A practical method of estimating the time-varying degree of vowel nasalization from acoustic features

Christopher Carignan[1]

[1]*Department of Speech, Hearing and Phonetic Sciences, University College London*[a)]

This paper presents a simple and easy-to-use method of creating a time-varying signal of the degree of nasalization in vowels, generated from acoustic features measured in oral and nasalized vowel contexts. The method is presented for separate models constructed using two sets of acoustic features: (1) an uninformed set of 13 Mel-frequency cepstral coefficients (MFCCs) and (2) a combination of the 13 MFCCs and a phonetically-informed set of 20 acoustic features of vowel nasality derived from previous research. Both models are compared against two traditional approaches to estimating vowel nasalization from acoustics: A1-P0 and A1-P1, as well as their formant-compensated counterparts. Data include productions from six speakers of different language backgrounds producing 11 different qualities within the vowel quadrilateral. The results generated from each of the methods are compared against nasometric measurements, representing an objective "ground truth" of the degree of nasalization. The results suggest that the proposed method is more robust than conventional acoustic approaches, generating signals which correlate strongly with nasometric measures across all vowel qualities and all speakers and which accurately approximate the time-varying change in the degree of nasalization. Finally, a experimental example is provided to help researchers implement the method in their own study designs.

Keywords: vowel nasalization; A1-P0; A1-P1; nasometry; PCA regression; methodology

[a] c.carignan@ucl.ac.uk;

I. INTRODUCTION

During the production of vowel nasalization, the velum lowers to allow air to flow past the velopharyngeal (VP) port and through the nasal cavity, thereby acoustically coupling the oropharyngeal and nasal cavities. This coupling results in a wide range of modifications to the acoustic output of a VP-coupled vowel in comparison to its non-VP-coupled (i.e. oral) counterpart. These modifications include: reduction of formant amplitudes, widening of formant bandwidths, modulation of formant frequencies, shifting of spectral energy toward lower frequencies, and addition of poles (i.e. formants) and zeros (i.e. anti-formants) to the acoustic spectrum (Carignan, 2018; Chen, 1997; Feng and Castelli, 1996; Fujimura and Lindqvist, 1971; Maeda, 1993; Styler, 2017, *inter alia*). Due to the myriad acoustic effects of VP-coupling, many acoustic metrics have been proposed to capture and characterize the degree of nasalization in vowels. Some have focused on general spectral shape (Feng and Castelli, 1996; Pruthi and Espy-Wilson, 2004), others have focused on spectral modifications in specific regions (Carignan, 2018; Hawkins and Stevens, 1985; Stevens *et al.*, 1987), and yet others have focused on identifying nasal poles in the spectrum (Chen, 1997; Maeda, 1993).

Recently, Styler (2017) compared the efficacy of 22 acoustic features in distinguishing oral and nasal(ized) vowels in both English and French. Three features were found to be the most effective in capturing nasalization in the acoustic signal: A1-P0, F1 bandwidth, and spectral tilt. However, the author observed that even those features varied considerably across speakers and between the two languages, concluding that "the acoustic nature of vowel nasality is both language- and speaker-specific" (*ibid.*, abstract). Of particular interest is the observation that A1-P0 emerged as an effective acoustic correlate of nasalization, given that it is the most widely and frequently used acoustic metric in the literature on vowel nasality. A1-P0 has been used to make substantial advances in our understanding of how vowel nasality is implemented across languages (Garellek *et al.*, 2016; Khattab *et al.*, 2018), speakers (Kim and Kim, 2019) and listeners (Zellou, 2017), how it is affected both by prosodic factors (Cho *et al.*, 2017; Jang *et al.*, 2018; Zellou and Scarborough, 2012) and lexical factors (Scarborough, 2013; Scarborough and Zellou, 2013), and how it can serve as a catalyst for sound change (Beddor, 2009; Zellou and Tamminga, 2014).

A. Exploring the poles: A1-P0 and A1-P1

Capitalizing on the introduction of (nasal) poles to the acoustic spectrum of VP-coupled vowels, Chen (1997) proposed two measures based on the relationship between the amplitudes of oral and nasal poles, as determined by spectral harmonics: A1-P0 and A1-P1. A1 refers to the amplitude of the highest harmonic within F1, whereas P0 and P1 refer to the respective amplitudes of harmonics associated with nasal poles. Chen (1997) proposed a range of 250-450 Hz for the location of P0 (250-400 Hz was proposed by Maeda, 1993 for males), and a range of 790-1100 Hz for the location of P1. As the degree of nasalization increases, the amplitudes of oral poles are expected to decrease while the amplitudes of nasal poles are expected to increase. Thus, the difference in these oral-nasal pole amplitudes is expected to exhibit an inverse relationship with the degree of nasalization. While A1-P0 (i.e. the difference between the amplitude of the most prominent F1 harmonic and the amplitude of the low-frequency nasal pole) was introduced as the more robust of the two measures (and is indeed the most commonly used in the literature on vowel nasality), Chen (1997) noted that there are some cases where A1-P0 may fail—in particular, vowels with low F1 frequency, for which A1 and P0 may in fact be associated with the same harmonic. She thereby proposed A1-P1 as an appropriate substitute for high vowels. Correction functions were also given to rectify such problems by taking into account the relative position and bandwidths of nearby formants.

Figure 1 shows overlaid spectra measured in 50 ms windows of oral /a/ (red, solid line) and nasalized /a/ (blue, dotted line), extracted from a token produced by speaker S1 of the current study. In this particular token, P0 is estimated at 230 Hz and A1 is estimated at 800 Hz. In comparing the oral and nasalized spectra, the amplitude of P0 is similar, but the amplitude of A1 is reduced in the nasalized spectrum. Thus, A1-P0 is lower for nasalized /a/ than oral /a/, as expected.

Figure 2 shows overlaid spectra measured within 50 ms windows of oral /i/ (red, solid line) and nasalized /i/ (blue, dotted line), also extracted from a token produced by speaker S1. In this particular token, P0 is estimated at 330 Hz; however, given the low F1 for /i/, A1 is also estimated at 330 Hz. This token therefore represents an example where the A1-P0 measure would fail, since the same harmonic is chosen for both P0 and A1. Thus, A1-P1 is a more appropriate measure in this case. However, determining which harmonic represents P1 is not entirely straightforward, since none of them is particularly prominent. The 7$^{\text{th}}$ harmonic is at 760 Hz, the 8$^{\text{th}}$ at 870 Hz, the 9$^{\text{th}}$ at 980 Hz, and the 10$^{\text{th}}$ at 1090 Hz, with a monotonic decrease in amplitude across the four harmonics. Thus, each of these harmonics

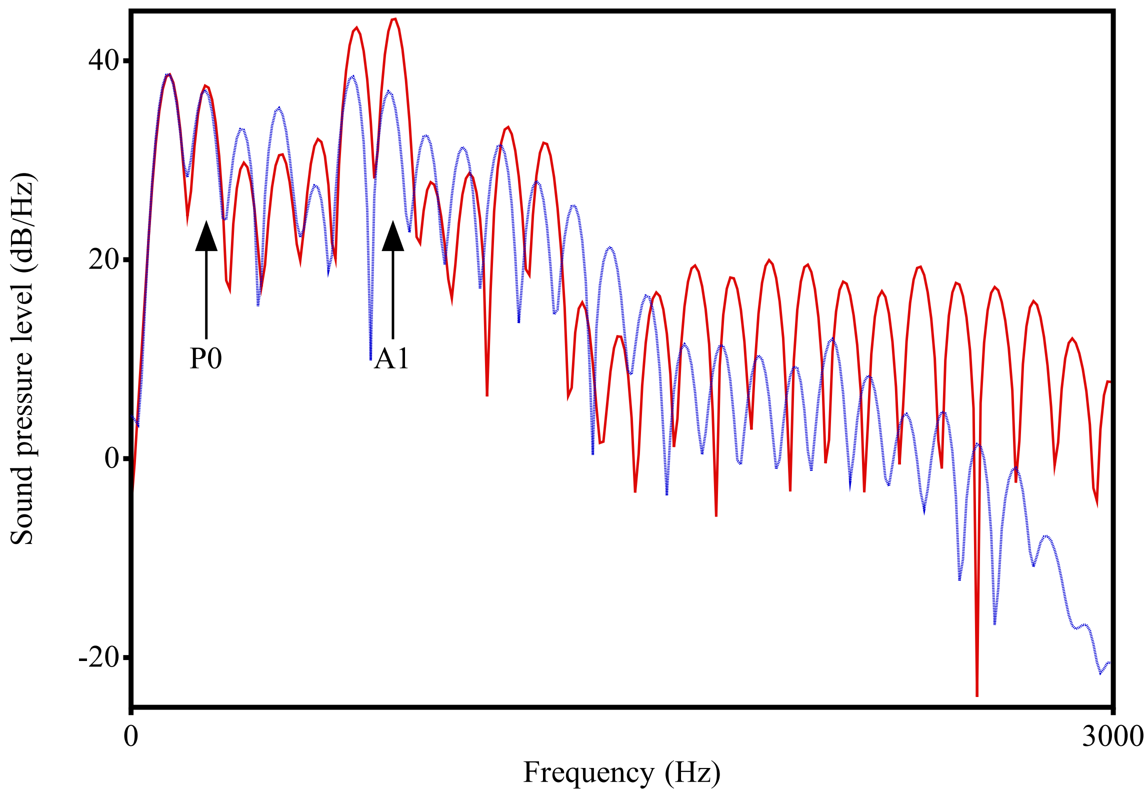

FIG. 1. (color online) Overlaid spectra of oral (red, solid line) and nasalized (blue, dotted line) /a/.

is arguably a potential candidate within the estimated range of 790-1100 Hz proposed for the location of P1 (Chen, 1997).

It may be clear by this point that comparing amplitudes of A1, P0, and P1 can quickly become an exercise in counting harmonics, a task that is not always as straightforward as might be assumed. This task is further complicated by the dynamic nature of speech harmonics: as the fundamental frequency (F0) changes, the frequencies of the harmonics of F0 change as well. Thus, individual harmonics that are relatively prominent at one F0 may become less prominent at another, since the amplitude of each harmonic is dependent

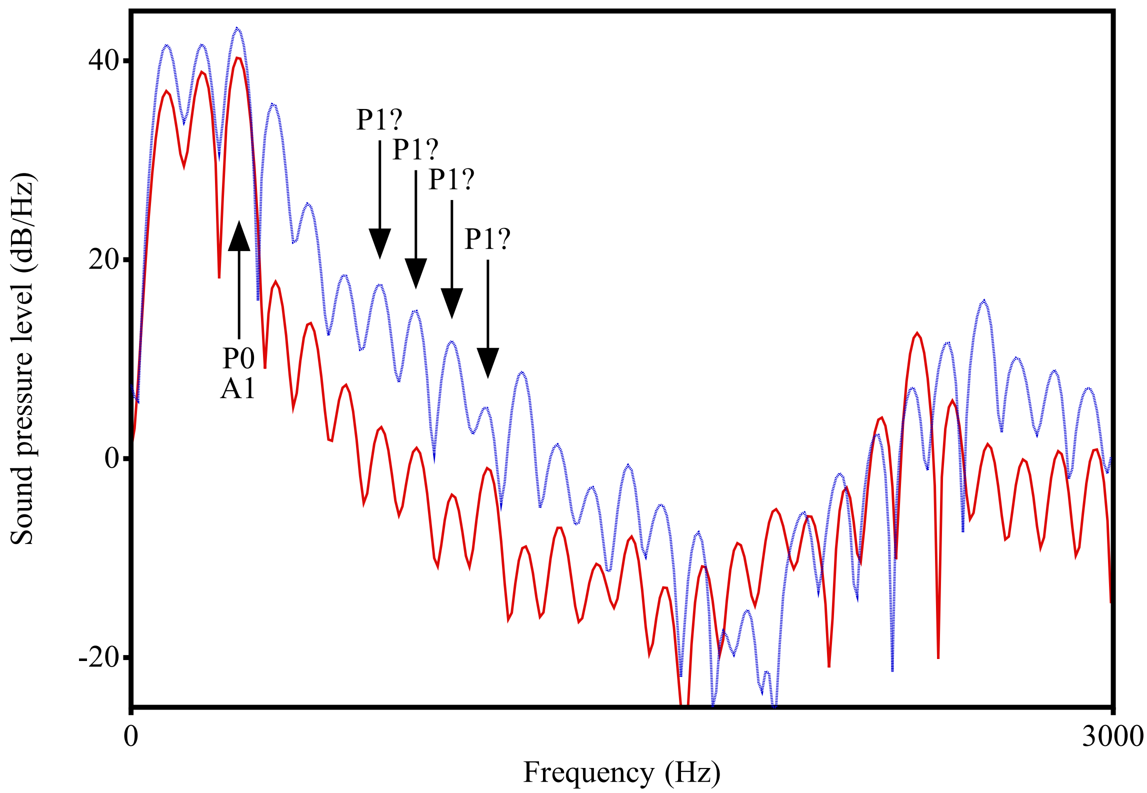

FIG. 2. (color online) Overlaid spectra of oral (red, solid line) and nasalized (blue, dotted line) /i/.

upon its frequency "location" within the acoustic transfer function. This issue is further exacerbated as F0 increases (e.g., relatively high F0 of female speakers or children), causing the harmonics to spread throughout the acoustic transfer function due to increased inter-harmonic spacing.

Ultimately, focusing on a single acoustic metric is likely not the best approach to characterizing the degree of vowel nasalization in a way that is both accurate and robust across different vowel qualities, speakers, and languages. Instead of focusing on a *single* acoustic correlate of nasalization, the current study proposes a method of generating an estimate of the degree of nasalization from speaker-specific models trained on a *wide range* of possible acoustic correlates. In this way, the fidelity of the resulting metric is not diminished by the fact that the accuracy of any given individual correlate may vary across different speakers and contexts. Throughout the paper, we will refer to the proposed method as the **NAF** method (**N**asalization from **A**coustic **F**eatures).

II. METHODOLOGY

The data and code used to generate the models and figures appearing in this article are available at https://github.com/ChristopherCarignan/NAF.

A. Nasometry and nasalance

Nasometry was used in order to obtain an objective measure of the degree of nasalization in a way that does not alter or impede upon speech acoustics (cf. aerodynamic measurements of nasal airflow which use a mask placed over the mouth and/or nose). The data used here come from nasometry recordings from Carignan (2018), collected using a Glottal Enterprises H-SEP-MU, which consists of two directional microphones located on either side of an acoustic baffle that surrounds the speaker's upper lip. The microphone above the baffle thus captures the acoustic energy radiating from the nose, while the microphone below the baffle captures the acoustic energy radiating from the mouth. The two signals were combined in order to create a single, merged audio signal for acoustic analysis (Section II C). Separate amplitude (dB) tracks for the oral and nasal signals were created in Praat (Boersma and Weenink, 2017), and a measurement of the degree of nasalization (called "nasalance") was derived by calculating the proportional nasal amplitude, i.e. $A_{nasal}/(A_{oral} + A_{nasal})$; see Dow (2020) for discussion of the Differential Energy Ratio, an alternative approach to quantifying nasometric data. Throughout this paper, this nasalance measure will be referred to as the "ground truth", as it represents an objective (albeit indirect) measurement of the changing degree of nasalization. For each speaker, nasalance measures were z-score normalized for each vowel category (Section II B), in order to control for variation in intra-oral airflow impedance arising from differences in tongue position across vowels.

B. Speakers and experimental task

Native speakers of six different languages/dialects participated in the study (American English, Australian English, Mandarin, Cantonese, French, and Hungarian): four males and two females, with a mean age of 31.3 years (SD = 7.5). All speakers were either graduate students or professional academics in phonetics and/or phonology. The speakers were instructed to produce 20 sustained repetitions of each of the 11 vowels /i, ɪ, e, ɛ, æ, a, ɑ, ɔ, o, ʊ, u/; the repetitions were carried out in individual blocks for each vowel, proceeding in the order indicated above. For each repetition, the speaker was instructed to sustain phonation of an oral quality of the vowel, then subsequently lower the velum during the sustained phonation while attempting to maintain tongue posture. During the productions, the experimenter monitored tongue posture on a GE LOGIQ *e* ultrasound system. If the experimenter judged the tongue posture to have changed substantially, the item was repeated; this process continued until 20 repetitions of each vowel were obtained that each displayed minimal change in tongue posture. Due to the relatively difficult nature of the experiment, speakers were sometimes unable to achieve the task for a particular vowel or were unable to obtain 20 repetitions; in these cases, the speaker was instructed to advance to the next target vowel in the set.

C. Acoustic features

A total of 33 acoustic features were obtained in Praat at 5 ms intervals within a 500 ms window centered on the point of maximum velocity in the nasalance signal of each token, i.e. the point of the most rapid change from oral to nasal. This point of maximum velocity will be referred to as the "onset of nasalization" for the sake of simplicity, even though it does not correspond to traditional estimates of gesture onset, e.g., 20% velocity thresholds (Kroos, 1996). Thus, the first 50 samples of each token (i.e. 0-250 ms) correspond to an oral portion of the vowel up to the onset of nasalization, and the last 50 samples (i.e. 250-500 ms) correspond to a nasalized portion of the vowel beginning with the onset of nasalization.

18 acoustic features of nasality were measured using the Nasality Automeasure Praat script[1]: the frequency, amplitude, and bandwidth of F1-F3; P0 and P1 amplitude; P0 prominence; A1-P0 and A1-P1, as well as their formant-compensated analogs; A3-P0; and H1-H2. The script was run in "Full-Auto" mode with defaults for all parameters, with the exception of the formant estimation range, which was set at 5000 Hz for males and 5500 Hz for females (default: 5300 Hz). Although A1-P0 and A1-P1 are expected to exhibit an inverse relationship with the degree of nasalization, the values were inverted for the purposes of this study—i.e. so that an increase corresponds to an increase in the degree of nasalizaiton, and vice versa—for easier comparison with both the nasalance signal and

the signal generated by the NAF method. Additionally, the center of gravity within the region of 0-5000 Hz (Styler, 2017) and a measure of nasal murmur—quantified as the ratio of low frequency (0-320 Hz) amplitude to high frequency (320-5360 Hz) amplitude (Pruthi and Espy-Wilson, 2004)—were made in order to capture broad spectral changes. In addition to these 20 phonetically-informed acoustic features, 13 Mel-frequency cepstral coefficients (MFCCs) were calculated in Praat, representing a set of phonetically-uninformed features. MFCCs are widely used as features in speech recognition, including automatic classification of vowel nasalization (Liu *et al.*, 2019).

D. Principal components transformation of features

The basic principle of the NAF method is first to determine the speaker-specific mapping of acoustic features to the realization of nasality via statistical modeling of a training set of data, and then to use this speaker-specific model to predict the degree of nasalization in a testing set of data. Since most of the acoustic features used here were created with the express purpose of identifying and characterizing vowel nasalization, a high degree of multicollinearity in the feature set is expected. As such, the features cannot be used as-is as predictor variables in statistical modeling, since the collinearity would inflate the standard error of the individual dimensions in the training data, leading to instability of the partial regression coefficients and, subsequently, the inability to use the model to accurately predict the degree of nasalization in the testing data. Thus, for each speaker, principal components analysis (PCA) was carried out in order to de-correlate the acoustic features. The resulting (orthogonal) PC scores were then used as independent variables in linear regression,[2] as described in the following section (Section II E). Two PCA models were created for each speaker: one for the uninformed set of acoustic features (the 13 MFCCs) and one for a combination of the 13 MFCCs and the set of 20 acoustic features of vowel nasality described in Section II C. Throughout the paper, these feature sets and their corresponding results will be referred to as the "reduced NAF method" (13 features) and the "full NAF method" (33 features).

E. Speaker-wise training of acoustic features

For each speaker, the total number of tokens for each of the vowel qualities was split via random sampling into training and testing sets using a 75%-25% training-testing ratio. Nasalance data from the training tokens were separated into the lower quantile (i.e. bottom 25% of nasalance values) and the upper quantile (i.e. top 25% of nasalance values) for each speaker; these observations were considered as "oral" data and "nasal" data, respectively, for the purposes of model training. This resulted in an average of 3154 data points in each of the oral and nasal categories for each speaker ($SD = 833$). Oral data were coded as '0' and nasal data were coded as '1'; this coding was used as a *numeric* dependent variable in a linear regression model, with a linear combination of the corresponding PC scores used as predictors. Although logistic (i.e. binomial) regression can be used in a similar manner (with 0 and 1 used as *categorical* contrast coding), logistic regression runs the risk of perfect separation of the training data due to the large number of acoustic features of nasalization. In other words, the problem of separating the oral and nasal tokens becomes *too easy* when fitting a binomial distribution to dimensions which do not have enough overlapping values for the two categories. Thus, a linear model with 0 and 1 as numeric values instead of categorical factors was used here. This therefore assumes that the acoustic mapping between the (PCA-transformed) acoustic features and the degree of nasalization is linear: features with values that are half-way between those associated with oral (0) and nasal (1) correspond to a half-degree of nasalization (0.5), and so forth. Any non-linearity in the mapping between acoustic features and the degree of nasalization is thus expected to result in reduced model accuracy.[3]

F. Generating a time-varying nasalization signal

The regression models were then used to predict response scores for the testing set, using the PC scores of the test tokens as predictors (mean = 4183 data points, $SD = 1044$). The resulting predictions thus form a time-varying estimate of the degree of nasalization with a sampling rate of 200 Hz (i.e. the sampling rate of the original acoustic data). Tukey's Running Median Smoothing, implemented using the `smooth()` function from the default *stats* package in R (R Core Team, 2020), was applied to the 100 samples of each token for six experimental metrics: the full NAF method, the reduced NAF method, A1-P0, formant-compensated A1-P0, A1-P1, and formant-compensated A1-P1. Finally, each of these six metrics, as well as the nasalance values, were *z*-scale normalized for each speaker in order to compare the relative magnitudes across the different signals.

Examples of signals generated from the full NAF method (dashed line), the reduced NAF method (dotted line), and formant-compensated A1-P0 (dash-dotted line) are shown in Figure 3, with the nasalance signal (solid line) shown for reference. Example A shows a case where each of the three experimental metrics matches the nasalance profile relatively well: each signal is low during the oral phase, high during the oral phase, and increases between the two phases at approximately the same time as the nasalance signal. However, whereas both reduced NAF and A1-P0 over- and/or under-estimate in the oral and nasal phases, the full NAF signal closely matches the nasalance profile throughout the entire token. Example B shows a case where all three methods match the oral phase well but fail to various extents in the nasal phase; however, the full NAF signal nevertheless approximates the nasalance signal more closely in the nasal phase than either A1-P0 or reduced NAF. Example C shows a case where both the reduced NAF signal and the A1-P0 signal fail in different ways: the reduced NAF signal severely under-estimates in the nasal portion and implies a fluctuation of nasalization that is not present in the nasalance signal, and the A1-P0 signal is an inverse of the nasalance ground truth (i.e. the A1-P0 signal is higher in the oral portion and lower in the nasal portion). However, the full NAF signal matches the nasalance signal relatively well throughout the entire token.

G. Performance assessment

The performance of the NAF method will be compared against the performance of A1-P0 and A1-P1, as well as their formant-compensated counterparts. Performance will be assessed in a number of ways: the overall relationship between a given metric and the ground-truth nasalance, how well a given metric estimates the temporal onset of nasalization, and how well a given metric captures global change in nasalization over time. For the estimates of the temporal onset of nasalization, an "onset time-lag" measure was created, which is the

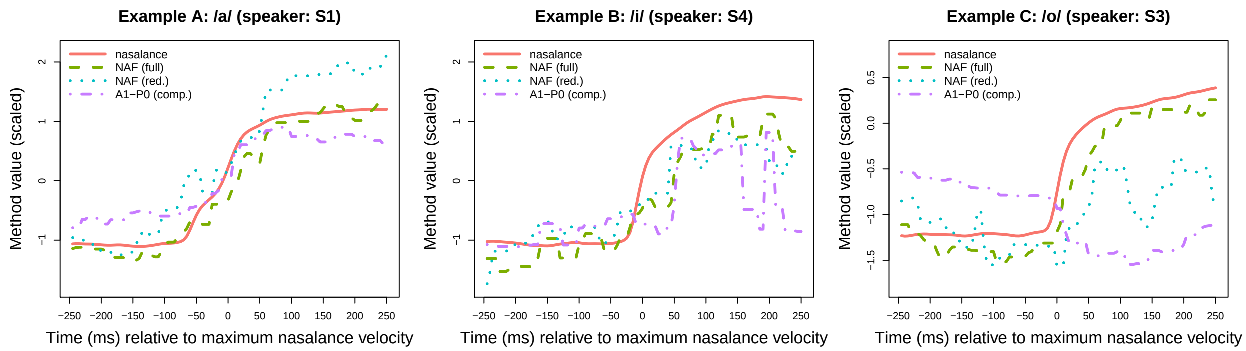

FIG. 3. (color online) Examples of nasalance values for individual tokens, along with corresponding values generated from the full NAF method, the reduced NAF method, and formant-compensated A1-P0.

difference (ms) between the point of maximum velocity in the nasalance signal and the point of maximum velocity in the signal generated by a given experimental method.

1. Correlations with nasalance

The first, basic test of the performance of the experimental methods in accurately characterizing the degree of nasalization is the strength of the correlations between the ground-truth nasalance signal and the estimate of the degree of nasalization generated by each of the methods. Two sets of Pearson's product moment correlation tests were constructed for each method, one set that included a separate test for each vowel (averaging over speakers), and one set that included a separate test for each speaker (averaging over vowels). Additionally, R^2 was used as an estimate of the variance explained in the correlation tests.

2. Bayesian regression models

In order to test for differences between the six experimental methods, Bayesian generalized mixed regression models (BRMs) were constructed using the *brms* package (Bürkner, 2020; Stan Development Team, 2017) in R. Two models were constructed: one for R^2 of the correlation between a given method and the ground-truth nasalance and one for the onset time-lag for a given method. Random intercepts for both speaker and vowel were included, together with random slopes for method over both speaker and word. The models were run with four Markov chain Monte Carlo (MCMC) chains, 2000 iterations, and 1000 warm-up samples. Model convergence was reached in all model parameters ($\hat{R} = 1$) and no divergences in the MCMC chains were observed. For each model, marginal posterior distributions were calculated for each of the six experimental methods, and both 95% and 66% credible intervals were generated from these posteriors. A credible interval (differently from a frequentist confidence interval) can be interpreted as the percentage probability that a parameter lies within that interval range; in the current study, this corresponds to an interval of possible values for the parameter μ.

The BRM for R^2 values was built using a Gaussian distribution and weakly informative priors corresponding to the belief that the mean R^2 for the intercept (the full NAF method) lies somewhere between 0 and 1 and that this value changes between -1 and $+1$ for any of the other experimental methods, at 95% confidence. In other words, since R^2 values can only be between 0 and 1, these priors allow for any possible range of values for each of the six experimental metrics. A $HalfCauchy(0,\ 0.01)$ distribution was used for the model and random intercept standard deviations, which corresponds to a 95% HDI = [0, 0.25] R^2. An $LKJ(2)$ distribution was used for the correlation between random effects, as recommended by Vasishth *et al.* (2018).

The BRM for the onset time-lag values was built using a Gaussian distribution and weakly informative priors corresponding to the belief that the mean time-lag for the intercept (the full NAF method) lies somewhere between −200 ms and +200 ms of the "true" onset of nasalization (in the ground-truth nasalance signal) and that the time-lag for each of the other experimental methods is somewhere between −100 ms and +100 ms of the time-lag of the full NAF method, at 95% confidence. A $HalfCauchy(0, 2)$ distribution was used for the model and random intercept standard deviations, which corresponds to a 95% HDI = [0, 51] ms. An $LKJ(2)$ distribution was used for the correlation between random effects.

3. Generalized additive mixed models

Generalized additive mixed models (GAMMs) were constructed in R using the `bam()` function of the *mgcv* package (Wood, 2019). Like any time-series data, it is expected that individual samples in a token will correlate with each other for the time-varying metrics examined here, resulting in correlation of the model residuals and therefore violating the model assumption of independent errors. The `bam()` function includes an optional feature intended to reduce autocorrelation using a ρ parameter, which was set to the value of the autocorrelation function (ACF) at lag = 1, i.e. ACF[2]. The first sample of each token was used as the 'AR.start' commencement point.

Random factor smooths were included for both speaker and word over time, as a function of the method used, thereby corresponding to the random effect structure used in the BRMs (Section II G 2). The model fit was assessed using the `gam.check()` function of the *mgcv* package: the k-index for all terms was ≥ 1 with a large p-value, indicating that the default number of basis functions was appropriate for these data. The parameter m was set to 1, resulting in penalization of the first derivative of the smooth (velocity) rather than the default second derivative of the smooth (acceleration), effectively acting as shrinkage of the random effects in the time-varying magnitude data used here (Wieling, 2018).

III. RESULTS

A. Global correlation with nasalance

1. Inter-vowel accuracy

Table I displays coefficients for linear correlations between each of the six experimental methods and the ground-truth nasalance signal, separated by vowel quality. These numbers thus represent the correlations for each vowel, averaged across the six speakers. The correlation strength is displayed both in text and color saturation in the top half of each cell (0: white, 1: red), and standard deviation is displayed in the bottom half. In order to enhance legibility, text appearing in more saturated cells (stronger correlations) is colored white, while text appearing in less saturated cells (weaker correlations) is colored black; the cutoff for this text color choice is 0.7—i.e. correlations ≥ 0.7 appear in white text while correlations < 0.7 appear in black text—which helps to determine at a glance which correlations meet the threshold that is conventionally used to denote a "strong" correlation. The right-most column displays values that are averaged across the 11 vowel-specific columns.

Method	/i/	/ɪ/	/e/	/ɛ/	/æ/	/a/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	Avg.
A1-P0	0.62	0.57	0.52	0.66	0.78	0.76	0.72	0.89	0.70	0.74	0.28	0.66
	(0.19)	(0.33)	(0.71)	(0.54)	(0.22)	(0.30)	(0.41)	(0.11)	(0.21)	(0.20)	(0.67)	0.35
A1-P0	0.72	0.73	0.53	0.66	0.78	0.76	0.72	0.88	0.70	0.73	0.32	0.68
(comp.)	(0.14)	(0.20)	(0.71)	(0.54)	(0.22)	(0.29)	(0.41)	(0.11)	(0.23)	(0.20)	(0.69)	0.34
A1-P1	0.83	0.68	0.66	0.71	0.37	0.05	0.29	0.25	0.13	0.27	-0.13	0.37
	(0.12)	(0.21)	(0.25)	(0.40)	(0.66)	(0.43)	(0.34)	(0.47)	(0.51)	(0.51)	(0.63)	0.41
A1-P1	0.83	0.67	0.66	0.71	0.40	0.16	0.14	0.10	0.36	0.33	0.07	0.40
(comp.)	(0.12)	(0.22)	(0.24)	(0.40)	(0.65)	(0.40)	(0.56)	(0.48)	(0.27)	(0.42)	(0.63)	0.40
NAF	0.87	0.88	0.85	0.83	0.73	0.86	0.85	0.88	0.89	0.85	0.81	0.85
(red.)	(0.11)	(0.07)	(0.11)	(0.13)	(0.12)	(0.08)	(0.04)	(0.06)	(0.05)	(0.11)	(0.05)	0.09
NAF	0.88	0.95	0.92	0.91	0.89	0.92	0.93	0.93	0.94	0.91	0.90	0.92
(full)	(0.12)	(0.02)	(0.05)	(0.08)	(0.08)	(0.04)	(0.02)	(0.03)	(0.03)	(0.08)	(0.05)	0.05

TABLE I. Vowel-specific correlations with nasalance.

With regard to A1-P0, the average correlation strength is 0.66 with an average standard deviation of 0.35. This corresponds to a moderate-to-strong correlation between A1-P0 and the ground-truth nasalance measurement. There is significant variability in the correlation strength across the different vowel qualities, with the strongest correlations occurring for low vowels and mid-back vowels (strongest: 0.89 for /ɔ/) and the weakest correlations occurring for high vowels and mid-front vowels (weakest: 0.28 for /u/). These results are as expected, given the difficulties in distinguishing A1 and P0 for high vowels (Section I A). Thus, we should expect increased accuracy for high vowels using the formant-compensated A1-P0 measurement. Indeed, in comparison with the base A1-P0 signal, the correlation coefficients for the formant-compensated A1-P0 signal are higher for /i, ɪ, u/ and remain largely unchanged for the other eight vowels, resulting in a minor increase in the average overall correlation from 0.66 to 0.68.

With regard to A1-P1, the average correlation strength is 0.37 with an average standard deviation of 0.41. This corresponds to a weak correlation between A1-P1 and the ground-truth nasalance measurement. It is evident from Table I that the poor overall performance of this metric is due to the large discrepancy between the moderate-to-strong correlations for /i, ɪ, e, ɛ/ and the no-to-weak correlations for the other seven vowel qualities. Again, these results are as expected, given the fact that A1-P1 is a metric that is used specifically for high vowels, where A1-P0 is difficult to measure (Section I A). In comparison with the base A1-P1 signal, the formant compensated A1-P1 signal results in a minor increase in the average overall correlation from 0.37 to 0.40.

With regard to the NAF method, the average correlation strength for the reduced acoustic feature set is 0.85 with an average standard deviation of 0.09. This corresponds to a strong correlation between the (reduced) NAF signal and the ground-truth nasalance measurement. Unlike A1-P0 and A1-P1, there are strong correlations for each of the 11 vowel qualities, with the weakest correlation (0.73) occurring for /æ/. However, when using the full acoustic feature set (i.e. the full NAF method), the correlation coefficients increase for all vowels, resulting in very strong correlations for each of the 11 vowel qualities, ranging from 0.88 (for /i/) to 0.95 (for /ɪ/). The average correlation strength for the full NAF method is 0.92 with an average standard deviation of 0.05, corresponding to a very strong and consistent correlation between the (full) NAF signal and the ground-truth nasalance measurement.

In summary, the NAF method produces values that correlate strongly with the ground-truth nasalance values, for both the reduced features set of 13 MFCCs and the full feature set that includes phonetically-informed acoustic measures of nasality. Both feature sets well outperform A1-P0 and A1-P1 in that the NAF method results in high accuracy for all vowel qualities, whereas A1-P0 is more accurate for low vowels and mid-back vowels and A1-P1 is more accurate for high- and mid-front vowels. Indeed, the NAF method is even highly accurate for /u/ ($R = 0.90$), a vowel that posed obvious difficulties for both A1-P0 ($R = 0.32$) and A1-P1 ($R = 0.07$).

2. Inter-speaker accuracy

Table II displays coefficients for linear correlations between each of the six experimental methods and the ground-truth nasalance signal, separated by speaker. These numbers thus represent the correlations for each speaker, averaged across the 11 vowel qualities. The formatting of the table, including the representation of correlation strengths via both text and color saturation, is the same as for Table I. The right-most column displays values that are averaged across the six speaker-specific columns. The average correlations and standard deviations for the six methods are largely the same as for Table I, with only minor deviations arising from the averages being calculated across vowels rather than across speakers. Thus, as observed in the previous section, A1-P1 yields weak correlations with nasalance, A1-P0 yields moderate-to-strong correlations, and NAF yields very strong correlations.

Method	S1	S2	S3	S4	S5	S6	Avg.
A1-P0	0.64	0.36	0.71	0.62	0.85	0.81	0.67
	(0.52)	(0.57)	(0.20)	(0.33)	(0.18)	(0.18)	0.33
A1-P0	0.66	0.42	0.75	0.63	0.85	0.84	0.69
(comp.)	(0.52)	(0.57)	(0.19)	(0.33)	(0.18)	(0.13)	0.32
A1-P1	0.48	0.08	0.10	0.48	0.47	0.55	0.36
	(0.55)	(0.53)	(0.49)	(0.57)	(0.41)	(0.29)	0.47
A1-P1	0.47	0.06	0.15	0.64	0.53	0.46	0.38
(comp.)	(0.53)	(0.53)	(0.48)	(0.35)	(0.32)	(0.48)	0.45
NAF	0.88	0.82	0.75	0.87	0.89	0.84	0.84
(red.)	(0.07)	(0.09)	(0.10)	(0.06)	(0.10)	(0.10)	0.09
NAF	0.94	0.89	0.86	0.92	0.94	0.94	0.92
(full)	(0.03)	(0.06)	(0.11)	(0.05)	(0.03)	(0.03)	0.05

TABLE II. Speaker-specific correlations with nasalance.

With regard to inter-speaker variation, there are substantial differences in accuracy for A1-P0 and A1-P1, while the NAF method produces high accuracy for each of the six speakers. As expected, the formant-compensated measures for A1-P0 and A1-P1 yield similar or increased correlations compared to their respective base metrics. Nonetheless, even these formant-compensated metrics yield poor cross-speaker performance: formant-compensated A1-P0 ranges from weak correlation ($R = 0.42$ for S2) to strong correlation ($R = 0.85$ for S5) and formant-compensated A1-P1 ranges from no correlation ($R = 0.06$ for S2) to moderate correlation ($R = 0.64$ for S4), whereas the full NAF method results in strong correlations for all speakers (range: 0.86–0.94). Indeed, the NAF method is even highly accurate for S2 ($R = 0.89$), a speaker who posed obvious difficulties for both A1-P0 ($R = 0.42$) and A1-P1 ($R = 0.06$).

B. BRM results

1. R^2 of correlation with nasalance

Figure 4 shows the marginal posteriors of the mean values of R^2 for the six methods. For this figure (as well as Figure 5, below), separate density distributions are shown for the posterior values of the six methods. Beneath each distribution, the 95% credible interval (CI) is denoted by the thin horizontal line, the 66% CI is denoted by the thick horizontal line, and the median is denoted by the dot. The results for each method are given in the text with respect to the full NAF method as the model intercept. Ranges for the estimates of the means are given for the 95% CIs of the marginal posterior distributions.

With respect to correlation with the ground-truth nasalance signal, the NAF method produces an average of 0.75-0.92 R^2 using the full acoustic feature set ($\hat{\theta} = 0.83$, $SD = 0.04$) and is changed between -0.07 and $+0.08$ R^2 using the reduced acoustic feature set

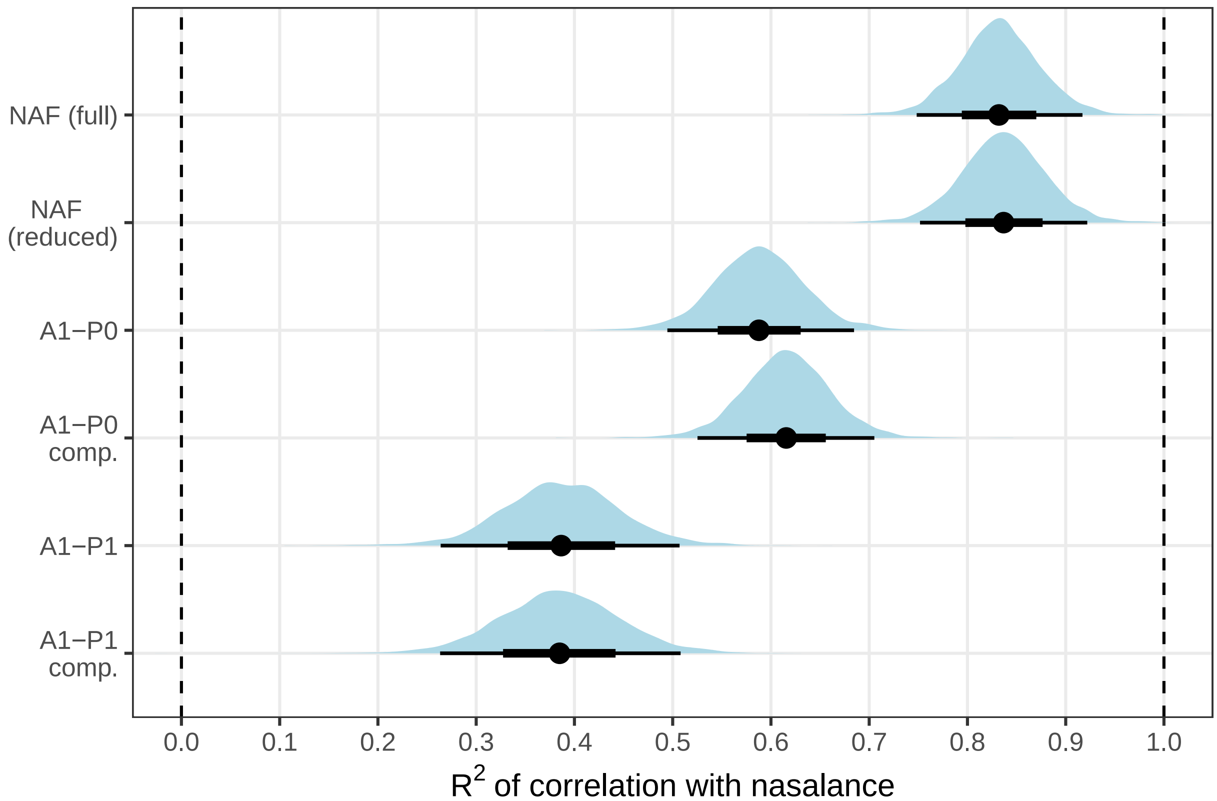

FIG. 4. Posteriors of R^2 means. The [0, 1] limits of possible R^2 values are denoted by the dashed lines.

($\hat{\theta} = 0.00$, $SD = 0.04$); in other words, the model suggests that there is no difference in R^2 between the two implementations of the NAF method. In comparison to the full NAF method, A1-P0 produces an average reduction of 0.16-0.33 R^2 ($\hat{\theta} = -0.24$, $SD = 0.04$), formant-compensated A1-P0 produces an average reduction of 0.14-0.30 R^2 ($\hat{\theta} = -0.22$, $SD = 0.04$), A1-P1 produces an average reduction of 0.33-0.56 R^2 ($\hat{\theta} = -0.45$, $SD = 0.06$), and formant-compensated A1-P1 produces an average reduction of 0.33-0.57 R^2 ($\hat{\theta} = -0.45$, $SD = 0.06$). Thus, with respect to correlation with the ground-truth nasalance signal, although A1-P0 performs better than A1-P1, both of these conventional acoustic metrics of vowel nasalization account for significantly less variance when compared to the NAF method.

2. *Estimations of the temporal onset of nasalization*

Figure 5 shows the marginal posteriors of the average time-lag of the estimated onset of nasalization for the six metrics—in other words, the error of the estimated onset with respect to the "true" onset determined from the nasalance data. With reference to the true onset of nasalization, the full NAF method estimates a time point that is between 6.91 ms early and 27.29 ms late ($\hat{\theta} = 10.51$, $SD = 8.70$). In comparison to the full NAF method, the reduced NAF method yields an estimate of the onset of nasalization that is 3.33-45.58 ms earlier ($\hat{\theta} = -24.44$, $SD = 10.77$), A1-P0 yields an estimate that is between 0.52 ms earlier and 38.08 ms later ($\hat{\theta} = 18.10$, $SD = 9.83$), formant-compensated A1-P0 yields an estimate that is between 6.10 ms earlier and 33.20 ms later ($\hat{\theta} = 13.80$, $SD = 9.75$), A1-P1 yields an estimate that is between 17.55 ms earlier and 22.59 ms later ($\hat{\theta} = 2.68$, $SD = 10.21$), and formant-compensated A1-P1 yields an estimate that is between 5.10 ms earlier and 32.50 ms later ($\hat{\theta} = 13.69$, $SD = 9.77$).

In summary, with respect to the true onset of nasalization, the full NAF method estimates a point that is on average 10.51 ms late (2.10 data samples), the reduced NAF method estimates a point that is 13.93 ms early (2.79 samples), A1-P0 estimates a point that is 29.21 ms late (5.84 samples), formant-compensated A1-P0 estimates a point that is 24.31 ms late (4.86 samples), A1-P1 estimates a point that is 13.19 ms late (2.64 samples), and formant-compensated A1-P1 estimates a point that is 24.2 ms late (4.84 samples). Thus,

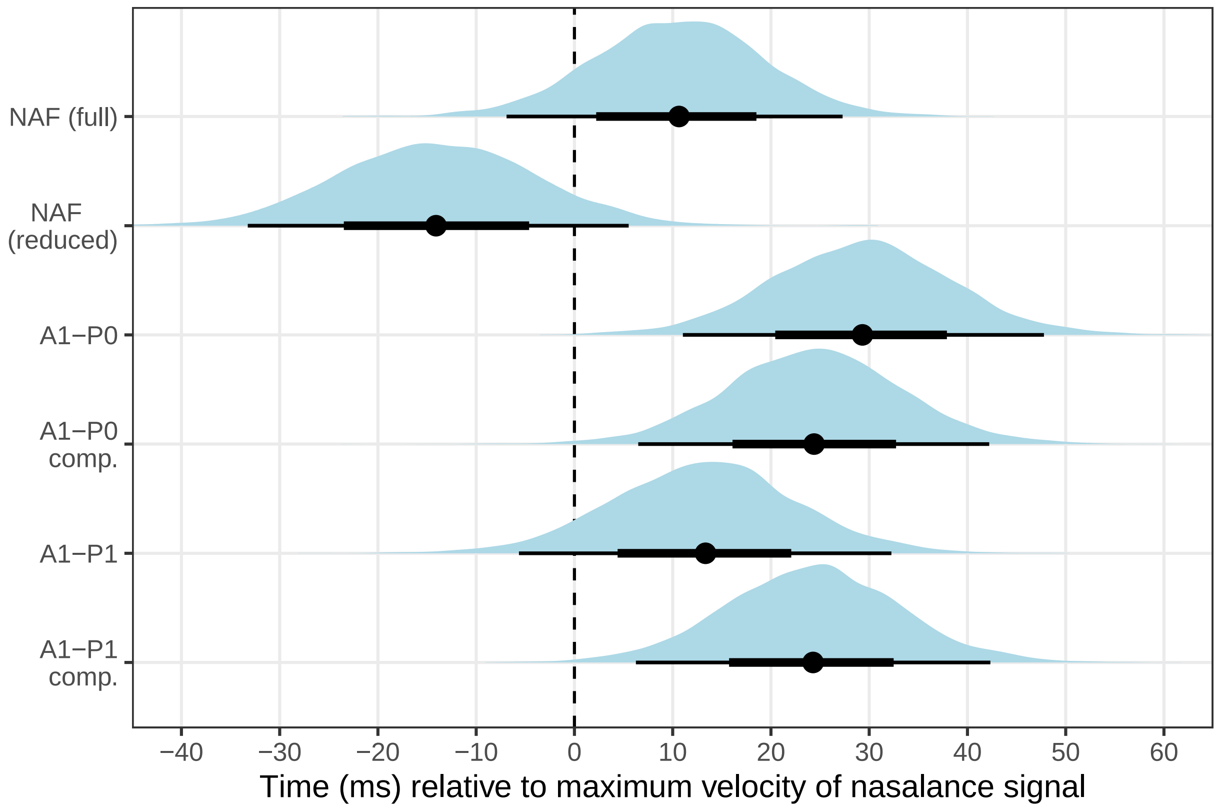

FIG. 5. Posteriors of onset time-lag means. The time point of maximum nasalance velocity is denoted by the dashed line.

each of the methods produced estimates that are, on average, within 6 samples away from the true onset at the 200 Hz sampling rate used here. However, the NAF method produces estimates with overall smaller average error ($\approx$ 2-3 samples; 10-15 ms) compared to A1-P0 ($\approx$ 4-6 samples; 20-30 ms) and A1-P1 ($\approx$ 3-5 samples; 15-25 ms).

C. GAMM results

Figure 6 displays the category fits for the GAMM constructed to examine differences between ground-truth nasalance, the full NAF method, the reduced NAF method, and formant-compensated A1-P0, with regard to the magnitude and timing of the degree of nasalization over the entire token interval. For the sake of legibility, only the ribbons corresponding to the standard error (SE) of the mean are shown (i.e. the means are not displayed here). Areas where two SE ribbons do not overlap along the y-axis are interpreted as regions of significant difference between the respective methods.

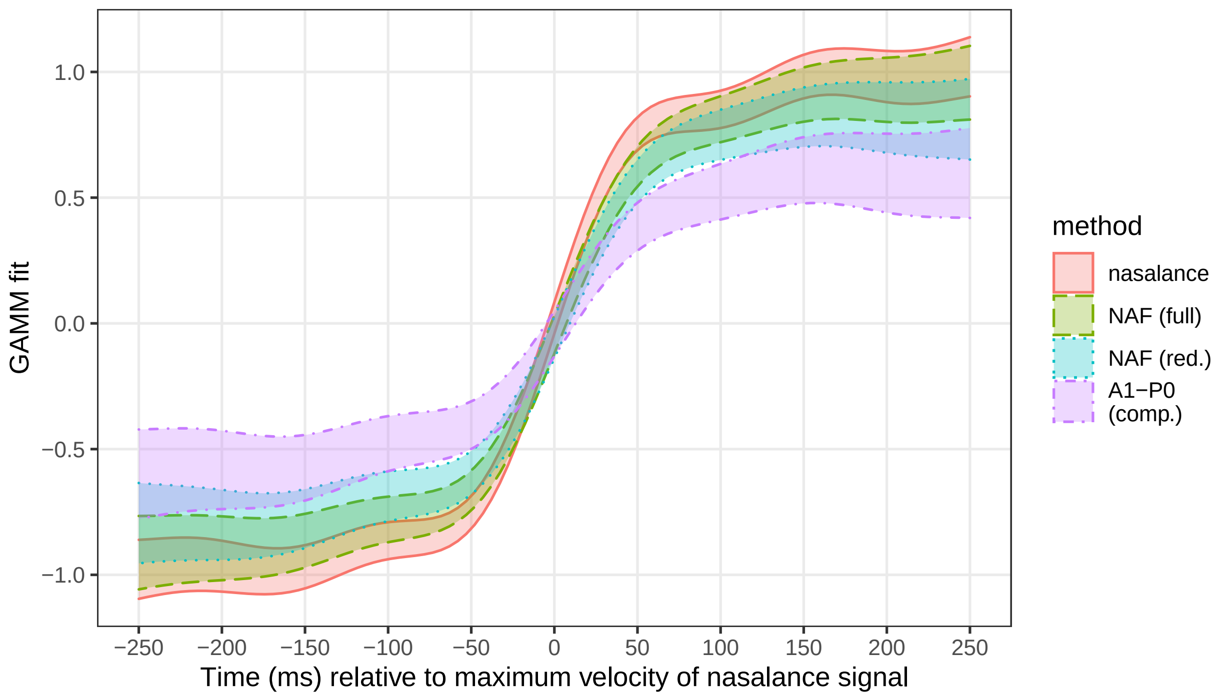

FIG. 6. (color online) GAMM fits for the entire 500 ms interval data.

Before turning to the results for the individual methods, a global observation can be made that each of the three experimental methods produces signals that approximate the nasalance profile in relative terms: the signals are all relatively low in the oral phase, relatively high in the nasal phase, and reach the point of maximum velocity near the 0 point along the x-axis (i.e. the point of maximum velocity in the nasalance signal). When using the full acoustic feature set, there are no significant differences between the NAF method and ground-truth nasalance: the SE ribbon for nasalance (red, solid lines) and the SE ribbon for the full NAF method (green, dashed lines) overlap throughout the entire token interval. When using the reduced acoustic feature set, there are no significant differences between the NAF method (blue, dotted lines) and the ground-truth nasalance, with the exception of two areas of very small difference in the regions immediately preceding the onset of nasalization (between ~ -90 ms and -50 ms) and immediately following the onset of nasalization (between $\sim +25$ ms and $+55$ ms). With regard to formant-compensated A1-P0 (purple, dash-dotted lines), there are significant differences throughout the entire oral phase and the entire nasal phase: with the exception of the region surrounding the point of maximum velocity where the respective SE bands cross (between ~ -20 ms and $+10$ ms), there is no overlap between the SE bands for nasalance and formant-compensated A1-P0.

An indication of inter-vowel and inter-speaker variation can be obtained from the model results for the random factor smooths, in which the magnitude of the F-statistic can be interpreted as the degree of variation (i.e. how important the variation is to the model); plots of the random smooths are also available in the supplementary Github material. With regard to nasalance, there is essentially no variation across vowels ($F(3, 98) = 0.56, p = 0.111$), indicating that the (vowel-scaled) nasalance profiles are similar across all 11 vowel qualities. There is a small amount of variation across vowels for the full NAF ($F(3, 98) = 1.12, p < 0.001$) and for the reduced NAF ($F(3, 98) = 1.53, p < 0.001$), and considerably more variation across vowels for formant-compensated A1-P0 ($F(3, 98) = 5.41, p < 0.001$). With regard to the by-speaker random smooth, there is a small amount of variation across speakers for each of the four metrics, with the largest amount of variation observed for formant-compensated A1-P0: nasalance ($F(3, 53) = 1.44, p < 0.001$), full NAF ($F(3, 53) = 1.82, p < 0.001$), reduced NAF ($F(3, 53) = 2, 21, p < 0.001$), and formant-compensated A1-P0 ($F(3, 53) = 3.01, p < 0.001$).

In summary, the GAMM results suggest that both implementations of the NAF method produce signals that accurately approximate the time-course of the changing degree of nasalization throughout the entire interval from 250 ms before to 250 ms after the onset of nasalization, although the full NAF method produces slightly more accurate results compared to the reduced NAF method. Conversely, formant-compensated A1-P0 does not approximate the time-course of nasalization as well as the NAF method: the signals generated from formant-compensated A1-P0 over-estimate the degree of nasalization when the velum is closed and under-estimate the degree of nasalization when the velum is open. Moreover, there is considerably more between-vowel variation in the formant-compensated A1-P0 signals in comparison with the NAF signals, suggesting yet again that this acoustic metric is not a reliable correlate of the degree of vowel nasalization for all vowel qualities.

IV. DISCUSSION

A. Comparing NAF to conventional acoustic methods

In this paper, the NAF method has been shown to approximate the ground-truth degree of nasalization (as determined using proportional nasal energy derived from nasometric recordings) more closely than A1-P0 or A1-P1. In correlations between each of the experimental methods and the ground truth, the average correlation for formant-compensated A1-P1 was 0.39, the average correlation for formant-compensated A1-P0 was 0.69, and the average correlation for the full NAF method was 0.92. This corresponds to a 33% increase in accuracy over A1-P0 and a 136% increase in accuracy over A1-P1. Moreover, according to the intercepts of the Bayesian regression model created to test for differences in the proportion of variance explained in these correlations, the average R^2 for formant-compensated A1-P1 was 0.38, the average R^2 for formant-compensated A1-P0 was 0.61, and the average R^2 for the full NAF method was 0.83. This corresponds to 36% increase in accuracy over A1-P0 and a 118% increase in accuracy over A1-P1.

The NAF method has been shown to be more robust across vowel qualities than A1-P0 or A1-P1. The range of correlation coefficients across the 11 vowels was between 0.88 and 0.95 for the full NAF method, indicating that even the "weakest" correlation with the ground-truth degree of nasalization is still very strong. In comparison, the range of correlation coefficients was between 0.32 and 0.88 for formant-compensated A1-P0 and between 0.07 and 0.83 for formant-compensated A1-P1. These results indicate that the NAF method is not only more accurate than these conventional methods in an overall sense (e.g., the strongest correlation for formant-compensated A1-P0 is equal to the weakest correlation for the full NAF method) but also more consistently accurate across different vowel qualities, even those that yielded poor performance for both A1-P0 and A1-P1 (e.g., /u/).

The NAF method has been shown to be more robust across speakers than A1-P0 or A1-P1. The range of correlation coefficients across the six speakers was between 0.86 and 0.94, indicating again that even the weakest correlation with the ground-truth degree of nasalization is still very strong. In comparison, the range of correlation coefficients was between 0.42 and 0.85 for formant-compensated A1-P0 and between 0.06 and 0.64 for formant-compensated A1-P1. These results indicate again that the NAF method is not only more accurate than these conventional methods in an overall sense but also more consistently accurate across different speakers, even those that yielded poor performance for both A1-P0 and A1-P1 (e.g., speaker S2).

The NAF method has been shown to capture the temporally changing profile of the degree of nasalization more accurately than A1-P0 or A1-P1. According to the intercepts of the Bayesian regression model created to test for differences in the time-lag between the point of maximum velocity in the ground-truth nasalance signal and the point of maximum velocity in the experimental methods, the average time-lag was +10.5 ms for the full NAF method, and the 95% CI included 0 (the "true" onset of nasalization). In comparison, the average time-lag was +24.3 ms for formant-compensated A1-P0 and +24.2 ms for formant-compensated A1-P1, and the respective 95% CIs did not include 0. Moreover, according to the generalized additive mixed model created to test for differences in the temporal profile of the ground-truth degree of nasalization, the NAF method, and A1-P0, there were no significant differences between the ground-truth and the full NAF method at any point in the temporal interval between 250 ms before and 250 ms after the temporal onset of nasalization. In comparison, formant-compensated A1-P0 over-estimated the degree of nasalization throughout the oral portion of the signal, but under-estimated the degree of nasalization throughout the nasal portion of the signal.

B. Methodological considerations for implementing NAF

Although the methodology involved in assessing the NAF method may have been somewhat complicated up to this point in the paper, the methodology involved in implementing the NAF method in a research project is considerably more simple: all that is needed is a setup for audio recording, a carefully constructed experimental corpus, and PCA regression of acoustic features.

All of the data used in this paper come from oral and nasalized productions of vowels. As such, there are no nasalance values (or corresponding acoustic measurements) for nasal consonants in either the training or testing items. The observed results are therefore representative of productions that are expected for vowels produced with a wide range of velum height, but no oral blockage, such as those found in naturalistic productions of oral vowels, nasalized vowels, and nasal vowels. Additionally, the range of nasalance values used here for the oral tokens (i.e. the lower 25% of nasalance values) and the range of nasalance values used here for the nasalized tokens (i.e. the upper 25% of nasalance values) are reasonable approximations of the ranges that may be encountered in naturalistic speech data for tokens produced with a raised velum and a lowered velum, respectively. Nevertheless, the data that have been used here to validate the NAF method come from a task that is physiological and arguably a-linguistic in nature, yet the objective of this paper is to present a methodological tool as a research technique to use in an actual linguistic context. The method should therefore be able to scale beyond contrived tasks such as controlled velum lowering.

I propose that the NAF method can be implemented into any experimental design simply by including NVN and CVC words as filler items in the study corpus. The velum is expected to be low during the V of NVN words due to nasal coarticulation, whereas the velum is expected to be high during the V of CVC words, especially if the surrounding Cs are both voiceless obstruents. These filler items can subsequently be used for speaker-specific model training, and they should therefore be balanced for vowel quality to match the experimental items the researcher would like to use for model testing.

An example of a (small) complete item list for implementing the NAF method is shown in Table III. In this example, the hypothetical researcher is interested in examining possible differences in the degree of anticipatory vowel nasalization in /Vnt/ vs. /Vnd/ sequences of American English, since it has been argued that vowel nasality has become phonologized to some extent in the former but not the latter context (Beddor, 2009; Mielke *et al.*, 2017; Solé, 2007; Zellou, 2017). For each of the vowel qualities to be tested in the /Vnd/-/Vnt/ pairs, a matching oral-nasal pair of filler items is included. In order to implement the NAF method, the acoustic features measured in all vowels in the data set (i.e. both training and testing items) are first submitted to a single PCA model for each speaker. Subsequently, the PC scores for the training items are used as the independent variables in a linear regression model with a vector of numeric values corresponding to 0 (for the oral items) and 1 (for the nasal items) as the dependent variable. It should be reiterated that the model family/fit should be linear, *not binomial*; as such, the coding of 0 and 1 should be implemented as a *numeric variable* rather than a categorical factor. This speaker-specific linear model can then be used to predict values for given time points of interest in the testing items, using the PC scores associated with those time points as predictor variables.

Vowel	Training		Testing	
	Oral	Nasal	/Vnd/	/Vnt/
/i/	*peep*	*meme*	*fiend*	**feent*
/ɪ/	*pit*	*min*	*wind*	*hint*
/æ/	*pat*	*man*	*band*	*pant*
/ɑ/	*pop*	*mom*	*pond*	*font*
/u/	*toot*	*noon*	*wound*	**woont*

TABLE III. An example of an item list used for implementation of the NAF method in research on the degree of anticipatory vowel nasalization in American English.

In order to test the validity of this proposed implementation of the NAF method, the word list in Table III was produced by the author while collecting nasometry data. Words appeared in the carrier phrase "Say *X* again", where *X* is the target word. 10 randomized blocks of the word list were produced, resulting in a total of 100 training items (50 oral, 50 nasal) and 100 testing items (50 /Vnd/, 50 /Vnt/). Vowel intervals were segmented manually in Praat using the broadband spectrogram of the combined nasalance audio data. Acoustic and nasalance measurements were made at 10 equidistant time points in each vowel, from 0% to 100% of the vowel interval. The accuracy of the full NAF method and formant-compensated A1-P0 was assessed by constructing a GAMM with six factor levels: a separate level for each combination of method (nasalance, NAF, A1-P0) and phonetic context (/Vnd/, /Vnt/). Random factor smooths were included for both vowel and item repetition over time, as a function of the method used. The model fit was assessed using the `gam.check()` function, and the k-index for all terms was ≥ 1 with a large p-value.

Figure 7 displays the SE ribbons of the category fits for the GAMM. For the sake of legibility, the category fits are separated into three plots corresponding to nasalance (left), the full NAF method (center), and formant-compensated A1-P0 (right). Inter-method comparisons can be made by examining whether, at any given point in the vowel interval, the SE ribbons for any two methods overlap along their respective y-axes. With regard to nasalance, the degree of nasalization begins rising from the very start of the vowel, reaching a plateau at $\approx 40\%$ of the vowel interval, and it exhibits an increase at the end of the vowel (i.e. near the nasal consonant). There are no significant differences between the /Vnd/ and /Vnt/ contexts. The signal generated by the full NAF method exhibits this same general profile except for the increase at the end of the vowel, which is not observed. However, in contrasting the SE bands between the two plots, there are no areas of significant difference between the nasalance profile and the NAF profile. In comparison, the signal generated by formant-compensated A1-P0 yields poor correspondence with the ground-truth nasalance, failing to capture the changing degree of nasalization at any point in the vowel.

C. Interpreting the NAF approach

The NAF method is a "brute force" or "shotgun" approach to estimating the degree of nasalization through speaker-specific machine learning of many different acoustic metrics, some of which may not independently correlate with nasality at all. This has two impor-

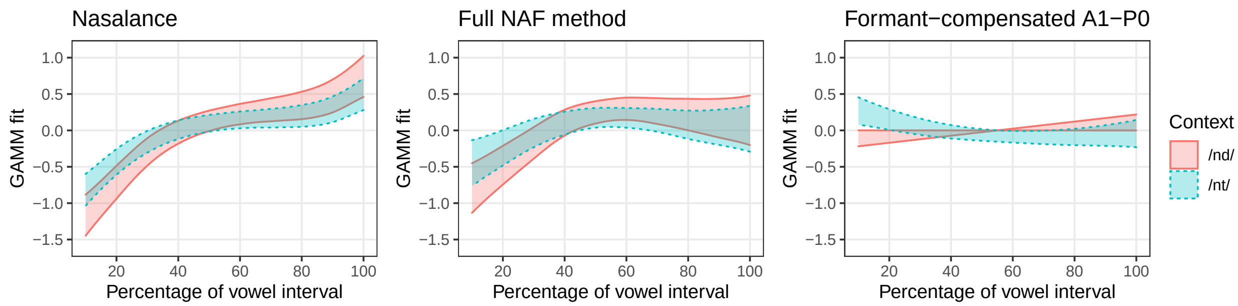

FIG. 7. (color online) GAMM fits for an example of an experimental implementation of the NAF method.

tant implications for interpretation of both the method and the NAF signal itself. The first implication is that the PCA transformation results in linear combinations of the acoustic features that are expressed as orthogonal dimensions; some of these dimensions will contribute strongly to generating the NAF signal while others will not. Dimensions that contribute strongly are interpreted as effectively capturing some aspect of vowel nasality; these components will yield large estimates in the regression model and will thus be important in generating the NAF signal. Conversely, acoustic features that do not correlate with nasality will be relegated to components that yield small estimates in the regression model, and they will therefore have little effect on the NAF signal. This means that *any set of acoustic features* can potentially be used in the NAF approach: the more acoustic features that are used, the better the chance of capturing underlying components that correlate with nasality. The second implication is that the PCA transformation of a large acoustic feature set means that the resulting NAF signal cannot be interpreted in any phonetically meaningful way other than as being a useful correlate of the degree of vowel nasalization. The goal of the NAF method is therefore not to contribute to our understanding of nasal acoustics but to provide researchers with a methodological tool to estimate the degree of vowel nasalization without the use of special instrumentation.

V. CONCLUSION

The NAF (Nasalization from Acoustic Features) method is a simple and easy-to-use approach to creating a time-varying signal of the degree of vowel nasalization in cases where articulatory measurements of nasalization (e.g., nasometry or nasal airflow) are not available to the researcher. The NAF method produces results that are significantly more accurate than A1-P0 and A1-P1, metrics that are widely used as acoustic correlates of nasalization. In comparison to these conventional metrics, the NAF method produces more reliable estimates of both the magnitude and the time-course of nasalization, and it produces estimates of nasalization that are far more robust across different vowel qualities and across different speakers. I therefore propose NAF as a methodological substitute for these traditional metrics in estimating the degree of vowel nasalization from acoustics alone, estimates which are accurate in magnitude and over the time-course of changing nasalization, and which are robust across different vowels, speakers, and languages.

ACKNOWLEDGEMENTS

The author would like to thank Pam Beddor and Georgia Zellou for comments on an early draft of this paper, as well as (virtual) attendees of the 17th Conference on Laboratory Phonology, where an implementation of the NAF method was presented in collaboration with Ander Egurtzegi. The author is also grateful to the three anonymous reviewers for their insight and suggestions.

[1]Available at: https://github.com/stylerw/styler_praat_scripts/tree/master/nasality_automeasure. Details about the script and its implementations of the acoustic measures of nasality can be found in Styler (2015).

[2]See Marx and Smith (1990) for background on PCA regression.

[3]The use of neural networks to map the acoustic features to the degree of nasalization was also explored, using the Keras R interface to the TensorFlow backend engine (Falbel *et al.*, 2020). However, the results were no better (and sometimes worse) than the linear PCA regression used in this paper.

Beddor, P. S. (**2009**). “A coarticulatory path to sound change,” Language **85**(4), 785–821, doi: `10.1353/lan.0.0165`.

Boersma, P., and Weenink, D. (**2017**). “Praat: doing phonetics by computer” Computer software program available from http://www.praat.org/.

Bürkner, P.-C. (**2020**). *brms: Bayesian Regression Models using ‘Stan’*, Computer software program available from https://cran.r-project.org/package=brms.

Carignan, C. (**2018**). “Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels,” Journal of the Acoustical Society of America **143**(5), 2588–2601, doi: `10.1121/1.5034760`.

Chen, M. Y. (**1997**). “Acoustic correlates of English and French nasalized vowels,” Journal of the Acoustical Society of America **102**, 2360–2370, doi: `10.1121/1.419620`.

Cho, T., Kim, D., and Kim, S. (**2017**). "Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English," Journal of Phonetics **64**, 71–89.

Dow, M. (**2020**). "A phonetic-phonological study of vowel height and nasal coarticulation in French," Journal of French Language Studies 1–36, doi: 10.1017/S0959269520000083.

Falbel, D., Allaire, J. J., Chollet, F., RStudio, Google, Tang, Y., Bijl, W. V. D., Studer, M., and Keydana, S. (**2020**). *keras: R Interface to 'Keras'*, Computer software program available from https://cran.r-project.org/package=keras.

Feng, G., and Castelli, E. (**1996**). "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization," Journal of the Acoustical Society of America **99**(6), 3694–3706.

Fujimura, O., and Lindqvist, J. (**1971**). "Sweep-tone measurements of vocal-tract characteristics," Journal of the Acoustical Society of America **49**, 541–558.

Garellek, M., Ritchart, A., and Kuang, J. (**2016**). "Breathy voice during nasality: A cross-linguistic study," Journal of Phonetics **59**, 110–121, doi: 10.1016/j.wocn.2016.09.001.

Hawkins, S., and Stevens, K. N. (**1985**). "Acoustic and perceptual correlates of the non nasal-nasal distinction for vowels," Journal of the Acoustical Society of America **77**, 1560–1575.

Jang, J., Kim, S., and Cho, T. (**2018**). "Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences," Journal of the Acoustical Society of America **144**(1), EL33–EL39, doi: 10.1121/1.5044641.

Khattab, G., Al-Tamimi, J., and Alsiraih, W. (**2018**). “Nasalisation in the production of Iraqi Arabic pharyngeals,” **75**(4), 310–348.

Kim, D., and Kim, S. (**2019**). “Coarticulatory vowel nasalization in American English: Data of individual differences in acoustic realization of vowel nasalization as a function of prosodic prominence and boundary,” Data In Brief **27**, 1–9.

Kroos, C. (**1996**). “Eingipflige und zweigipflige Vokale des Deutschen? Kinematische Analyse der Gespanntheitsopposition im Standarddeutschen,” Masters thesis, LMU Munich.

Liu, L., Hulden, M., and Scarborough, R. (**2019**). “RNN Classification of English Vowels: Nasalized or Not,” in *Proceedings of the Society for Computation in Linguistics: Vol. 2, Article 36*, doi: https://doi.org/10.7275/cnem-ab98.

Maeda, S. (**1993**). “Acoustics of vowel nasalization and articulatory shifts in French nasal vowels,” in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow, **5** of *Phonetics and Phonology* (Academic Press, San Diego), pp. 147–170, doi: 10.1016/B978-0-12-360380-7.50010-7.

Marx, B., and Smith, E. P. (**1990**). “Principal component estimation for generalized linear regression,” Biometrika **77**(1), 23–31, doi: https://doi.org/10.1093/biomet/77.1.23.

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,” Journal of the Acoustical Society of America **142**(1), 332–349.

Pruthi, T., and Espy-Wilson, C. Y. (**2004**). “Acoustic parameters for automatic detection of nasal manner,” Speech Communication **43**, 225–239.

R Core Team (**2020**). "R: A Language and Environment for Statistical Computing" Computer software program available from http://www.R-project.org.

Scarborough, R. (**2013**). "Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation," Journal of Phonetics **41**, 391–508.

Scarborough, R., and Zellou, G. (**2013**). "Clarity in communication: 'Clear' speech authenticity and lexical neighborhood density effects in speech production and perception," Journal of the Acoustical Society of America **134**, 3793–3807.

Solé, M.-J. (**2007**). "Controlled and mechanical properties in speech: a review of the literature," in *Experimental Approaches to Phonology*, edited by M.-J. Solé, P. Beddor, and M. Ohala (Oxford University Press, Oxford), pp. 302–321.

Stan Development Team (**2017**). "Stan: A C++ library for probability and sampling, version 2.14.0" http://mc-stan.org/.

Stevens, K. N., Fant, G., and Hawkins, S. (**1987**). "Some acoustical and perceptual characteristics of nasal vowels," in *In Honour of Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht), pp. 241–254.

Styler, W. (**2015**). "On the acoustical and perceptual features of vowel nasality," Ph.D. thesis, University of Colorado.

Styler, W. (**2017**). "On the acoustical features of vowel nasality in English and French," Journal of the Acoustical Society of America **142**(4), 2469–2482, doi: 10.1121/1.5008854.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (**2018**). "Bayesian data analysis in the phonetic sciences: A tutorial introduction," Journal of Phonetics **71**, 141–161.

Wieling, M. (**2018**). “Analyzing dynamic phonetic data using generalized additive mixed-modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English,” Journal of Phonetics **70**, 86–116.

Wood, S. N. (**2019**). “mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation” Computer software program available from https://cran.r-project.org/package=mgcv.

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

Zellou, G., and Scarborough, R. (**2012**). “Nasal coarticulation and contrastive stress,” in *Proceedings from INTERSPEECH 2012*, pp. 2686–2689.

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