frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 11 December 2023
DOI 10.3389/fcomm.2023.1307547

OPEN ACCESS

EDITED BY
Antonio Benítez-Burraco,
University of Seville, Spain

REVIEWED BY
Wendy Elvira-García,
University of Barcelona, Spain
Catarina Oliveira,
University of Aveiro, Portugal

*CORRESPONDENCE
Georgia Zellou
✉ gzellou@ucdavis.edu

RECEIVED 04 October 2023
ACCEPTED 22 November 2023
PUBLISHED 11 December 2023

CITATION
Gwizdzinski J, Barreda S, Carignan C and Zellou G (2023) Perceptual identification of oral and nasalized vowels across American English and British English listeners and TTS voices. *Front. Commun.* 8:1307547. doi: 10.3389/fcomm.2023.1307547

Perceptual identification of oral and nasalized vowels across American English and British English listeners and TTS voices

Jakub Gwizdzinski[1], Santiago Barreda[2], Christopher Carignan[1] and Georgia Zellou[2]*

[1]Department of Speech Hearing and Phonetic Sciences, University College London, London, United Kingdom, [2]Linguistics Department, University of California, Davis, Davis, CA, United States

Nasal coarticulation is when the lowering of the velum for a nasal consonant co-occurs with the production of an adjacent vowel, causing the vowel to become (at least partially) nasalized. In the case of anticipatory nasal coarticulation, enhanced coarticulatory magnitude on the vowel facilitates the identification of an upcoming nasal coda consonant. However, nasalization also affects the acoustic properties of the vowel, including formant frequencies. Thus, while anticipatory nasalization may help facilitate perception of a nasal coda consonant, it may at the same time cause difficulty in the correct identification of preceding vowels. Prior work suggests that the temporal degree of nasal coarticulation is greater in American English (US) than British English (UK), yet the perceptual consequences of these differences have not been explored. The current study investigates perceptual confusions for oral and nasalized vowels in US and UK TTS voices by US and UK listeners. We use TTS voices, in particular, to explore these perceptual consequences during human-computer interaction, which is increasing due to the rise of speech-enabled devices. Listeners heard words with oral and nasal codas produced by US and UK voices, masked with noise, and made lexical identifications from a set of options varying in vowel and coda contrasts. We find the strongest effect of speaker dialect on accurate word selection: overall accuracy is highest for UK Oral Coda words (83%) and lower for US Oral Coda words (67%); the lowest accuracy was for words with Nasal Codas in both dialects (UK Nasal = 61%; US Nasal = 60%). Error patterns differed across dialects: both listener groups made more errors in identifying nasal codas in words produced in UK English than those produced in US English. Yet, the rate of errors in identifying the quality of nasalized vowels was similarly lower than that of oral vowels across both varieties. We discuss the implications of these results for cross-dialectal coarticulatory variation, human-computer interaction, and perceptually driven sound change.

KEYWORDS

nasal coarticulation, perception, human-computer interaction, Bayesian, English dialects

1 Introduction

1.1 Nasal coarticulation

Nasalization refers to the production of speech sounds with a lowered velum, which allows air to resonate through the nasal cavity. Of the languages represented in the World Atlas of Language Structures, nearly all (98%) possess phonemic nasal consonants (Maddieson, 2013), such as /m/ (as in *mouse*) or /n/ (as in *nose*) in English. In contrast, just

over a quarter of languages, maintain a phonemic contrast between oral vowels and nasal vowels (Hajek, 2013), e.g., the French word pair *beau* /bo/ "beautiful" and *bon* /bõ/ "good" (Styler, 2017). Other languages, such as English, do not use vowel nasality as a phonologically contrastive feature, and nasalized vowels are instead the result of nasal coarticulation, e.g., *bun* /bʌn/ [bʌ̃n]. Anticipatory coarticulation, specifically, occurs when the lowering of the velum for a nasal consonant starts during the production of the preceding vowel, causing the vowel to become nasalized as well.

Different languages show different amounts of nasal coarticulation: nasal coarticulation has been shown to be different in degree, extent, and direction in Greek (Diakoumakou, 2004), Thai (Onsuwan, 2005), Ikalanga (Beddor, 2007), Bininj Kunwok (Stoakes et al., 2020), and Arabana (Carignan et al., 2023). There is also evidence of cross-dialectal differences in nasal coarticulation within a language: for instance, Dominican Spanish anticipatory nasal coarticulation is extensive while Argentinian Spanish has less extensive nasalization (Bongiovanni, 2021). The degree of nasal coarticulation has been shown to depend on the specific variety of English, e.g., listeners tend to judge American English and Australian English as sounding more nasal than British English (Beddor, 1993; Hartley and Preston, 2002; Burridge and Kortmann, 2008). In Australia, strong vowel nasalization is associated particularly with so-called "broad" accents (Pittam, 1987), including the tensing of /æ/ before nasal consonants (Cox and Palethorpe, 2014), a characteristic that has also been observed widely in North American dialects, but absent in Newfoundland and British English (Mielke et al., 2017). Nasal coarticulation also varies with age and social group (Zellou and Tamminga, 2014). For instance, it has been shown that younger English speakers from Philadelphia use far less nasal coarticulation than older speakers, particularly older men (Tamminga and Zellou, 2015).

During the production of vowel nasalization, the acoustic resonances of both the oral cavity and the nasal cavity are merged, creating an acoustic signal that contains the resonant frequencies of both cavities (Styler, 2017) and results in spectral properties that are substantially different from those associated with the oral cavity alone (Carignan, 2018a,b). Using nasalance and ultrasound to separate the effects of both cavities on formant frequencies, Carignan (2018a) observed that the independent acoustic effects of vowel nasalization on vowels produced by speakers from different language backgrounds include the lowering of F1 for non-high vowels and the lowering of F2 for non-front vowels. Considering the perceptual relation between these formants and vowel quality, this result suggests that nasalized non-high vowels tend to sound higher whereas nasalized non-front vowels tend to sound backer than their respective oral counterparts; indeed, these patterns have been observed in perceptual studies (Wright, 1975, 1986; Beddor, 1993; Delvaux, 2009). Because of these acoustic modifications, nasalized vowels are also closer together in the acoustic space, which may further complicate perception of vowel quality due to reduced phonetic distinctiveness (Krakow et al., 1988; Beddor, 2009).

Although these effects of nasal coarticulation may lead to perceptual confusion of the vowel, nasal coarticulation on vowels has been shown to aid perception of the nasal consonant itself. Listeners are sensitive to nasal coarticulation as it helps them identify nasal consonants and better contrast minimal pairs of words, such as *bet* /bɛt/ [bɛt] and *bent* /bɛnt/ [bɛ̃t] (Zellou, 2017; Zellou and Dahan, 2019), and this perceptual benefit is used by listeners to identify codas as nasal as soon as the nasalization of the preceding vowel begins (Beddor et al., 2013). Although this sensitivity occurs in the pre-nasal vowel environment, listeners do not attribute the effects to the vowel itself; rather, listeners compensate for the presence of nasal coarticulatory effects by attributing the phonetic vowel nasality to its source, the nasal consonant itself (Ohala, 1993). This process of perceptual compensation for nasal coarticulation suggests that the vowel quality shifts outlined above may be ascribed by listeners to their source, in that the presence of these spectral modifications of the vowel acoustics may be correctly interpreted by listeners as arising from the coupling of the oral and nasal cavities.

However, listeners may not fully compensate for coarticulation (Beddor and Krakow, 1999). Then, the acoustic effects of vowel nasalization may result in perceptual reanalysis of the vowel quality. For instance, lowering of F1 in a vowel can come about not only due to nasal resonances (as observed in Carignan, 2018c), but also due to raising of the tongue body (Wright, 1975, 1986; Krakow et al., 1988). Therefore, listeners may potentially misidentify the source of the lowered F1 as due to a raised tongue instead of nasalization. Using English listener imitations of French speaker oral and nasal vowel productions, Carignan (2018c) observed that the English imitators successfully matched F1 resonances of the native French nasal vowel productions, although they did so by employing a lesser degree of nasalization and, instead, a raised tongue body, in comparison with the French speakers. This suggests that the English listeners at least partially misattributed the spectral effects of nasalization as due to inherent properties of the vowel itself, leading to a perceptual reanalysis of the vowel arising from perceptual ambiguity (De Decker and Nycz, 2012; see also Carignan, 2014; Zellou et al., 2020; Zellou and Brotherton, 2021).

1.2 Nasal coarticulation in English

Since the French stimuli used by Carignan (2018c) were phonologically contrastive nasal vowels, it is not clear whether the reanalysis exhibited by the English listeners in that study might also extend to phonetic vowel nasality, e.g., in their native English language. The current study seeks to address this issue by focusing on the interaction between vowel quality and coda nasality in the perception of words with oral and nasalized vowels in two different varieties of English.

Scarborough and Zellou (2012) analyzed the mean formant values of four American English vowels /æ, ɛ, ɑ, ʌ/, in both oral and nasal coda contexts, and also asked listeners to determine the identity of the vowels. In the nasal, compared to the oral, context, all four vowels were realized with a lowered F1, the front vowels were realized with a raised F2, and the back vowels were realized with a lowered F2; these findings are consistent with the language-independent effects of nasalization observed by Carignan (2018a,b). With regard to the perceptual results, they found that listeners were better at identifying oral vowels than nasalized vowels and took

more time to identify nasalized vowels than they did to identify oral vowels, suggesting that, although nasal coarticulation may aid the perception of the nasal consonant source in VN sequences, it can at the same time cause perceptual ambiguity as to the quality of the vowel itself.

Most of the previous literature on both the production and perception of vowel nasality in English has focused on varieties of American English. Limited research has suggested that VN sequences in British English, for example, are characterized by a smaller degree of nasal coarticulation than American English (e.g., Hartley and Preston, 2002; Hosseinzadeh et al., 2015). Since a greater degree of nasal coarticulation on VN sequences facilitates the correct identification of nasal codas (Ohala and Ohala, 1995; Beddor and Krakow, 1999; Beddor et al., 2013; Zellou, 2017), one might speculate that nasal codas in American English would be easier to identify than those in British English. At the same time, since a greater degree of nasalization can also result in greater acoustic and perceptual ambiguity of vowel quality in these same VN sequences (Beddor et al., 1986; Krakow et al., 1988; Scarborough and Zellou, 2012; Carignan, 2018a,b), one might also speculate that the vowel quality of nasalized vowels in American English is more difficult to identify than those in British English. We test these predictions in the current study.

1.3 Speech perception during human-computer interaction

We are in a new digital age: humans are producing language to, and understanding speech from, voice-enabled devices, such as Amazon's Alexa, Apple's Siri, and Google Assistant. While cross-dialect communication is a major topic in speech perception (e.g., Clopper, 2014), exploring issues in cross-dialect perception during human-*computer* interaction is an understudied area. Indeed, the customizability of modern voice-AI devices allows users who otherwise might not interact with people who speak a different dialect than them to change their voice setting to an extra-local accent. Indeed, recent work has found that slightly less than half of US users report that they switch their voice-enabled devices to an extra-local accent (British, Australian, Indian English) (Bilal and Barfield, 2021). Other work has found results suggesting that listeners find some non-local accents more likable than local accents (Dodd et al., 2023). Therefore, asking how same- vs. cross-dialect perception of various acoustic features and coarticulatory cues will be an increasingly relevant focus of study as speech technology becomes even more integrated into daily life. In the current study, we use TTS voices of the highest quality–neural speech synthesis–to generate our stimuli in multiple voices from speaker datasets of native US- and UK-accented talkers.

1.4 Current study

In the current study we investigate patterns of correct identification, as well as coda and vowel errors, for words with nasal and oral codas (i.e., containing coarticulatory nasalized and oral vowels, respectively) across American English (US) and British English (UK) TTS voices, as perceived by both US and UK listeners.

2 Methods

2.1 Stimulus materials

The target items used were 12 monosyllabic words with the onset/b/, one of six non-high vowels as the nucleus (/æ/, /ɛ/, /eI/, /ʌ/, /ɑ/ [US] or /ɒ/ [UK], and /oʊ/ [US] or /əʊ/ [UK]), and either the oral coda/d/or the nasal coda/n/: *bad, ban, bed, ben, bade, bane, bod, bon, bun, bud, bode, bone* (we focus on non-high vowels following prior work on nasal coarticulatory patterns in American English, Beddor and Krakow, 1999; Zellou, 2022, as well as consistent acoustic effects observed for non-high vowels cross-linguistically, Carignan, 2018a). The stimuli were generated using Amazon Polly, an online service that provides a Neural TTS system to produce high quality natural and human-like sounding text-to-speech voices (Amazon, 2022). Two synthetic female speakers were selected for each dialect: "Salli" and "Joanna" were chosen as the US voices and "Amy" and "Emma" were chosen as the UK voices. The generated stimuli were amplitude-normalized to 70 dB and mixed with white noise at a 0 dB SNR (Miller and Nicely, 1955), and the stimuli were padded with 250 ms of noise on both ends.

To examine the patterns of nasal coarticulation across the dialects of these stimuli, the acoustic nasality properties of the vowels in the items were measured. The degree of acoustic nasalization is assessed using a spectral measure: A1-P0, or the difference in the amplitudes of the first formant spectral peak (A1) and the lowest frequency nasal formant peak (P0) (Chen, 1997). A1-P0 is, thus, the relative difference between the oral and nasal formants. As the relative degree of nasalization increases, the amplitude (in dB) of nasal formant peaks increases, while oral formant peaks become dampened. Smaller A1-P0 values are associated with greater vowel nasalization. We measured A1-P0 at vowel midpoint for each item.

The mean A1-P0 values for oral and nasalized vowels from the US and UK TTS stimuli are provided in Figure 1. A mixed effects linear regression was run on the A1-P0 data with fixed effects of dialect (UK, US) and vowel type (oral, nasalized). Effects were sum-coded. The model also included by-speaker random effects.

As seen in Figure 1, nasalized vowels have overall lower A1-P0 values than oral vowels (coef. $= -0.9$, SE $= 0.3$, $t = -3.1$, $p < 0.01$), indicating that they are more acoustically nasalized. Furthermore, an interaction between dialect and vowel type (coef. $= 0.6$, SE $= 0.3$, $t = 2.1$, $p < 0.05$) indicates that the relative difference in degree of nasalization between oral and nasalized vowels is smaller for UK than for US items. This indicates that US nasal coarticulated vowels contain greater degree of nasality (lower A1-P0 values) than UK nasalized vowels. These patterns confirm prior reports that US speakers produce greater coarticulatory nasality than UK speakers.

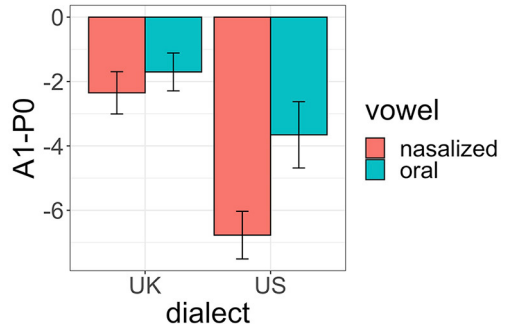

FIGURE 1
Relative degree of acoustic nasalization (in A1-P0 dB) for oral and nasalized vowels in the UK and US stimuli.

2.2 Listeners

Sixty participants, who were native US and UK listeners between the ages of 18 and 24, completed the study. The US listener group consisted of 29 participants from California (mean age: 20.4 years old; 19 female, 2 non-binary/genderqueer, 8 male) and the UK listener group consisted of 31 participants from the south of England (mean age: 21.9 years old; 21 female, 0 non-binary, 10 male). Participants in both groups were recruited through prolific, an online platform where they could voluntarily sign up to take part in the study and receive a payment for their contribution. The respondents were native English speakers in the specified age range, had grown up in either California or the south of England, and had not lived abroad for more than a year.

2.3 Procedure

The listeners completed a six-option forced-choice word identification paradigm. On a given trial, listeners were presented with a single stimulus item and instructed to identify the word in a forced-choice paradigm. After the item was presented auditorily, the six word options were presented to listeners on the screen. The options included: (1) the target word (e.g., "bad") [ACC], (2) a word contrasting in coda nasality (e.g., "ban") [C], (3) a word contrasting in vowel quality along the height dimension (e.g., "bed") [V], (4) a word contrasting in both coda nasality and vowel quality (e.g., "ben") [VC], (5) another vowel height contrast (e.g.," bade") [V], (6) the second vowel contrast with a coda nasality change (e.g., "bane") [VC].

Table 1 provides all the options provided for each target word. The participants were instructed to select the word they heard out of the 6 options. If the target word contained a front vowel, the competitors also contained front vowels; likewise, if the target word contained a non-front vowel, the competitors also contained non-front vowels. Each word was generated by both TTS voices of both dialects, resulting in a total of 48 stimuli (12 words * 2 voices * 2 dialects), and each audio stimulus was presented twice throughout the experiment, resulting in 96 trials per listener.

2.4 Statistical analysis

Listeners were effectively asked to identify the vowel and final consonant for each stimulus word, resulting in four possible outcomes: both the vowel and coda are incorrect, the vowel is incorrect, the coda is incorrect, both the vowel and coda are correct (accurate). Responses were modeled using a Bayesian multinomial regression model using the brms package (Bürkner, 2017) and Stan (Stan Development Team, 2023) in R (R Core Team, 2021). These models predict the probability of observing each of the four possible outcomes on any given trial as a function of the predictor variables. These models are similar conceptually to logistic regression models, with scores replacing logits and the softmax function replacing the logistic function (for more information see Barreda and Silbert, 2023, Ch. 13). Our model predicted the probability of observing each outcome as a function of the following fixed effects: speaker dialect (US or UK), listener dialect (US or UK), whether the coda consonant was oral or nasal, and the identity of the TTS speaker (four in total, two for each dialect). Sum coding was used for all fixed effects. Listener (29 US, 31 UK) and word (12 levels) were also included as random effects. The formula used for this model is presented in equation 1.

$$\text{outcome} \sim \text{speaker dialect} * \text{listener dialect} * \text{coda nasality type} + \text{speaker} + (1|\text{listener}) + (\text{speaker dialect}|\text{word}). \quad (1)$$

Logistic regression models predict the probability of observing one of two possible outcomes, "successes". Multinomial models predict J-1 outcomes for J categories, meaning in our case we will have three versions of each parameter, where the fourth (the "reference" category) is fixed to equal zero. In this case, we used the inaccurate vowel and coda outcome as the reference level.

Bayesian inference relies on the inspection of the posterior distribution of model parameters. These distributions reflect the most probable values of our model parameters given our data and model structure, including the prior probabilities of the parameters. Posterior distributions will be presented using their means, their standard deviations (analogous to the standard error), and the 95% credible interval. The 95% credible interval is analogous to the 95% confidence interval but with one important distinction: whereas a 95% credible interval is an interval that has a 0.95 probability of containing the value of the parameter, 95% confidence intervals are expected to contain the value of the true parameter in 95% of replications.

3 Results

The results are presented in Figure 2. Consideration of effects in the probability space can be somewhat misleading, especially in a situation with several probabilities near zero such as this one. However, some things are clear from the raw aggregated data. First, listeners were overall quite accurate in identifying these sounds, with no mistakes being much more likely than both mistakes (as seen in Figure 2C, overall accuracy for words with oral codas is 74% and accuracy for words with nasal codas is 60%). It also appears as though vowel errors are somewhat more

10.3389/fcomm.2023.1307547

TABLE 1 The target words and the competitor options.

Target word [ACC]	Coda nasality change option [C]	Vowel 1 change option [V]	Coda and vowel 1 change option [VC]	Vowel 2 change option [V]	Coda and vowel 2 change option [VC]
BAD	ban	bed	ben	bade	bane
BAN	bad	ben	bed	bane	bade
BED	ben	bad	ban	bade	bane
BEN	bed	ban	bad	bane	bade
BADE	bane	bed	ben	bad	ban
BANE	bade	ben	bed	ban	bad
BOD	bon	bud	bun	bode	bone
BON	bod	bun	bud	bone	bode
BUD	bun	bod	bon	bode	bone
BUN	bud	bon	bod	bone	bode
BODE	bone	bod	bon	bud	bun
BONE	bode	bon	bod	bun	bud

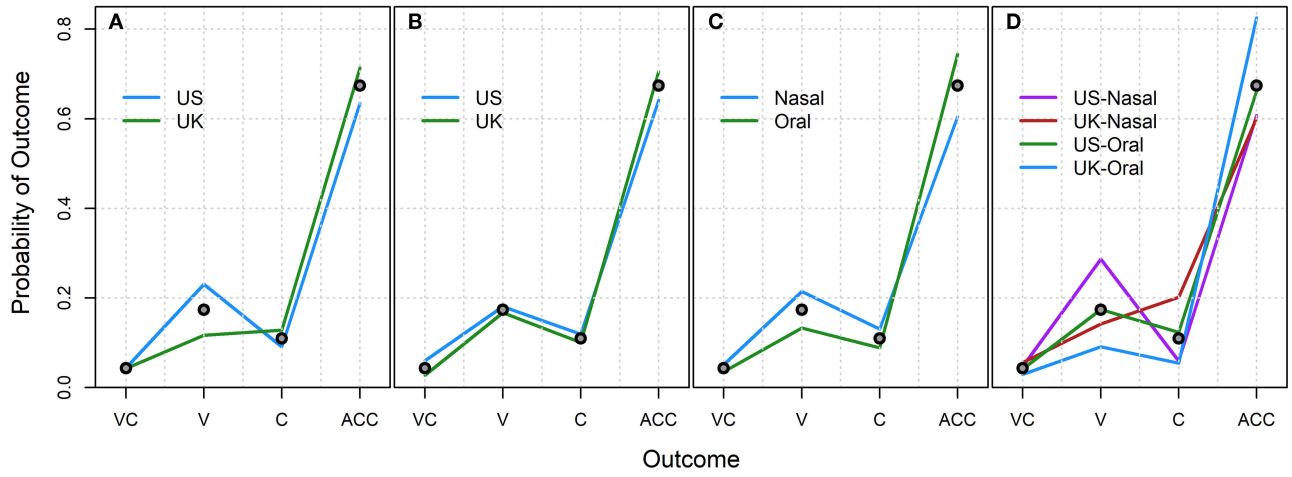

FIGURE 2
Probability of each outcome separated by **(A)** speaker dialect, **(B)** listener dialect, **(C)** coda nasality, and **(D)** speaker dialect by coda nasality, averaged across all trials. Gray points indicate the marginal probability of observing each outcome across all conditions (V, vowel error; C, coda error; ACC, accurate).

likely than consonant errors (for Oral Coda words: Vowel errors = 13% vs. Coda errors = 9%; for Nasal Coda words: Vowel errors = 21% vs. Coda errors = 13%). Although listener dialect does not seem to matter much in response patterns, speaker dialect does, as does coda nasality. There also appears to be variation in accuracy and error patterns across speaker dialects, as seen in Figure 2D. In particular, overall accuracy is highest for UK Oral Coda words (83%) and lowest for words with Nasal Codas in both dialects (UK Nasal = 61%; US Nasal = 60%). Accuracy for US Oral Coda words is 67%. Vowel errors were most likely for US Nasal words (21%), followed by UK Nasal words (14%), then Oral Coda words in both dialects (US = 13%; UK = 9%). Coda errors were highest for UK Nasal words (20%) and lowest for US Oral (6%) and UK Oral (5%) words. US Nasal word coda errors were 12%.

The full output of the Bayesian analysis is provided in the Supplementary material. Figure 3 presents the posterior distribution of parameter fixed effects, grouped by type of effect (e.g., intercepts, speaker dialect) across modeled response variables. As seen, the highest proportion of responses were Accurate. Vowel errors were the next most likely response, followed by Consonant errors.

As seen in Figure 3, the posterior probability distributions and credible intervals for each category outcome for the interaction between Speaker Dialect and Coda Nasality are also quite large and do not cross zero. Figure 4 presents predicted probabilities for each outcome, divided according to speaker dialect and coda nasality condition. Table 2 presents pairwise comparisons of subsets of these predicted probabilities. Firstly, we observe that there is a negative estimate for differences in Accurate responses

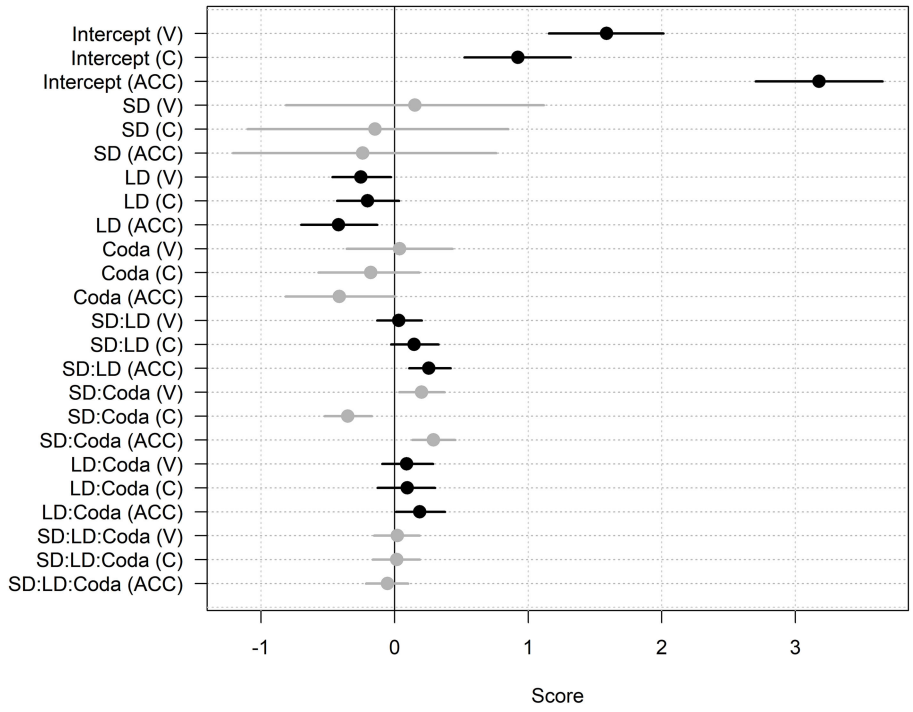

FIGURE 3
Posterior probability distributions for all model fixed effects. Points indicate posterior means, lines indicate the 95% credible intervals for each parameter. Colors reflect groups of parameters, one for each modeled outcome category (V, vowel error; C, coda error; ACC, accurate).

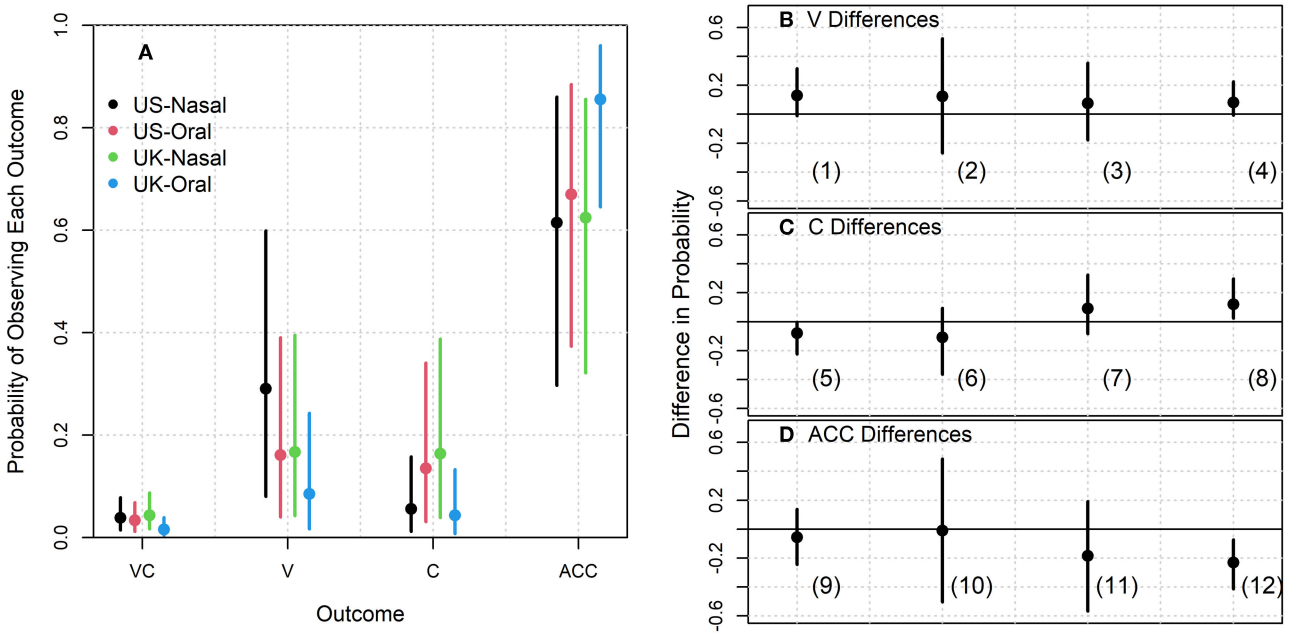

FIGURE 4
(A) Modeled probability of observing each outcome (VC, vowel and coda error; V, vowel error; C, coda error; ACC, accurate) as a function of speaker dialect and coda nasality. Points represent posterior means, lines reflect 95% credible intervals for each parameter. **(B–D)** Posterior means and credible intervals for selected pairwise comparisons of the outcome probabilities presented in **(A)**. Values reflect comparisons presented in Table 2. Each panel compares differences from outcomes for vowel errors, coda errors, and accurate outcomes independently.

between words with nasal and oral codas produced by the UK speakers. This is seen in Figure 4, particularly point (12) in the right panel, showing that words with oral codas produced by the UK speakers have the overall highest accurate responses.

TABLE 2 Posterior means, standard deviations (sd), 2.5% and 97.5% credible intervals (CI) for the pairwise differences presented in Figures 4B, C.

	Mean	sd	2.5% CI	97.5% CI
1. V Error: US Nasal vs. US Oral	0.13	0.09	−0.01	0.32
2. V Error: US Oral vs. UK Oral	0.12	0.20	−0.27	0.52
3. V Error: US Nasal vs. UK Nasal	0.08	0.13	−0.18	0.35
4. V Error: UK Nasal vs. UK Oral	0.08	0.06	−0.006	0.23
5. C Error: US Nasal vs. US Oral	−0.08	0.061	−0.22	−0.002
6. C Error: US Oral vs. UK Oral	−0.11	0.12	−0.36	0.09
7. C Error: US Nasal vs. UK Nasal	0.09	0.10	−0.08	0.32
8. C Error: UK Nasal vs. UK Oral	0.12	0.07	0.02	0.30
9. ACC: US Nasal vs. US Oral	−0.06	0.09	−0.24	0.14
10. ACC: US Oral vs. UK Oral	−0.01	0.25	−0.50	0.48
11. ACC: US Nasal vs. UK Nasal	−0.19	0.19	−0.57	0.19
12. ACC: UK Nasal vs. UK Oral	−0.23	0.09	−0.41	−0.07

Values represent the difference in outcome probabilities (V, vowel error; C, coda error; ACC, accurate) based on speaker dialect and coda nasality.

Accuracy decreases for words with nasal codas produced by UK speakers.

The comparisons also reveal that the likelihood of a coda identification error was greater for CVN words produced by the UK speakers than for their CVC words [point (8) on Figure 4, right panel]; in other words, nasalized vowels are likely to be misrecognized as signaling oral coda contexts than oral vowels are to be misidentified as coming from nasal coda contexts when produced by UK speakers. In contrast, the effect is in the opposite direction for US speakers [point (5) on Figure 4, right panel]: A negative coefficient for the comparison between coda errors for CVN and CVC words produced by US speakers indicates that listeners are *less likely* to make a coda error for nasal vowels than they are for oral vowels.

Comparisons of vowel errors, meanwhile, show similar patterns of confusions based on coda nasality across dialects. For both UK [point (4) in the right panel of Figure 4] and US [point (1)] speakers, there is a greater likelihood of making a vowel error for CVN words than for CVC words.

4 General discussion

The goal of this study was to investigate perceptual confusions for words (presented in noise) containing non-high oral and nasalized vowels across American English and British English TTS voices for American English and British English listeners. Both the US listeners and the UK listeners identified the presented words correctly most of the time and there were similar types of error patterns across both listener groups.

Yet, there were systematic differences in patterns of confusions based on coda type and speaker dialect. Overall, listeners were more accurate at identifying words containing oral vowels than words containing nasalized vowels, which confirms prior work showing greater confusion of vowel quality in nasal contexts compared to oral contexts (Beddor et al., 1986; Krakow et al., 1988; Scarborough and Zellou, 2012; Carignan, 2014).

With respect to coda nasality confusions, in particular, we found differences in error patterns for UK and US speakers. CVN words were more likely to be incorrectly identified as being CVC items than CVC items were to be misidentified as CVN for UK speakers, while the reverse pattern was observed for US speakers. Why was this the case? It is well established in the literature that a greater degree of nasal coarticulation facilitates the correct identification of nasal codas as the nasality of the vowel is attributed to its phonetic environment, that is the following nasal consonant (Beddor and Krakow, 1999; Scarborough, 2013; Scarborough and Zellou, 2013; Zellou, 2022), often to the point where listeners can tell whether a coda is oral or nasal just by listening to the preceding vowel (Ohala and Ohala, 1995; Beddor et al., 2013). As American English is claimed to possess greater nasal coarticulation than British English (Hartley and Preston, 2002), as was also observed using A1-P0 measurements of our TTS voices in these two dialects, it is expected that there would be fewer errors involving coda misidentification in the words with nasal codas produced by the US speakers than in those produced by the UK speakers as the greater degree of nasal coarticulation present on US vowels would aid the listeners in classifying the presented codas as nasal.

It is also well established in the literature that a greater degree of nasal coarticulation makes it harder for listeners to identify the vowel quality of nasalized vowels as the merger of the acoustic transfer functions of the oral and nasal cavity causes a great deal of modifications to the acoustic signal reducing the distances between nasalized vowels and thus making them more acoustically similar to one another (Beddor et al., 1986; Krakow et al., 1988; Scarborough and Zellou, 2012; Carignan, 2018a,b). As American English is claimed to possess more nasal coarticulation than British English (Beddor, 1993; Hartley and Preston, 2002), it was predicted that there would be more errors involving vowel misidentification in the words with nasal codas produced by the US speakers than in those produced by the UK speakers. Yet, nasalized vowels had larger rates of vowel errors than oral vowels for both speaker groups (and there was not a difference in vowel error rates for nasalized vowels across US and

UK speakers). Thus, even though less coarticulatory nasalization in UK English makes it harder for listeners to identify the upcoming coda, it does not make vowel quality identification less challenging in noise.

The goal of this study was to examine perceptual confusions in cross-dialectal perception for TTS voices, since human-computer interaction is increasing in modern society and speech communication between human and voice-AI systems is a growing area of scientific interest (e.g., Zellou et al., 2021; Cohn et al., 2022), including in cross-cultural contexts (Gessinger et al., 2022). In particular, focusing on intelligibility disparities across and within TTS voices can provide practical suggestions for how to improve speech technology (Cohn and Zellou, 2020; Aoki et al., 2022). Therefore, the results of the present study can be considered in terms of implications for speech and language use with speech technology. For one, the lack of effects of *listener* dialect suggest that perceptual patterns for TTS voices of different dialects can be generalized across speech communities for a language. Moreover, the results of this study can be used to improve the intelligibility of TTS voices. In particular, focusing on disparities for specific word types (in this case, for instance, enhancing the coarticulatory patterns for UK nasal coda words) could improve the intelligibility of the voices.

There were also limitations of the present study, which open avenues for future work. This study only investigated perceptual confusions for oral and nasalized vowels across different varieties of English. Future studies could also investigate differences in production, ideally using nasalance and ultrasound, which have been proven to be useful in separating the effects of VP coupling on the formant frequencies (Carignan, 2018a). Moreover, the current study only used TTS voices to produce the stimuli; future studies could examine if there is a difference between American English and British English speakers in naturally produced speech as well (cf. Zellou et al., 2016). In order to obtain a set of minimal pairs, the current study made use of both real words, e.g. bad or ban, and pseudo-words, e.g., bod or bon. It is known that different American English speakers may use different degrees of nasal coarticulation and hyperarticulation for real words and pseudo-words (Scarborough, 2012, 2013). Therefore, future work could also investigate this phenomenon in British English while researchers duplicating this study could also compare and contrast nasal coarticulation in real words or pseudo-words in the two varieties of English to see if there is a difference.

The current study also does not differentiate between front vowels and non-front vowels, even though there is a documented difference in regard to the independent acoustic effects of vowel nasalization on vowel quality between the two. Non-front vowels tend to have a lower F2 when nasalized (Carignan, 2018a,b), which means that they tend to sound backer (Beddor, 1993; Delvaux, 2009). Moreover, it has been found that the American English low back vowels, /ɑ/ and /ʌ/, have a lower F2 when nasalized while the low front vowels, /æ/ and /ɛ/, have a higher F2 when nasalized (Scarborough and Zellou, 2012; Zellou et al., 2020). Therefore, not only are non-front vowels acoustically backer when nasalized but also front vowels are acoustically fronter. Ideally, a future study could examine the effects of vowel nasalization on all the phonemes of American English and British English based on their height and backness. As has been noted, there is currently very little data on differences in the production and perception of nasal coarticulation between different varieties of English. Therefore, future studies should investigate this phenomenon in other varieties of English as well. As an example, it is also commonly claimed that, similarly to American English, Australian English also utilizes more nasal coarticulation than British English, particularly its broad varieties (Pittam, 1987; Burridge and Kortmann, 2008; Cox and Palethorpe, 2014). Future work could examine the differences in the perception and production of nasal coarticulation between American English, British English, and Australian English.

5 Conclusion

In conclusion, this study examined whether the patterning of the degree of nasal coarticulation varies across speakers of American English and British English by examining perceptual confusions for oral and nasalized non-high vowels across listeners of both varieties of English. We focus on TTS voices, exploring the effect of phonetic differences across and within voices on the perceptual consequences for human-computer interaction. Our findings support claims regarding a greater confusion of vowel quality in the pre-nasal context compared to oral coda contexts as well as the facilitation of the correct identification of nasal codas that anticipatory nasal coarticulation provides. Participants in this study had more problems correctly identifying nasal codas in the words produced by the UK speakers than in those produced by the US speakers. This finding partially supports the claim that American English indeed possesses a greater degree of nasal coarticulation than British English. Yet, across both varieties, listeners showed greater likelihood of incorrectly identifying the quality of nasalized vowels than oral vowels. Therefore, even the minimal amount of co-articulatory nasalization present in UK speakers' productions leads to difficulties for listeners in correctly identifying the vowel of the target word.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by UC Davis Institutional Review Board (IRB). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JG: Data curation, Writing – original draft. SB: Formal analysis, Writing – original draft. CC: Conceptualization, Methodology, Supervision, Writing – original draft. GZ: Conceptualization,

Formal analysis, Methodology, Software, Supervision, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2023.1307547/full#supplementary-material

References

Amazon (2022). *Amazon Web Services*. Available online: http://aws.amazon.com/es/ec2/ (accessed November 12, 2021).

Aoki, N. B., Cohn, M., and Zellou, G. (2022). The clear speech intelligibility benefit for text-to-speech voices: Effects of speaking style and visual guise. *JASA Express Lett.* 2, 4. doi: 10.1121/10.0010274

Barreda, S., and Silbert, N. (2023). *Bayesian Multilevel Models for Repeated Measures Data: A Conceptual and Practical Introduction in R*. Oxfordshire: Taylor and Francis.

Beddor, P. S. (1993). "The perception of nasal vowels," in *Nasals, Nasalization, and the Velum, Phonetics and Phonology vol. 5*, eds M. K. Huffman and R. A. Krakow (Cambridge, MA: Academic Press).

Beddor, P. S. (2007). "Nasals and nasalization: The relation between segmental and coarticulatory timing," in *Proceedings of the 16th International Congress of Phonetic Sciences*, *Saarbrücken*, 249–254.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language* 165, 785–821. doi: 10.1353/lan.0.0165

Beddor, P. S., and Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *J. Acoust. Soc. Am.* 106, 2868–2887. doi: 10.1121/1.428111

Beddor, P. S., Krakow, R. A., and Goldstein, L. M. (1986). Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology* 3, 197–217. doi: 10.1017/S0952675700000646

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013). The time course of perception of coarticulation. *J. Acoust. Soc. Am.* 133, 2350–2366. doi: 10.1121/1.4794366

Bilal, D., and Barfield, J. K. (2021). Hey there! what do you look like? user voice switching and interface mirroring in voice-enabled digital assistants (VDAs). *Proc. Assoc. Inform. Sci. Technol.* 58, 1–12. doi: 10.1002/pra2.431

Bongiovanni, S. (2021). Acoustic investigation of anticipatory vowel nasalization in a Caribbean and a non-Caribbean dialect of Spanish. *Linguist. Vangu.* 7, 20200008. doi: 10.1515/lingvan-2020-0008

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *J. Statist. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Burridge, K., and Kortmann, B. (2008). *Varieties of English. The Pacific and Australasia*. Berlin: De Gruyter Mouton. doi: 10.1515/9783110208412.0.23

Carignan, C. (2014). An acoustic and articulatory examination of the oral in nasal: the oral articulations of French nasal vowels are not arbitrary. *J. Phonet.* 46, 23–33. doi: 10.1016/j.wocn.2014.05.001

Carignan, C. (2018a). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *J. Acoust. Soc. Am.* 143, 2588–2601. doi: 10.1121/1.5034760

Carignan, C. (2018b). *An Examination of Oral Articulation of Vowel Nasality in the Light of the Independent Effects of Nasalization on Vowel Quality*. Padua: Associazione Italiana Scienze della Voce.

Carignan, C. (2018c). Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Lab. Phonol.* 9, 1. doi: 10.5334/labphon.136

Carignan, C., Chen, J., Harvey, M., Stockigt, C., Simpson, J., and Strangways, S. (2023). An investigation of the dynamics of vowel nasalization in Arabana using machine learning of acoustic features. *Lab. Phonol.* 14, 1. doi: 10.16995/labphon.9152

Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *J. Acoust. Soc. Am.* 102, 2360–2370. doi: 10.1121/1.419620

Clopper, C. G. (2014). Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Lab. Phonol.* 5, 69–90. doi: 10.1515/lp-2014-0004

Cohn, M., Segedin, B. F., and Zellou, G. (2022). Acoustic-phonetic properties of Siri-and human-directed speech. *J. Phonet.* 90, 101123. doi: 10.1016/j.wocn.2021.101123

Cohn, M., and Zellou, G. (2020). "Perception of concatenative vs. neural text-to-speech (TTS): Differences in intelligibility in noise and language attitudes," in *Proceedings of Interspeech.*

Cox, F., and Palethorpe, S. (2014). "Phonologisation of vowel duration and nasalised/æ/in Australian English," in *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, 33–36.

De Decker, P. M., and Nycz, J. R. (2012). Are tense [æ]s really tense? The mapping between articulation and acoustics. *Lingua* 122, 810–821. doi: 10.1016/j.lingua.2012.01.003

Delvaux, V. (2009). Perception du contraste de nasalité vocalique en français. *J. French Lang. Stud.* 19, 25–59. doi: 10.1017/S0959269508003566

Diakoumakou, E. (2004). *Coarticulatory Vowel Nasalization in Modern Greek*. Ann Arbor: University of Michigan.

Dodd, N., Cohn, M., and Zellou, G. (2023). Comparing alignment toward American, British, and Indian English text-to-speech (TTS) voices: Influence of social attitudes and talker guise. *Front. Comp. Sci.* 5, 1204211. doi: 10.3389/fcomp.2023.1204211

Gessinger, I., Cohn, M., Zellou, G., and Möbius, B. (2022). Cross-cultural comparison of gradient emotion perception: human vs. Alexa TTS voices. *Proc. Interspeech* 2022, 4970–4974. doi: 10.21437/Interspeech.2022-146

Hajek, J. (2013). "Vowel nasalization," in *The World Atlas of Language Structures Online*, eds. M. S. Dryer and M. Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Hartley, L. C., and Preston, D. R. (2002). "The names of US English: Valley girl, cowboy, yankee, normal, nasal and ignorant," in *Standard English*. London: Routledge, 207–238.

Hosseinzadeh, N. M., Kambuziya, A. K. Z., and Shariati, M. (2015). British and American phonetic varieties. *J. Lang. Teach. Res.* 6, 647–655. doi: 10.17507/jltr.0603.23

Krakow, R. A., Beddor, P. S., Goldstein, L. M., and Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *J. Acoust. Soc. Am.* 83, 1146–1158. doi: 10.1121/1.396059

Maddieson, I. (2013). "Absence of common consonants," in *The World Atlas of Language Structures Online*, eds. M. S. Dryer, and M. Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Mielke, J., Carignan, C., and Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal/æ/-raising in English: an ultrasound study. *J. Acou. Soc. Am.* 142, 332–349. doi: 10.1121/1.4991348

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acou. Soc. Am.* 27, 338–352. doi: 10.1121/1.1907526

Ohala, J. J. (1993). Coarticulation and phonology. *Lang. Speech* 36, 155–170. doi: 10.1177/002383099303600303

Ohala, J. J., and Ohala, M. (1995). Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. Phonology and phonetic evidence. *Papers in Lab. Phonol.* IV, 41–60. doi: 10.1017/CBO9780511554315.004

Onsuwan, C. (2005). *Temporal Relations Between Consonants and Vowels in Thai Syllables*. Ann Arbor: University of Michigan.

Pittam, J. (1987). Listeners' evaluations of voice quality in Australian English speakers. *Lang. Speech* 30, 99–113. doi: 10.1177/002383098703000201

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed November 1, 2022).

Scarborough, R. (2012). Lexical similarity and speech production: neighborhoods for nonwords. *Lingua* 122, 164–176. doi: 10.1016/j.lingua.2011.06.006

Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: relating coarticulation and hyperarticulation. *J. Phonet.* 41, 491–508. doi: 10.1016/j.wocn.2013.09.004

Scarborough, R., and Zellou, G. (2012). Acoustic and perceptual similarity in coarticulatorily nasalized vowels. *Interspeech* 2012, 1408–1411. doi: 10.21437/Interspeech.2012-669

Scarborough, R., and Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *J. Acou. Soc. Am.* 134, 3793–3807. doi: 10.1121/1.4824120

Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual, Version*. Available online at: https://mc-stan.org (accessed November 1, 2022).

Stoakes, H. M., Fletcher, J. M., and Butcher, A. R. (2020). Nasal coarticulation in Bininj Kunwok: an aerodynamic analysis. *J. Int. Phonetic Assoc.* 50, 305–332. doi: 10.1017/S0025100318000282

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *J. Acou. Soc. Am.* 142, 2469–2482. doi: 10.1121/1.5008854

Tamminga, M., and Zellou, G. (2015). "Cross-dialectal differences in nasal coarticulation in American English," in *ICPhS*.

Wright, J. T. (1975). "Effects of vowel nasalization on the perception of vowel height," in *Nasálfest: Papers from a Symposium on Nasals and Nasalization,* eds. C. A. Ferguson, L. M. Hyman, and J. J. Ohala. Palo Alto, CA: Stanford University Language Universals Project, 373–388.

Wright, J. T. (1986). "The behavior of nasalized vowels in perceptual vowel space," in *Experimental Phonology*, J. J. Ohala, and J. J. Jaeger. New York: Academic Press, 45–67.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *J. Phonet.* 61, 13–29. doi: 10.1016/j.wocn.2016.12.002

Zellou, G. (2022). *Coarticulation in Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/9781009082488

Zellou, G., and Brotherton, C. (2021). Phonetic imitation of multidimensional acoustic variation of the nasal split short-a system. *Speech Commun.* 135, 54–65. doi: 10.1016/j.specom.2021.10.005

Zellou, G., Cohn, M., and Block, A. (2021). Partial compensation for coarticulatory vowel nasalization across concatenative and neural text-to-speech. *J. Acou. Soc. Am.* 149, 3424–3436. doi: 10.1121/10.0004989

Zellou, G., and Dahan, D. (2019). Listeners maintain phonological uncertainty over time and across words: The case of vowel nasality in English. *J. Phonet.* 76, 100910. doi: 10.1016/j.wocn.2019.06.001

Zellou, G., Scarborough, R., and Kemp, R. (2020). Secondary phonetic cues in the production of the nasal short-a system in California English. *Interspeech* 2020, 25–29. doi: 10.21437/Interspeech.2020-1322

Zellou, G., Scarborough, R., and Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *J. Acou. Soc. Am.* 140, 3560–3575. doi: 10.1121/1.4966232

Zellou, G., and Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *J. Phonet.* 47, 18–35. doi: 10.1016/j.wocn.2014.09.002