.12.002>
- Zellou, G. (2022). *Coarticulation in Phonology*. Elements in Phonology Series. Cambridge University Press: Cambridge. DOI: <https://doi.org/10.1017/9781009082488>
- Zellou, G., & Brotherton, C. (2021). Phonetic imitation of multidimensional acoustic variation of the nasal split short-a system. *Speech Communication*, 135, 54–65. DOI: <https://doi.org/10.1016/j.specom.2021.10.005>
- Zellou, G., & Chitoran, I. (2023). Lexical competition influences coarticulatory variation in French: comparing competition from nasal and oral vowel minimal pairs. *Glossa: a journal of general linguistics*, 8(1). DOI: <https://doi.org/10.16995/glossa.9801>
- Zellou, G., Pycha, A., & Cohn, M. (2023). The perception of nasal coarticulatory variation in face-masked speech. *The Journal of the Acoustical Society of America*, 153(2), 1084–1093. DOI: <https://doi.org/10.1121/10.0017257>

