The effect of vocal tract parameters on aspiration noise discrimination

Ilse B. Labuschagne, and Valter Ciocca

Citation: The Journal of the Acoustical Society of America **147**, 1239 (2020); doi: 10.1121/10.0000756
View online: https://doi.org/10.1121/10.0000756
View Table of Contents: https://asa.scitation.org/toc/jas/147/2
Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

A model of speech production based on the acoustic relativity of the vocal tract
The Journal of the Acoustical Society of America **146**, 2522 (2019); https://doi.org/10.1121/1.5127756

Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape
The Journal of the Acoustical Society of America **146**, 4190 (2019); https://doi.org/10.1121/1.5134784

Comparison of volitional opposing and following responses across speakers with different vocal histories
The Journal of the Acoustical Society of America **146**, 4244 (2019); https://doi.org/10.1121/1.5134769

Interaural recalibration of phonetic categories
The Journal of the Acoustical Society of America **147**, EL164 (2020); https://doi.org/10.1121/10.0000735

Exploration of excitation source information for shouted and normal speech classification
The Journal of the Acoustical Society of America **147**, 1250 (2020); https://doi.org/10.1121/10.0000757

Editorial—JASA 2020
The Journal of the Acoustical Society of America **147**, 1262 (2020); https://doi.org/10.1121/10.0000786

JASA ARTICLE

The effect of vocal tract parameters on aspiration noise discrimination[a)]

Ilse B. Labuschagne[b)] and Valter Ciocca
School of Audiology and Speech Sciences, The University of British Columbia, 2177 Wesbrook Mall, Vancouver, British Columbia V6T 1Z3, Canada

ABSTRACT:
Previous research showed that aspiration noise difference limens in moderately breathy /a/ vowels decreased as the spectral slope of the glottal source spectrum became increasingly steep [Kreiman and Gerratt, J. Acoust. Soc. Am. **131**(1), 492–500 (2012)]. The current study investigated whether discrimination of aspiration noise levels was affected by differences in spectral shape due to vowel quality (/æ/ and /i/) and speaker identity (three male speakers) when the slope of the glottal source spectrum was fixed. The results showed that discrimination performance was worse overall for /i/ than /æ/, but the result may have resulted from relatively poor performance for the /i/ vowel of one speaker. Acoustic analyses of the stimuli were performed to estimate the association between acoustic properties and the perceptual outcomes. The results showed that both the smoothed cepstral peak prominence and the harmonic energy level between 2 and 5 kHz may account for the observed differences in aspiration noise discrimination among speakers within each vowel, but not for differences between vowel categories. It is possible that the relationship between the aspiration noise discrimination and aforementioned acoustic properties may be modulated by the spectral distribution of energy across frequency. © *2020 Acoustical Society of America*.
https://doi.org/10.1121/10.0000756
(Received 29 August 2019; revised 30 January 2020; accepted 30 January 2020; published online 25 February 2020)
[Editor: Jody Kreiman] Pages: 1239–1249

I. INTRODUCTION

Breathiness is a perceptual quality that can be used to discriminate between voices (Kreiman *et al.*, 1994; Kreiman, *et al.*, 1990), to ascertain possible voice pathology (in cases of severe breathiness) (Fukazawa *et al.*, 1988; Hartl *et al.*, 2003), or to attribute meaning in languages that contrast breathy and non-breathy vowels (Ladefoged, 1982). Breathiness occurs during phonation when either the vocal folds never completely close or when the vocal folds are open for a relatively long time compared to the period of time that the vocal folds are closed. These patterns of vocal fold vibration lead to an increased rate of turbulent airflow and are perceived as a noisy quality of the speech output. The perception of breathiness has often been found to correlate with the presence of noise in the acoustic speech signal. For example, Klatt and Klatt (1990) showed that adding noise to synthesized vowels increased listeners' perception of breathiness. Similarly, Samlan *et al.* (2013) found that the harmonics-to-noise ratios (HNRs) of synthesized stimuli correlated well with ratings of breathiness ($r = -0.884$). Considering the salience of noise as a cue for the perception of breathiness, relatively few studies have investigated the noise difference limens (DLs) in vowels, that is, an estimate of the smallest detectable difference in noise levels between two vowels. While previous studies focused on the effects of changing glottal source parameters on noise discrimination, the current study aimed to improve our understanding of the effects of vocal tract parameters on noise discrimination in vowel stimuli.

Shrivastav and Sapienza (2006) investigated aspiration noise DLs for five-formant /a/ vowels that were synthesized to model utterances from three male and three female speakers. Vowels were synthesized with different levels of noise-to-harmonics ratios (NHRs). Results indicated that aspiration noise DLs were systematically affected by NHRs: At −30, −20, and −10 dB NHR, average DLs were found to be about 21, 14, and 11 dB, respectively. Additionally, they found significant differences in aspiration noise DLs for utterances produced by different speakers. Because their stimuli were synthesized with co-varying glottal and vocal tract parameters, Shrivastav and Sapienza were unable to attribute the differences in aspiration noise DLs to differences in specific synthesis parameters. However, acoustic analyses of their stimuli suggested that higher amounts of noise energy between 2 and 5 kHz might be associated with lower aspiration noise DLs. In a subsequent study, Kreiman and Gerratt (2012) varied the steepness of the spectral slope of the glottal waveform of two five-formant /a/ vowels (one male and one female) in 3-dB/octave increments, from −3 to −12 dB/octave. NHR was also varied from −10 to −40 dB NHR in 10-dB increments. Like Shrivastav and Sapienza (2006), they found that aspiration noise DLs were smaller for stimuli with larger NHRs. However, the aspiration noise DLs reported by Kreiman and Gerratt were about 2 to 3 times smaller than those obtained by Shrivastav and

[a)]Portions of this work were presented in "Discrimination of aspiration noise in breathy vowels," *Proceedings of Acoustics Week in Canada*, Vancouver, Canada, September 2016.
[b)]Electronic mail: ilse.labuschagne@alumni.ubc.ca

Sapienza. In the stimuli used by Shrivastav and Sapienza, the spectrum of the glottal noise source generated by the Klatt synthesizer was flat. Kreiman and Gerratt hypothesized that differences in the spectral shape of the noise source after it was filtered by the vocal tract might be responsible for the discrepancy in the size of DLs between the two studies. They found that the spectral shape of the aspiration noise affected aspiration noise DLs, such that flat noise spectra resulted in smaller DLs than falling noise spectra. Therefore, Kreiman and Gerratt concluded that this effect could not explain the large DL differences between the two studies. One aim of the current study was to obtain further evidence about the size of aspiration noise DLs, by way of the sensitivity measure d' that measures how well a listener can discriminate between two vowel sounds with a fixed difference in aspiration noise levels. The synthesizer and noise source that was used by Shrivastav and Sapienza was used in the current study, and the glottal source parameters were controlled across the stimuli.

Kreiman and Gerratt also showed that glottal waveforms (that is, the harmonic energy content) with steeper spectral slopes resulted in smaller aspiration noise DLs. They observed that stimuli with steeper glottal spectra have relatively little harmonic energy at higher frequencies and that, for these stimuli, high-frequency harmonic energy might not mask aspiration noise energy as effectively as for stimuli with relatively flat glottal spectra. Figure 1 displays the effect of glottal slope on aspiration noise DLs using six-formant vowel stimuli created with the MATLAB software (The Mathworks, Inc., 2017). This figure shows separately the spectra of the harmonic energy (grey) and of the noise energy (black) for the glottal waveforms and the output spectra of the vowels /æ/ and /i/. Panels (a) and (b) show the spectra for glottal waveforms with a spectral slope of −3 and −12 dB/octave, respectively. Panels (c) and (d) display the spectra of the /æ/ vowels synthesized with glottal slopes of −3 and −12 dB/octave, respectively; similarly, panels (e) and (f) show the spectra of the /i/ vowels obtained with the two glottal slopes. The NHR was −30 dB NHR for both vowel qualities. The aspiration noise that is generated at the glottis was modeled as additive Gaussian white noise (flat spectrum across all frequencies). For both vowels, above about 2 kHz the level of the noise compared to the harmonic energy is much higher (greater NHR) for the vowels in panels (d) and (f) than for the vowels in panels (c) and (e). Therefore, it is likely that high frequency harmonics in vowels synthesized with a steeper glottal slope [panels (d) and (f)] do not mask noise energy at these frequencies as effectively as in the stimuli in panels (c) and (e). Lower masking of high-frequency noise possibly resulted in lower aspiration noise DLs. In summary, the results of both Kreiman and Gerratt (2012) and Shrivastav and Sapienza (2006) showed that aspiration noise DLs were affected by the level of the aspiration noise relative to the harmonic energy across all frequencies, and by the level of aspiration noise relative to the harmonic energy within frequency bands. Shrivastav and Sapienza (2006) also provided evidence that aspiration noise DLs can differ substantially among stimuli modelled after different speakers.

While both Kreiman and Gerratt (2012) and Shrivastav and Sapienza (2006) measured aspiration noise DLs with varying characteristics of the glottal spectrum, the current study focused on determining the discrimination of aspiration noise levels for different vocal tract parameters. Discrimination differences among speakers were investigated for different vowels, /æ/ and /i/, while controlling the glottal source parameters. We hypothesized that discrimination of aspiration noise would differ among speakers and between vowels, because differences in formant frequencies for each speaker-vowel combination result in unique distributions of harmonic and noise energy, which in turn might affect aspiration noise discrimination. Such differences might be observed even though the NHR in different frequency bands is virtually the same across stimuli. A study by Gockel *et al.* (2002) provided evidence for this hypothesis. They measured absolute noise thresholds in harmonic series maskers whose components had been added in either cosine or random phase. They varied the harmonic masker level [40 to 70 dB sound pressure level (SPL)]. The harmonic series co-varied in fundamental frequency (F0; 62.5 or 250 Hz) and frequency range (625–5000 Hz for the 62.5-Hz F0; 2500–5000 Hz for the 250-Hz F0). They found that absolute noise thresholds relative to the masker level decreased as masker level increased only for the 62.5-Hz harmonic series with components in cosine phase. This finding showed that, at the lowest F0, noise thresholds relative to the masker level were not constant across harmonic masker levels. Whether noise discrimination thresholds in vowels are also affected by the level of harmonic masker energy is an empirical question that was investigated by the present study. An acoustic analysis of the experimental stimuli was also performed to investigate whether aspiration noise discrimination is affected by the energy level in the 2–5 kHz range. Since stimuli were synthesized with equal NHR across frequency bands, any observed differences in the energy level between 2 and 5 kHz among stimuli were expected to result from differences in vocal tract properties (that is, formant frequency parameters). In addition to measuring the energy levels in different frequency bands, the smoothed cepstral peak prominence (CPPS) was calculated for the experimental stimuli, following the procedure described by Hillenbrand and Houde (1996). CPPS is an index of the periodicity of a waveform; this measure (or the related cepstral peak prominence) had previously been found to negatively correlate with breathiness (Hillenbrand and Houde, 1996; Samlan *et al.*, 2013; Shrivastav and Sapienza, 2003). Finally, the amplitude difference between the first harmonic and the second harmonic (H1-H2) was calculated for all the stimuli. This measure reflects glottal pulse characteristics associated with the production of breathy voices such as decreases in the abruptness of vocal fold closure and increases of the open quotient [Gauffin and Sundberg (1989), Hanson (1997), Klatt and Klatt (1990), however, see Hartl *et al.* (2003) and Samlan *et al.* (2013) for

https://doi.org/10.1121/10.0000756

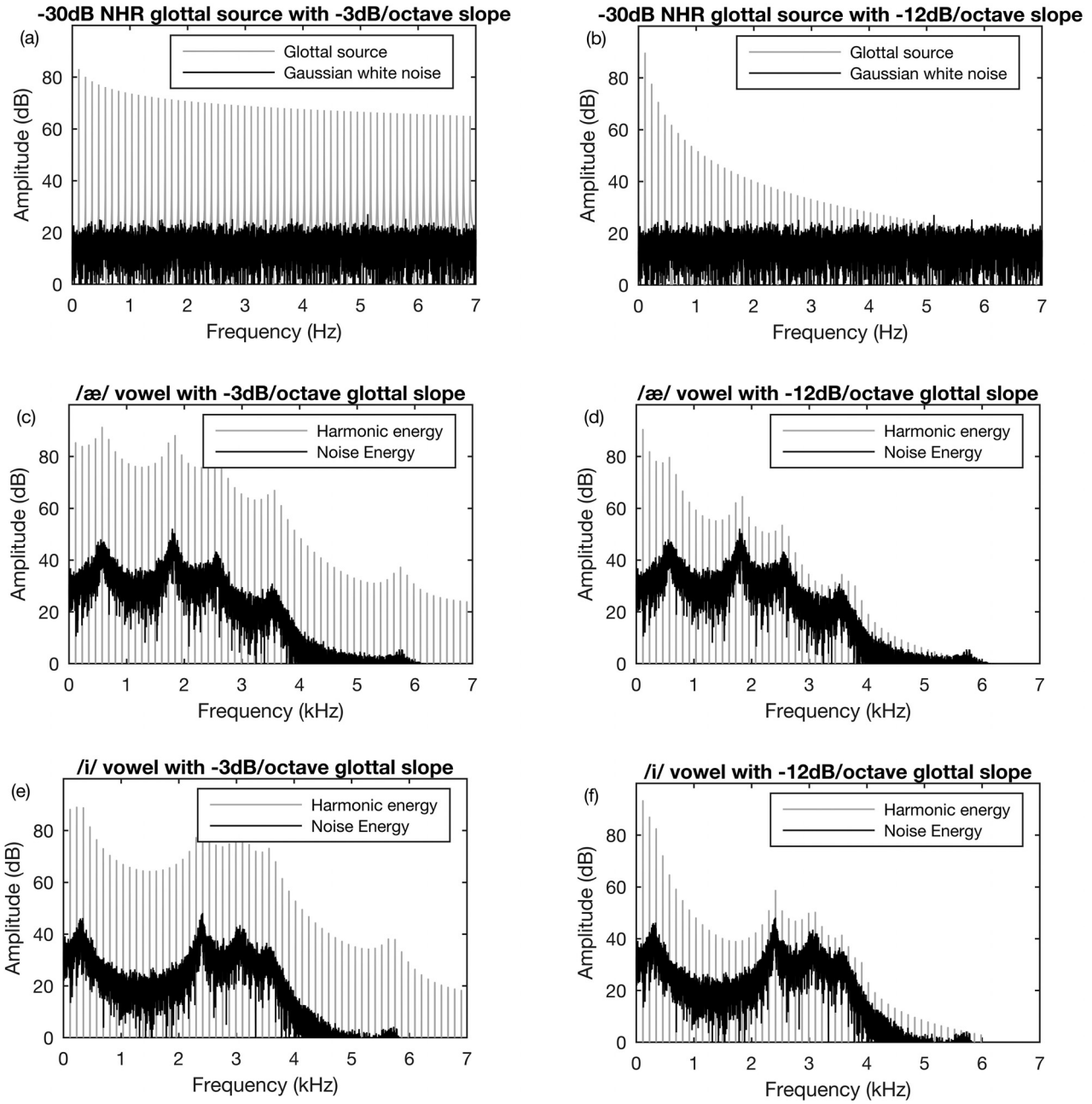

FIG. 1. This figure shows /æ/ and /i/ vowels that were synthesized with different glottal spectra. In each figure, the harmonic and inharmonic energy are shown separately in grey and in black, respectively. The first row shows −3 dB/octave [left—panel (a)] and −12 dB/octave [right—panel (b)] harmonic glottal spectra with aspiration noise generated at the glottis. The second and third rows show the spectra of vowels synthesized with the glottal spectra displayed in the first row. The second row shows spectra for /æ/ vowels [panels (c) and (d)]; /i/ vowels are displayed in the third row [panels (e) and (f)]. Within each column, the NHR across all frequencies are the same. Across rows, the absolute amount of harmonic and noise energy differ.

contrasting findings]. Since glottal parameters were controlled, any observed differences between H1-H2 were expected to result from differences in vocal tract properties.

II. METHOD

A. Participants

Eighteen participants (13 females and 5 males) between the ages of 21 and 40 (average age of 26) participated in the study. Participants were screened to ensure that their hearing thresholds were 15 dB hearing level lower at octave frequencies from 250 to 4000 Hz. The hearing screening took place inside a double-walled sound-attenuated booth, using a Maico MA-39 air-conduction audiometer. Participants were compensated for their participation in the study. The study was approved by the Behavioural Research Ethics Board of the University of British Columbia.

B. Stimuli

The stimuli were based on utterances extracted from the Hillenbrand database of vowel recordings (Hillenbrand, 2003;

Hillenbrand *et al.*, 1995). This database consists of a collection of vowel recordings in /h-vowel-d/ format, accompanied by vowel duration information and the center frequencies of the first three formants at eight time-points (10% to 80% of the vowel duration, in 10% increments). The vowels /æ/ and /i/ from three speakers (S08, S30, and S44) were selected because they were perceived by the authors as having relatively neutral voice qualities (that is, not particularly breathy, nasal, or rough), and because they had similar duration. The duration of the /æ/ vowels ranged from 289 to 302 ms, with an average duration of 296 ms; the duration of the /i/ vowels ranged from 276 to 291 ms, with an average duration of 283 ms. In order to resynthesize these vowels, the fundamental frequency (F0), the voicing amplitude envelope, and the fourth to sixth formant frequencies were estimated at the same eight time-points that were used for formant estimates in the Hillenbrand database. These estimates were calculated using the PRAAT software (Boersma and Weenink, 2017). These parameters were used, along with the first three formant values provided in the Hillenbrand database, to synthesize new vowels using the parallel branch of the Klatt synthesizer (Klatt and Klatt, 1990). The values of the voicing amplitude envelope of each utterance were rescaled to an average intensity of 70 dB. After the initial synthesis of a vowel, the formant locations and bandwidths, and formant levels were adjusted to achieve a good match between the spectra of the original and synthesized stimuli. After each spectral adjustment, the authors perceptually compared synthesized and original vowels to ensure that the synthesized vowels sounded both natural, and increasingly similar in both vowel and voice quality to the original recordings. This process was repeated until both authors were satisfied that spectral and perceptual characteristics of the synthesized vowels closely matched those of the original vowels. An example of the spectral matches between original and resynthesized vowels are shown in Fig. 2 for vowels /æ/ and /i/ of speaker S44.

Each panel of this figure shows the spectrum of a 100-ms center segment of the original vowel recording (black) and the spectrum of a 100-ms center segment of the synthesized vowels (grey) after the completion of the spectral-perceptual matching process. An online supplement[1] (Sec. 1) lists the values of the duration, the F0, the formants (frequency, bandwidth, and level), and the amplitude envelope used for each of the synthesized vowel stimuli. All stimuli were synthesized with glottal pulse waveforms represented by the equation $flow(t) = t^3 - t^4$, for $0 <= t <= 1$; the open quotient was set to 0.4. Except for the level of aspiration noise, all other glottal pulse parameters (flutter—also referred to as jitter, collision phase, double pulsing, spectral tilt, and breathiness) had a value of 0 (the default setting within the PRAAT software). Although default values of zero did not necessarily reflect natural speech—for example, voices naturally have a non-zero amount of jitter—the synthesized vowel tokens still sounded natural. Furthermore, using the default values of zero prevented these parameters from influencing the listener's perception of breathiness or other perceptual voice qualities like roughness [for example,

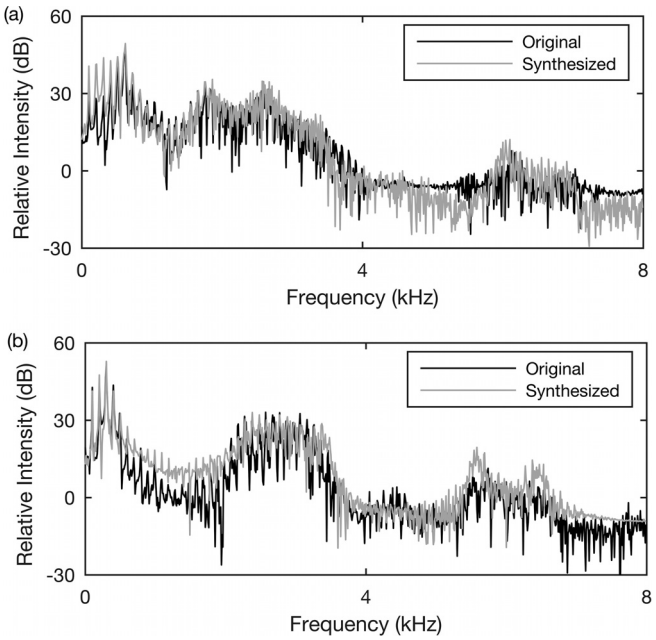

FIG. 2. Amplitude spectra of 100-ms segments from the centre of the synthesized vowels (grey lines) and of the original vowels (black lines) of speaker S44. Panel (a) shows the /æ/ vowels; panel (b) shows the /i/ vowels.

Shrivastav and Sapienza (2003) found a high correlation of 0.863 between the % of jitter and perceptual ratings of breathiness]. Each of the six stimuli (three speakers and two vowels) was synthesized with three levels of aspiration noise: 35, 37, and 39 dB. All of these aspiration noise values resulted in natural-sounding vowels. The 35-dB level resulted in a small, yet natural amount of breathiness and was used as the reference aspiration noise level. Increments of 2 dB were used on the basis of pilot studies that showed that aspiration noise DLs were likely to be approximately 2 to 4 dB for these stimuli. Each of the synthesized vowels was multiplied by an amplitude envelope with linear rise and decay time of 10 ms. Stimuli were synthesized with a sampling rate of 44.1 kHz and 16-bit resolution.

A separate, smaller set of synthesized training stimuli was used to familiarize the participants with the experimental procedure and with differences in aspiration noise levels (representing different levels of breathiness). This set of synthesized vowels was modelled after the /æ/ and /i/ vowels of speakers S01 and S02 from the Hillenbrand database. The same synthesis procedure described previously was used, but using 0, 33, and 38-dB aspiration noise levels. These levels simulated low, moderate and high levels of breathiness, respectively. The training set of stimuli consisted of 12 utterances (two vowels × two speakers × three aspiration noise levels).

C. Procedure

After the hearing screening, participants completed a two-part training session. During part one, the first author provided a verbal explanation of "breathiness." Following the explanation, participants listened to examples of low,

https://doi.org/10.1121/10.0000756

moderate and high breathiness (training set stimuli). Participants listened to each of the 12 training stimuli as many times as they wanted by clicking "play" buttons on a computer window using a mouse. Stimuli were presented binaurally through Oppo PM-3 headphones using a Grace Design m9xx digital-to-analog converter, at a peak-level of 70 dBA. In part two of the training session, participants performed a two-alternative forced-choice (2AFC) task. In each trial, listeners heard a sequence of two stimuli: one stimulus synthesized with 0-dB aspiration noise (the "standard"), and the same vowel synthesized with either 33- or 38-dB of aspiration noise ("comparison stimuli"). Each stimulus pair was presented twice (standard first or second), for a total of 16 pairs (two vowels × two speakers × two levels of aspiration noise differences × two presentation orders). Custom experimental software, written with the LiveCode package (LiveCode, 2016), controlled the presentation of the stimuli in randomized order. The software ran on an iMac desktop computer (Apple Inc.). During each trial, two stimuli were presented sequentially, with a 1-s inter-stimulus interval. Participants had to choose which of the two stimuli was "more breathy" by pressing a key on a computer keyboard ("1" for the first vowel; "9" for the second vowel). It is unlikely that listeners were able to use overall presentation level differences to discriminate between stimuli, because the largest presentation level difference between stimulus pairs was measured as 0.06 dB SPL. This difference is much smaller than the difference of about 1 dB SPL that is required for intensity discrimination with wideband stimuli presented at 70 dB SPL (Jesteadt *et al.*, 1977). Feedback was given after each response; after feedback, a screen with the text "Get ready…" was displayed for 2 s, and then the next trial started automatically. Participants that correctly selected the interval with the 33- or 38-dB aspiration noise at least 12 times out of 16 were allowed to continue to the experimental session. All participants achieved this criterion.

The 2AFC task in the experimental session was the same as the one used in the training session. During each trial, listeners heard a pair of stimuli that differed only in the level of aspiration noise. For each of the six experimental stimuli (two vowels by three speakers), the standard (35-dB aspiration noise) was paired with either the 37- or the 39-dB aspiration noise comparison stimuli. These combinations resulted in differences of +2 or +4 dB in aspiration noise level. Each stimulus pair was presented in two orders (standard first followed by comparison, or vice versa) within each block. Stimulus presentation was blocked by vowel; there was a short break between each block. The order of presentation of the vowels was counterbalanced across participants. For each vowel, 12 pairs of stimuli (three speakers × two levels of aspiration noise differences × two presentation orders) were repeated ten times for a total of 120 trials per vowel. The hearing screening, training and experimental sessions were completed within a single hour, and took place in a double-walled sound attenuating booth.

D. Acoustic analysis

The acoustic analysis comprised three acoustic correlates of breathiness: NHR, CPPS, and H1-H2. Previous research hypothesized that aspiration noise DLs depended on the NHR above 2 kHz (Kreiman and Gerratt, 2012; Shrivastav and Sapienza, 2006). In the current study, NHR was calculated for the following frequency bands: 20–400 Hz (band 0, or B0); 400–2 kHz (band 1; B1); 2–5 kHz (band 2; B2); 5–10 kHz (band 3; B3). These frequency bands were the same as those used by Samlan *et al.* (2013). NHR was estimated with the PRAAT software for the standard of each vowel and each speaker, following the procedure used by Shrivastav and Sapienza (2006). This procedure involved the synthesis of two new sets of stimuli. The harmonic stimuli were synthesized with 0-dB aspiration noise and 70-dB voicing. A Fourier transform was performed on each of harmonic stimuli using the function "To spectrum." The resulting spectra were filtered with Hann passband filters using the function "Filter (pass Hann band)" with 20-Hz smoothing. This function band-pass filtered the sounds by multiplying an amplitude window function in the spectral domain with the spectrum of the sound. The band-pass filter window function had raised-cosine transitions, and resulted in a total transition width of 40 Hz on each boundary of the band-pass filter. Finally, the harmonic energy in each band was calculated using the function ("Get band energy"). The noise-only stimuli were synthesized with the aspiration noise level set to 35 dB, and the voicing amplitude set to −60 dB (to approximate no voicing). From these stimuli, the noise energy in each band (N0, N1, N2, N3) was calculated using the same procedure described above. NHR was then calculated for each band as the dB difference between the noise and the harmonic energy of the corresponding harmonic and noise-only stimuli. Because the glottal parameters were fixed, the NHR in each band was expected to be very similar across all stimuli. One of the aims of the study was to investigate if the amount of energy (noise and harmonic) in different bands affected aspiration noise discrimination. Therefore, in addition to the NHR within each band, the total energy in each band for each of the standard stimuli was calculated using the "Get band energy" function.

CPPS was calculated using the function "Get CPPS" with a 10-frame (20-ms) time-smoothing window, and a 10-bin (0.05 ms) quefrency smoothing window. PRAAT's implementation of Theil's robust regression (Theil, 1950) was used to fit a straight line to the smoothed cepstrum, after which cubic interpolation was used for cepstral peak detection. The CPPS was calculated for stimuli that were low-pass filtered with a cutoff frequency of 5 kHz.

To calculate H1-H2, each vowel stimulus was first segmented into ten segments of equal duration. Each segment was multiplied by a Hanning window in the temporal domain. H1-H2 was determined from the spectrum of each of the segments. The final H1-H2 value was calculated as the average of the H1-H2 values for the ten segments.

III. RESULTS

A. Perceptual results

Sensitivity scores (d') were calculated for every condition and listener. Sensitivity scores based on signal detection theory were used rather than defining a DL according to a proportion correct score or using an adaptive procedure that converge on a DL for three reasons. First, d' scores vary depending on the experimental procedure. Klein (2001) points out that DLs based on a proportion correct score would only be stable across experiments if the assumption of high threshold theory was supported by evidence, which is not the case. Second, d' scores are unaffected by listener response bias (Macmillan and Creelman, 2005, p. 7–9). Third, d' scores allowed appropriate comparisons to the DLs of previous studies that used different criteria and experimental procedures to obtain DLs (Macmillan and Creelman, 2005, p. 173). The following equation was used to calculate sensitivity scores:

$$d'_{2AFC} = [Z\{H\} - Z\{F\}]/\sqrt{2}, \quad (1)$$

where $Z\{H\}$ and $Z\{F\}$ are the z-transformations of the hit and the false-alarm rates, respectively (Macmillan and Creelman, 2005, p. 168). For each trial, a hit was recorded when the participant identified the first stimulus of the pair to be more breathy, and that stimulus had a higher NHR; a false-alarm was recorded when the participant identified the first stimulus to be more breathy and the second stimulus had a higher NHR. When a participant achieved a perfect score for any condition (i.e., a hit rate of 1 and/or a false alarm rate of 0), a correction rule was applied that adjusted the proportion of 1 down by 1/2n, and the proportion 0 up by 1/2n (with n = 10) (Miller, 1996). This correction was necessary, because the z-transformation of response proportions 0 and 1 lead to $-\infty$ and $+\infty$, respectively. The maximum and minimum values of d' were 2.33 and -2.33. Section 2 of an online supplement[1] shows the individual d' scores of all listeners for every condition. A commonly adopted empirical threshold of $d' = 1$, which corresponds to moderate sensitivity, was used as a criterion to estimate aspiration noise DLs [Klein (2001), Macmillan and Creelman (2005), p. 119—although note that Klein (2001) suggested that setting the d' criterion larger than 1 yielded more stable thresholds estimates]. For the constant stimulus method used in the current study, the $d' = 1$ threshold corresponded to a proportion of correct responses of 0.76 for a listener without a response bias (where the first stimulus in the trial pair were chosen in equal proportion as the second stimulus).

A Bayesian approach was used for the main statistical analysis of d' scores. The outcome of a Bayesian data analysis is a posterior distribution ("posterior," hereafter) composed of estimated probabilities of specific hypotheses or parameter values, rather than the point estimates of model parameters that are obtained through classical (frequentist) statistics (Dienes, 2011; Kruschke and Liddell, 2017; Van de Schoot *et al.*, 2014). As a result, Bayesian statistics is well equipped to deal with small sample sizes, because the results are valid for any sample size given the model and the observed data (Hox *et al.*, 2012; Lee and Song, 2004). Unlike classical statistics, Bayesian data analysis procedures incorporate prior knowledge (or lack of prior knowledge) as well as information about possible parameter values. The use of prior information for making scientific inferences is consistent with the progressive accumulation of knowledge within scientific disciplines (see, for example, Kruschke, 2010). A Bayesian multilevel model, analogous to a frequentist linear mixed-effect regression model, was fitted to the d' scores using the *brms* package (Bürkner, 2017) for the R statistical programming language (R Core Team, 2017). This package interfaces with *Stan*, a separate statistical programming language with Bayesian inference capability that uses a Markov Chain Monte Carlo (MCMC) sampler (Carpenter *et al.*, 2017). Participant-specific intercepts and slopes (also called "random" or "group-level" effects) were included in alternative models together with fixed effects (also referred to as "population-level" effects). The fixed effects were aspiration noise levels (+2 and +4 dB), vowel quality (/i/ and /æ/), and speaker (S08, S30, and S44). Although stimuli from different speakers are often included in models of perceptual studies as random intercepts and slopes [see, for example, Barr *et al.* (2013) and Sorensen *et al.* (2016)], they were included as fixed effects in the current study in order to directly estimate how the acoustic properties of speaker-specific utterances might account for any differences in d' scores.

A null, intercept-only model and twelve alternative models were considered; these models are listed in Sec. 3 of the online supplement.[1] Trends observed in the data informed decisions about whether or not to include different factors in alternative models (Barr *et al.*, 2013). Alternative models considered main effects of aspiration noise level, vowel quality and speaker, three two-way interactions, and the three-way interaction, and allowed for listener-specific slopes for each of the three fixed-effect factors. Prior distributions for each model parameter were specified on the basis of the results of a pilot study. Pilot results showed that the average d' score across all conditions was about 1, with a standard deviation (SD) of about 1. The prior distribution for the model intercept was therefore defined as a normal distribution with a mean of 1 and an SD of 1. Other priors were also investigated, but the resulting posteriors did not change substantially from the results presented below (see Sec. 6 of the online supplement[1] for more details). For the fixed effects coefficients, the prior distributions were defined as normal distributions with means of 0 and SDs of 1. Default prior distributions (half student's t distributions with 3 degrees of freedom, means of 0, and SDs of 2) were used for the SDs of the random effects, and for the correlation matrix for the random effects (Bürkner, 2017). This correlation matrix reflected the reasonable assumption that listeners who achieved higher d' scores overall would not necessarily obtain either lower or higher random slopes for any of the

https://doi.org/10.1121/10.0000756

fixed effects. To estimate the model parameters, four MCMC chains were obtained; each chain consisted of 20 000 iterations of which 4000 were burn-in iterations. All model parameters converged with a minimum effective number of samples (also known as effective sample size) of 10 000 or larger for each parameter. Because Bayesian analysis produces a posterior for all model parameters, it is possible to calculate the smallest interval that contains a specified percentage of the posterior draws for a parameter (highest density interval). This interval, hereafter referred to as uncertainty interval (UI), indicates that there is a specified amount of probability that the parameter has a value between the interval limits. The UIs reported in this section contain 90% of the probability mass for a given parameter, and their limits are enclosed within square brackets following their median values.

Models were compared using the widely applicable information criterion [or Watanabe-Akaike information criterion (WAIC)] (Gelman *et al.*, 2014; Vehtari *et al.*, 2017; Watanabe, 2010). This measure calculates the predictive accuracy of a statistical model by estimating the loss of information that would occur when the model is fit to new data. A model with a smaller WAIC score—the model with smaller information loss—has better predictive accuracy. New models were compared with the previous best model using the ΔWAIC statistic ($\Delta\text{WAIC} = \text{WAIC}_{\text{previous best model}} - \text{WAIC}_{\text{new model}}$) obtained from the *waic* function within the *brms* package. If ΔWAIC was positive and greater than the standard error of the difference between paired estimates from each model (SE), the new model was preferred. If ΔWAIC was either positive but smaller than the SE, or negative, the previous best model was retained. The ΔWAIC and SE values for each new model that was evaluated are listed in Sec. 3 of the online supplement.[1]

Of all the models considered, the model selected on the basis of the criteria specified above included main effects of the aspiration noise level, vowel, and speaker, interaction effects of vowel by aspiration noise level and vowel by speaker, unique listener intercepts, and unique listener slopes for noise level and vowel [dprime ∼ noise_level + vowel + speaker + noise_level:vowel + vowel:speaker + (1 + noise_level + vowel|listener)—see Sec. 4 of the online supplement for a summary of the model parameters[1]]. It is also worth noting that the other four models with the highest predictive accuracy (see models 7, 9, 10, and 12 in Sec. 3 of the online supplement[1]) all had the same fixed effects as the selected model (Sec. 7 of the online supplement[1] shows that the model that best fit the data using a frequentist analysis also contained these fixed effects). The median of the SD of the listener-specific intercepts ($\text{SD}_{\text{Intercepts}} = 0.54\,[0.36, 0.74]$) suggested that listeners differed in how well they performed overall. The relatively smaller medians of the SDs of the listener-specific slopes for aspiration noise level and vowel ($\text{SD}_{\text{noise level slopes}} = 0.24\,[0.07, 0.42]$ and $\text{SD}_{\text{vowel slopes}} = 0.26\,[0.04, 0.44]$, respectively) suggested that listeners had somewhat similar sensitivity differences when comparing one aspiration noise level to the other, and when comparing one vowel to another. There was no evidence that there were non-zero correlations between the listener-specific intercepts and listener-specific slopes.

From the selected model, comparisons among conditions were made by calculating the posteriors of the "strictly standardized mean difference" (SSMD) (Zhang, 2010). SSMD is an effect size measure that is analogous to calculating the effect size for contrasts in classical statistics. To calculate the posterior SSMD for a contrast, the *BEST* package for R (Meredith and Kruschke, 2018) was first used to estimate posterior probability distributions for the mean and the standard deviation of the paired differences of d' scores between the relevant conditions. One hundred thousand samples were obtained for each contrast. Then, the SSMD distribution was calculated by dividing the posterior distribution of the mean by the posterior distribution of the standard deviation. This procedure corresponds to the method-of-moment equation for estimating the SSMD between two dependent groups (Zhang, 2010). Sawilowsky's (2009) revised rules of thumb were used to interpret the effect size in SSMD units as "very small" (0.01), "small" (0.2), "medium" (0.5), "large" (0.8), "very large" (1.2), or "huge" (≥ 2.0). In the following paragraphs, the contrasts will be evaluated with reference to the UIs of the effects size estimates. Section 5 of the online supplement[1] provides the medians and the UIs for the posteriors of the contrasts described below.

Figure 3 shows the medians and the UIs of the marginal posteriors for each of the conditions in the study. This figure shows larger differences in d' scores between aspiration noise levels for /æ/ than for /i/ vowels. For each of the three speakers, the medians of the effect size posteriors for differences between aspiration noise levels were large to very large for /æ/ vowels ($\text{SSMD}_{\text{S08,/æ/,(4dB−2dB)}} = 1.7\,[1.0, 2.4]$; $\text{SSMD}_{\text{S30,/æ/,(4dB−2dB)}} = 1.3\,[0.8, 1.9]$; $\text{SSMD}_{\text{S44/æ/,(4dB−2dB)}} = 1.0\,[0.5, 1.5]$). For /i/ vowels, the median SSMDs were moderate to large ($\text{SSMD}_{\text{S08,/i/,(4dB−2dB)}} = 0.9\,[0.4, 1.4]$; $\text{SSMD}_{\text{S30,/i/,(4dB−2dB)}} = 0.8\,[0.3, 1.3]$; $\text{SSMD}_{\text{S44/i/,(4dB−2dB)}} = 0.7\,[0.2, 1.1]$). The differences in effect size for different vowels explains the inclusion of the vowel by aspiration noise interaction in the selected model. Second, the figure shows that the differences in performance among speakers were much smaller for /æ/ than for /i/, resulting in the vowel by speaker interaction. For example, for the /æ/, + 4-dB stimuli the largest differences were observed between speaker S44 and the other two speakers (both contrasts resulted in medium effect sizes: $\text{SSMD}_{\text{4dB/æ/,(S08−S44)}} = 0.5\,[0.0, 0.9]$; $\text{SSMD}_{\text{4dB/æ/,(S30−S44)}} = 0.4\,[0.0, 0.9]$). The SSMD posteriors of all other speaker contrasts for /æ/ vowels had medians smaller than 0.2, and UIs that included both negative and positive values. In comparison, the effect sizes of the contrasts between speakers for the /i/ vowels were substantial. Specifically, the largest SSMDs were observed between S08 and S30 (huge and very large at +2 and +4dB aspiration noise levels, respectively: $\text{SSMD}_{\text{2dB/i/,(S30−S08)}} = 2.0\,[1.2, 2.9]$; $\text{SSMD}_{\text{4dB/i/,(S30−S08)}} = 1.4\,[0.7, 2.0]$). All other contrasts for the /i/ vowels resulted in medium effect sizes ($0.45 < \text{SSMD} < 0.64$). These findings show that the effect of vowel depended on speaker identity. The largest effect size for the difference between vowels was observed

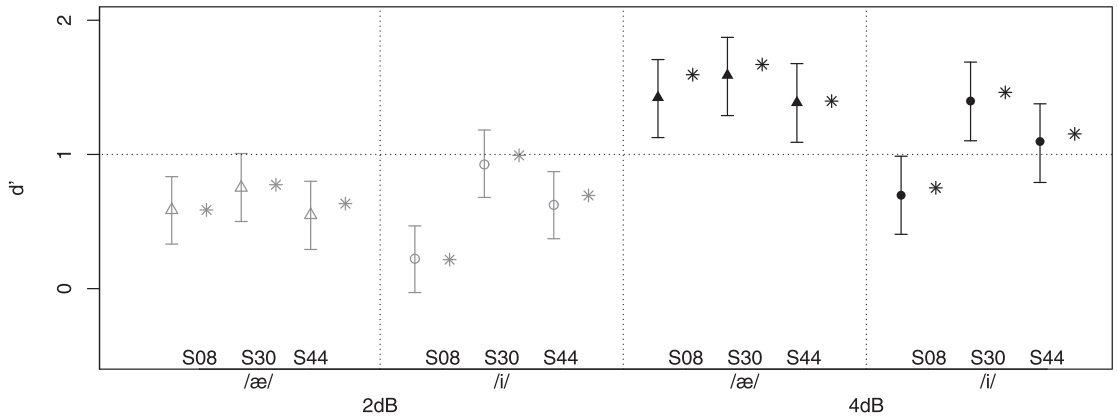

FIG. 3. Medians of the marginal posterior draws of all experimental conditions, with 90% UI bars (unfilled symbols); sample means are displayed as asterisks. The horizontal line at $d' = 1$ shows the level corresponding to the empirical threshold.

for speaker S08 with the +4-dB aspiration noise (very large effect size—$SSMD_{4dB,S08,/æ/-/i/} = 1.3\,[0.7, 1.8]$). Contrasts between the /i/ and /æ/ vowels of S08 at +2-dB aspiration noise and of S44 at +4-dB aspiration noise resulted in medium effect sizes (both had SSMD distributions with medians of 0.5, and UIs from 0.1 to 1.0). The remaining contrasts between vowels resulted in SSMD posteriors with small medians and UIs that included both negative and positive values with about equal probability, suggesting that there was uncertainty about the direction of the differences between vowels for these contrasts.

B. Acoustic analysis results

Table I shows the results of the acoustic analyses for the standard stimuli. This table shows that the NHRs for each band were similar across vowels and speakers. The distribution of noise and harmonic energy for the /i/ vowels differed compared to the /æ/ vowels. For /i/, the band that comprised the first formant (B0) had the highest energy level (68.5 to 69.9 dB). For /æ/, the band that contained the first and second formants (B1) had the highest levels (68.5 to 69.4 dB). Speaker-specific differences were also observed in the energy of B2 for the /i/ vowels. In this band, the lowest energy level (51.8 dB) was observed for speaker S08, which was much lower than the energy levels for speakers S44 and S30 (8.8 and 14.1 dB lower, respectively). For /æ/, the energy levels in B2 were similar across speakers (55.0 to 59.6 dB). As expected, the NHRs for each band of the comparison stimuli (not shown in the table) were about +2 and +4 dB higher for the +2- and +4-dB aspiration noise conditions, respectively (average + 2.1 dB and SD 0.4 dB, average + 3.8 dB and SD 0.6 dB, respectively). The NHRs were not exactly +2 and +4 dB higher because different aspiration noise samples were used to synthesize each stimulus.

CPPS values varied considerably across speakers and vowels, especially for the /i/ vowels. In particular, the CPPS values for S08 (20.8 for /i/ and 13.1 for /æ/) were higher than those of the other speakers (for whom values ranged from 6.2 to 10.9 dB). Relative to the standard stimuli, CPPS values were on average 1.4 lower for the +2-dB comparison stimuli (ranging from 1.1 to 1.7), and 3.1 lower for the +4-dB stimuli (ranging from 2.7 to 3.5); these values are not shown in Table I. Although the current study controlled the glottal synthesis parameters across all utterances, H1-H2 still varied across stimuli as a result of differences in vocal tract filters for the synthesis of different experimental stimuli. The differences in H1-H2 values among speakers were small for /æ/ (−4.1 to −4.6). For the /i/ vowel of speaker S08, the difference between H1 and H2 was much larger (−15.3 dB) than for the other speakers (−6.5 and −6.6 dB). The large H1-H2 difference value for the /i/ of S08 was likely due to the interaction between the filtering effects of a low average F1 frequency (321 Hz), and a higher average F0 (164 Hz) of this stimulus compared to the corresponding

TABLE I. Values for NHR and total energy within each frequency band, CPPS and H1-H2. Bn indicates the n-th band, where bands n = 0, 1, 2, 3 are delimited by 20–400 Hz, 400 Hz–2 kHz, 2 kHz–5 kHz, and 5 kHz–10 kHz, respectively.

	NHR (dB)				Total energy (dB)					
Stimulus	B0	B1	B2	B3	B0	B1	B2	B3	CPPS (dB)	H1-H2 (dB)
S08 /æ/	−35.1	−27.7	−19.1	−14.8	59.4	69.4	55.0	28.7	13.1	−4.6
S30 /æ/	−35.5	−26.0	−18.1	−14.1	63.0	68.6	58.6	33.9	9.0	−4.4
S44 /æ/	−33.7	−27.0	−17.8	−14.3	62.7	68.5	59.6	36.1	9.4	−4.1
S08 /i/	−37.4	−31.1	−18.7	−16.0	69.9	52.3	51.8	42.1	20.8	−15.3
S30 /i/	−34.4	−30.6	−17.2	−14.4	68.5	54.0	65.9	43.5	6.2	−6.5
S44 /i/	−35.9	−30.2	−17.4	−14.5	69.2	54.1	60.6	43.8	10.9	−6.6

https://doi.org/10.1121/10.0000756

values of the other speakers. H1-H2 for each of the comparison stimuli differed by less than 0.3 dB compared to the corresponding standard stimuli.

IV. DISCUSSION

Given the selection of d' value of 1 as the empirical threshold in the current study, aspiration noise DLs were estimated to be between +2- and +4-dB for all stimuli, except for the /i/ produced by speaker S08. For this utterance, the mean d' score remained below 1 at the +4-dB aspiration noise level. The estimated aspiration noise DLs were similar to those reported by Kreiman and Gerratt (2012), who found DLs of 3.64 to 4.35 dB for stimuli with similar glottal spectral slopes (−9 and −12 dB/octave) and NHR (−30 and −40 dB NHR standards) as the stimuli used in the current study. The present estimates were much lower than the DLs found by Shrivastav and Sapienza (2006), whose values were about 20 dB for stimuli with similar NHRs (−30 dB) as the stimuli used in the current study (−35 dB NHR). The study by Shrivastav and Sapienza and the current study differed in the experimental task. Shrivastav and Sapienza's task combined an adaptive procedure with a same-different response. Macmillan and Creelman (2005, pp. 168, 216, and 217) pointed out that same-different tasks are typically more difficult than 2AFC tasks. The use of a same-different task may have contributed to higher aspiration noise DLs in the study by Shrivastav and Sapienza. A d' value of about 2 with the current task, is equivalent to the target threshold for the same-difference task used by Shrivastav and Sapienza. This task used a proportion correct of 0.707 for the empirical threshold that theoretically correspond to a d' value of 2.3 if listeners use a differencing strategy, or a d' of 1.8 for an independent observation strategy (Macmillan and Creelman, 2005, pp. 217–225). More than a quarter of the d' scores for the +4-dB condition exceeded a value of 2, suggesting that the DLs would still be substantially lower about 20 dB, as observed by Shrivastav and Sapienza. Therefore, the reason for this discrepancy remains unclear.

In previous studies, differences in DLs were found for stimuli modelled after different speakers. Shrivastav and Sapienza (2006) suggested that relatively high noise levels above about 2 kHz were associated with smaller aspiration noise DLs. Similarly, Kreiman and Gerratt (2012) found empirically that relatively low levels of harmonic energy at high frequencies (stimuli with steeper slopes) resulted in smaller aspiration noise DLs, likely due to smaller amounts of masking of noise by high-frequency harmonics. The results from the present study also found credible differences in aspiration noise discrimination for different speakers even though the glottal source parameters were identical for all stimuli. Differences in the overall distribution of noise and harmonic energy across frequency likely affected aspiration noise discrimination in the absence of NHR differences. In other words, the same NHR in a specific band for two different stimuli might not lead to the same amount of noise being masked by harmonics, if the masker energy levels within the band differ between the stimuli. This conclusion is supported by the findings by Gockel *et al.* (2002) who showed that noise thresholds in harmonic series maskers depended on the presentation level of the maskers. To examine the perceptual importance of aspiration noise between 2 and 5 kHz (frequency band B2 in the present study), the results of the acoustic analysis were compared to the statistical model's posterior means in d' score units for the speaker-by-vowel interaction. Because the difference in the acoustic properties between standard and comparison stimuli varied in a predictable way for band energy and CPPS, and did not vary substantially for H1-H2, the discussion in the following paragraphs will focus on the acoustic properties of the standard stimuli. Given the small set of stimuli included in this study, and the fact that the acoustic measures covaried, the discussion of the association between individual acoustic measures and d' values is consequently descriptive. Therefore, the following observations about the relationships between the acoustic values and d' scores should be considered as suggestions for further investigations, rather than statements about the presence or absence of associations between acoustic parameters and perceptual responses.

B2 energy levels across speakers for the /i/ vowels (51.8, 60.6, 65.9 dB for S08, S44, S30, respectively) were proportionally related to the d' score medians pooled across aspiration noise levels ($d'_{/i/,S08} = 0.49$, [0.18, 0.80]; $d'_{/i/,S44} = 0.92$, [0.60, 1.23]; $d'_{/i/,S30} = 1.22$, [0.90, 1.53]). The energy levels between 2 and 5 kHz for the /æ/ vowels varied within a 4.6-dB range across the speakers (55.0, 58.6, 59.6 dB for S08, S30, S44, respectively). Consistent with the smaller variation in B2's energy levels, differences in d' scores between speakers were not as large for /æ/ as those for the vowel /i/. However, higher B2 energy levels were not associated with higher sensitivity for vowel differences. For example, the /æ/ and /i/ stimuli of speaker S08 had similar B2 energy levels (55.0 and 51.8 dB, respectively), yet there was a very large effect size at the +4-dB aspiration noise level ($SSMD_{4dB,S08,/æ/-/i/} = 1.3$ [0.7, 1.9]), and a medium effect size at the +2-dB aspiration noise level ($SSMD_{2dB,S08,/æ/-/i/} = 0.6$ [0.1, 1.0]) for this contrast. It is possible that the relative energy distribution across frequency bands may also contribute to the discrimination of aspiration noise rather than only the amount of energy within a specific band such as B2. Since the present study did not systematically vary the relative energy levels among the bands, it is not possible to draw further conclusions about how individual bands contributed to differences in aspiration noise discrimination.

For /i/ vowels, stimuli with larger CPPS values resulted in lower d' score means (higher DLs): /i/ stimuli for S30, S44, and S08 had CPPS values of 6.2, 10.9 and 20.8, and had d' score posterior medians of 1.2, 0.9, and 0.5, respectively, when pooled across aspiration noise levels. It appears that the large differences in CPPS across speakers account for the substantial effect size differences found when

contrasts were performed. This trend suggested that a small amount of noise added to a more periodic sound (larger CPPS) was more difficult to detect (resulting in higher DLs) than noise added to a less periodic signal. This finding is compatible with previous results that noise DLs were lower for sounds with greater levels of noise [and therefore lower CPPS values—Kreiman and Gerratt (2012); Shrivastav (2006)]. For the /æ/ vowels, similar CPPS values for S30 and S44 (9.0 and 9.4, respectively), and a slightly larger CPPS value for S08 (13.1) corresponded to similar d' score posterior medians (1.0, 1.2, and 1.0, respectively; these values were pooled across aspiration noise levels). Comparisons of the CPPS values across vowels were not made, because simulations by Fraile and Godino-Llorente (2014) showed that different vowels (/i/, /a/, and /U/) had distinctly different cepstral distributions and, therefore, inherently different CPPS values. CPPS is arguably one of the most robust measures of perturbations and aperiodicity in voice signals (Fraile and Godino-Llorente, 2014). In fact, CPPS is one of the acoustic measures that has been strongly associated with the perception of breathiness [Hillenbrand and Houde (1996), Samlan *et al.* (2013), Shrivastav and Sapienza (2003)—although see Hartl *et al.* (2003) for contrasting results]. In the current study, variations in CPPS appear to be associated with changes in d' scores within vowel categories.

No obvious relationship was observed between H1-H2 and d' scores across or within vowel categories. With the exception of the /i/ vowel for S08 (−15.3 dB), H1-H2 values across stimuli were all within a 2.5 dB range (−4.1 to −6.5 dB). Kreiman and Gerratt (2010) found that H1-H2 differences greater than 2.8 dB were required in order to discriminate between stimuli with glottal spectra similar to the stimuli used in the current study (approximately −9 dB/octave). Therefore, the differences in the H1-H2 value across the different stimuli in the current study may not have been large enough to be perceptually salient, with the exception of the /i/ vowel for speaker S08. It is not clear to what extent the H1-H2 value of the /i/ vowel of speaker S08 contributed to its low d' of 0.48. The large H1-H2 difference value can be attributed to the high F0 of this stimulus, because the filtering effect of the first formant resulted in a larger difference between H1 and H2 when the frequencies of H1 and H2 were further apart. It is possible that the higher amplitudes of the individual harmonic components masked the noise energy more efficiently than other stimuli, resulting in poorer discrimination performance. Generally, H1-H2 is a measure used to represent the spectral characteristics of the glottal spectrum: increases in H1-H2 suggest sinusoidal-like glottal waveforms that result from longer open quotients (Hanson, 1997). For this acoustic measure to represent more accurately the characteristics of the glottal waveform, researchers often use H1*-H2*, the difference between the first and second harmonic of the glottal spectrum, rather than H1-H2 (see, for example, Hanson, 1997, and Samlan *et al.*, 2013). H1*-H2* removes the filtering effect of the vocal tract on the amplitudes of H1 and H2. In the current study, all stimuli had the same glottal waveform (and the same H1*-H2*), and therefore differences in H1-H2 resulted only from the formant differences across stimuli.

V. CONCLUSIONS

The current findings showed that aspiration noise DLs were between 2 and 4 dB for non-pathological vowel standards. These results are consistent with the findings of Kreiman and Gerratt's (2012) study. Aspiration noise discrimination was affected by the vocal tract configuration, even when the glottal spectra were identical across speakers and vowels. While breathier voices tend to have higher amounts of aspiration noise at high frequencies *relative* to the amounts of harmonic energy, the relative increase in noise energy at high frequencies was not *necessary* to observe differences in aspiration noise discrimination. The differences in the total level of noise and harmonic energy in a frequency band were also associated with differences in aspiration noise discrimination, because the NHR in each frequency band was effectively the same across vowels and speakers. CPPS was associated with differences in aspiration noise discrimination in the current study. The values of H1-H2 were not consistently associated with differences in perceptual outcomes. Future studies that manipulate the distribution of the relative amount of energy in different spectral bands may further improve our understanding of the perception of aspiration noise in vowels.

ACKNOWLEDGMENTS

This research was supported by research funding provided to the second author by the Faculty of Medicine, University of British Columbia.

[1]See supplementary material at https://doi.org/10.1121/10.0000756 for: (i) values of the duration, the F0, the formants (frequency, bandwidth, and level), and the amplitude envelope used for each of the synthesized vowel stimuli (Sec. 1); (ii) d′ scores for each listener and each condition (Sec. 2); (iii) information on the linear mixed-effects model selection procedure (Sec. 3); (iv) parameters of the selected linear mixed-effects model (Sec. 4); (v) contrasts between relevant paired conditions (Sec. 5); (vi) information about how different priors affected the model selection procedure, the model parameters and contrasts (Sec. 6); and (vii) results of a frequentist statistical analysis (Sec. 7).

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**(3), 255–278.
Boersma, P., and Weenink, D. (**2017**). "Praat doing phonetics by computer" (version 6.0.30), www.praat.org (Last viewed 7/22/2017).
Bürkner, P. C. (**2017**). "An R package for Bayesian multilevel models using Stan," J. Stat. Softw. **80**(1), 1–28.
Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, A. B., Guo, J., Li, P., and Riddell, A. (**2017**). "Stan: A Probabilistic Programming Language," J. Stat. Softw. **76**(1), 1–32.
Dienes, Z. (**2011**). "Bayesian versus orthodox statistics: Which side are you on?," Prspect. Psychol. Sci. **6**(3), 274–290.
Fraile, R., and Godino-Llorente, J. I. (**2014**). "Cepstral peak prominence: A comprehensive analysis," Biomed. Signal Proces. **14**, 42–54.

Fukazawa, T., El-Assuooty, A., and Honjo, I. (**1988**). "A new index for evaluation of the turbulent noise in pathological voice," J. Acoust. Soc. Am. **83**(3), 1189–1193.
Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**(3), 556–565.
Gelman, A., Hwang, J., and Vehtari, A. (**2014**). "Understanding predictive information criteria for Bayesian models," Stat. Comp. **24**(6), 997–1016.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002**). "Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression," J. Acoust. Soc. Am. **111**(6), 2759–2770.
Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**(1), 466–481.
Hartl, D. A. M., Hans, S., Vaissière, J., and Brasnu, D. (**2003**). "Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia," Eur. Arch. Oto-Rhino-L. **260**(4), 175–182.
Hillenbrand, J. (**2003**). homepages.wmich.edu/~hillenbr/voweldata.html (Last viewed 31 January 2018.)
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.
Hillenbrand, J., and Houde, R. A. (**1996**). "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," J. Speech Hear. Res. **39**(2), 311–321.
Hox, J. J. C. M., van de Schoot, R., and Matthijsse, S. (**2012**). "How few countries will do? Comparative survey analysis from a Bayesian perspective," Surv. Res. Meth-Ger. **6**(2), 87–93.
Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**(1), 169–177.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.
Klein, S. A. (**2001**). "Measuring, estimating, and understanding the psychometric function: A commentary," Percept. Psychophys. **63**(8), 1421–1455.
Kreiman, J., and Gerratt, B. R. (**2010**). "Perceptual sensitivity to first harmonic amplitude in the voice source," J. Acoust. Soc. Am. **128**(4), 2085–2089.
Kreiman, J., and Gerratt, B. R. (**2012**). "Perceptual interaction of the harmonic source and noise in voice," J. Acoust. Soc. Am. **131**(1), 492–500.
Kreiman, J., Gerratt, B. R., and Berke, G. S. (**1994**). "The multidimensional nature of pathologic vocal quality," J. Acoust. Soc. Am. **96**(3), 1291–1302.
Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). "Listener experience and perception of voice quality," J. Speech Hear. Res. **33**(1), 103–115.
Kruschke, J. K. (**2010**). "What to believe: Bayesian methods for data analysis," Trends Cogn. Sci. **14**(7), 293–300.
Kruschke, J. K., and Liddell, T. M. (**2017**). "Bayesian data analysis for newcomers," Psychon. B. Rev. **25**(1), 155–177.
Ladefoged, P. (**1982**). "The linguistic use of different phonation types," UCLA Work. Pap. Phon. **54**, 28–39.
Lee, S., and Song, X. (**2004**). "Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes," Multivar. Behav. Res. **39**(4), 653–686.
LiveCode (**2016**). Version 8.1.1, https://downloads.livecode.com/livecode/ (Last viewed 6/22/2016).
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Psychology Press, Mahwah, NJ).
Meredith, M., and Kruschke, J. K. (**2018**). "BEST: Bayesian estimation supersedes the t-test" version 0.5.1, https://CRAN.R-project.org/package=BEST (Last viewed 10/19/2019).
Miller, J. (**1996**). "The sampling distribution of d′," Percept. Psychohys. **58**(1), 65–72.
R Core Team (**2017**). "R: A language and environment for statistical computing," http://www.R-project.org/ (Last viewed 7/5/2019).
Samlan, R. A., Story, B. H., and Bunton, K. (**2013**). "Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling," J. Speech Lang. Hear. Res. **56**(4), 1209–1223.
Sawilowsky, S. (**2009**). "New effect size rules of thumb," J. Mod. Appl. Stat. Meth. **8**(2), 597–599.
Shrivastav, R. (**2006**). "Multidimensional scaling of breathy voice quality: Individual differences in perception," J. Voice. **20**(2), 211–222.
Shrivastav, R., and Sapienza, C. M. (**2003**). "Objective measures of breathy voice quality obtained using an auditory model," J. Acoust. Soc. Am. **114**(4), 2217–2224.
Shrivastav, R., and Sapienza, C. M. (**2006**). "Some difference limens for the perception of breathiness," J. Acoust. Soc. Am. **120**(1), 416–423.
Sorensen, T., Hohenstein, S., and Vasishth, S. (**2016**). "Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists," Quant. Methods Psychol. **12**(3), 175–200.
The Mathworks, Inc. (**2017**). "Matlab," version 2017b (The Mathworks, Inc., Natick, MA).
Theil, H. (**1950**). "A rank-invariant method of linear and polynomial regression analysis part III," Proc. R. Neth. Acad. Arts Sci. **53**, 1397–1412.
Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., and Van Aken, M. A. (**2014**). "A gentle introduction to Bayesian analysis: Applications to developmental research," Child Dev. **85**(3), 842–860.
Vehtari, A., Gelman, A., and Gabry, J. (**2017**). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," Stat. Comput. **27**(5), 1413–1432.
Watanabe, S. (**2010**). "Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory," J. Mach. Learn. Res. **11**, 3571–3594.
Zhang, X. D. (**2010**). "Strictly standardized mean difference, standardized mean difference and classical t-test for the comparison of two groups," Stat. Biopharm. Res. **2**(2), 292–299.