Research Article

Categorization in the Perception of Breathy Voice Quality and Its Relation to Voice Production in Healthy Speakers

Yeonggwang Park,[a] Joseph S. Perkell,[a,b] Melanie L. Matthies,[a] and Cara E. Stepp[a,c,d]

Purpose: Previous studies of speech articulation have shown that individuals who can perceive smaller differences between similar-sounding phonemes showed larger contrasts in their productions of those phonemes. Here, a similar relationship was examined between the perception and production of breathy voice quality.
Method: Twenty females with healthy voices were recruited to participate in both a voice production and a perception experiment. Each participant produced repetitions of a sustained vowel, and acoustic correlates of breathiness were calculated. Identification and discrimination tasks were performed with a series of synthetic stimuli along a breathiness continuum. Categorical boundary location and boundary width were obtained from the identification task as a measurement of perception of breathiness. Spearman's correlation analysis

was performed to estimate associations between values of boundary location and width and the acoustic correlates of breathiness from the participants' voices.
Results: Significant correlations between boundary width ($r = -.53$ to $-.6$) and some acoustic correlates were found, but no significant relationships were observed between boundary location and the acoustic correlates.
Conclusions: Speakers with small boundary widths, which suggest higher perceptual precision in differentiating breathiness, had typical voices that were less breathy, as estimated with acoustic measures, compared to speakers with large boundary widths. Our findings may support a link between perception and production of breathy voice quality.
Supplemental Material: https://doi.org/10.23641/asha.9808478

T he link between speech perception and production has been investigated in previous studies (Fox, 1982; Franken, Acheson, McQueen, Eisner, & Hagoort, 2017; Ghosh et al., 2010; McAllister Byun & Tiede, 2017; Newman, 2003; Perkell et al., 2004), which have found that individuals' perceptual abilities to distinguish between similar-sounding phonemes were correlated across speakers with the amount of contrast between those phonemes that individuals produced in their own speech. Specifically, greater acuity was associated with a larger contrast in the sounds produced. This link between perception

and production has been demonstrated in vowels (Fox, 1982; Franken et al., 2017; Perkell et al., 2004), sibilants (Ghosh et al., 2010), and rhotic consonants (McAllister Byun & Tiede, 2017). Similarly, significant correlations were observed between acoustic measures of listeners' perceptual prototypes for phonemes and acoustic measures of their own productions of the phonemes; this relation was observed in voice onset time for stop consonants and frequencies of spectral peaks for voiceless fricatives (Newman, 2003). Many researchers have attributed this relationship between perception and production to the use of auditory targets and auditory feedback during speech production (Franken et al., 2017; Perkell et al., 2004).

The role of auditory feedback in the achievement of phonetic targets has been observed in experiments in which altered auditory feedback resulted in compensatory changes in speech output (Houde & Jordan, 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Purcell & Munhall, 2006). According to the Directions Into Velocities of Articulators (DIVA) model of speech production (Guenther, 2016), a speech sound map hypothesized in the left ventral premotor cortex stores and provides desired auditory targets for comparison with incoming auditory signals. When a mismatch

[a]Department of Speech, Language, and Hearing Sciences, Boston University, MA
[b]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge
[c]Department of Biomedical Engineering, Boston University, MA
[d]Department of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine, MA

Correspondence to Yeonggwang Park: yppark@bu.edu

occurs between the auditory target and the incoming signal, the auditory feedback control system in the DIVA model corrects for the errors in nearly real time and updates a feedforward motor command (Tourville & Guenther, 2011). Because both auditory targets and auditory feedback control contribute to how speech sounds are perceived, speech perception is inferred to influence speech production.

In this study, we examined whether there is also a relation between voice perception and production. Voice perception involves the analysis of auditory signals into vocal pitch, loudness, and quality; thereby, auditory feedback is thought to play an important role in maintaining typical pitch, loudness, and quality in voice production. Guenther (2016) posited that variations in prosodic parameters such as vocal pitch, loudness, and duration are controlled with feedforward and feedback mechanisms similar to those in segmental speech motor control. The role of auditory feedback in voice motor control has also been observed in experiments with pitch-shifted feedback, which resulted in compensatory changes in produced fundamental frequency (f_o) (Burnett, Freedland, Larson, & Hain, 1998; Donath, Natke, & Kalveram, 2002; Jones & Munhall, 2000; R. Patel, Niziolek, Reilly, & Guenther, 2011). However, much less is known about the interaction between perception and production of voice quality (e.g., roughness, breathiness, and strain), despite the importance of voice quality in communication (Ishikawa et al., 2017).

To characterize the perception of contrasts between individual speech sounds, researchers have used a series of speech stimuli that differ by constant physical amounts and span the range between two phonemes. Using these stimuli, they have performed classic tasks to evaluate categorical perception. These consist of identifying variants of two different phonemes and discriminating between two stimuli that are adjacent in the series of stimuli that range between the phonemes (Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni & Lazarus, 1974). A steep boundary in the identification curve and a peak in the discrimination rate (percent correct) at the categorical boundary have been noted as key features of categorical perception (Liberman et al., 1957). Likewise, most of the studies examining the relationship between speech perception and speech production also have used these identification and discrimination tasks to evaluate the characteristics of phoneme perception (McAllister Byun & Tiede, 2017; Perkell et al., 2004). From these tasks, they obtained the boundary width, which is determined by the slope of the identification curve, and the peak discrimination rate. In the current study, we applied a similar approach to characterizing individual perception of breathiness by using a series of synthetic stimuli along a "breathiness" continuum in identification and discrimination tasks.

The assumption behind our use of a boundary width in this study is that the perception of voice quality would involve the categorization of different voice qualities, similar to perception of phonemes. One account of how phoneme categories develop is the perceptual magnet effect, which holds that frequent exposure to a native language leads to warping of auditory perceptual spaces, forming categories of different phonemes (Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Kuhl, 1991). Neural correlates of the warping of auditory perceptual spaces were later demonstrated with a functional magnetic resonance imaging experiment; this showed that, when prototypical examples of /i/ were played frequently to participants, the size of the cortical representation of prototypical examples decreased, along with the discriminability of /i/ sounds near the prototype (Guenther & Bohland, 2002). The reduction in the size of the cortical representation could be interpreted as underlying a diminished response to unimportant variations among similar variants of the same phoneme (Goldstone & Hendrickson, 2010). As we are exposed to different voice qualities in everyday life, our auditory perceptual spaces for voice qualities may be adapted to varying demands by forming different categories. Depending on the nature of the maturational environment, different people may develop different characteristics and degrees of categorization for voice qualities. Therefore, we aimed to measure the characteristics of categories using boundary location and the degree of categorization with boundary width.

To examine the relationship between perception and production of voice quality, we chose breathiness. Breathiness was chosen to examine the relationship between perception and production of voice quality over other voice quality percepts (e.g., roughness) because it has a clear physiological basis and robust acoustic correlates. Breathiness is a vocal percept that is mainly the result of hearing the sound generated by the audible escape of air past a speaker's glottis due to incomplete glottal closure during voicing, as well as the effects of a source spectrum with attenuated periodic components at higher frequencies due to a more rounded, symmetrical glottal waveform (Hillenbrand, Cleveland, & Erickson, 1994). It can be related to voice pathologies or habitual voice patterns (Labuschagne & Ciocca, 2016). In healthy voices, different degrees of vocal fold adduction and patterns of vocal fold vibration can result in differences in perceived breathiness: Less adducted, more gradually closing vocal folds produce a breathier sound (Hanson, 1997). Breathiness has also shown strong correlations with several acoustic measures (Hillenbrand et al., 1994; Labuschagne & Ciocca, 2016), leading to the inference that those measures can be used as acoustic estimates of the degree of breathiness in a speaker's voice.

In this study, we tested the hypotheses that speakers' perceptual boundary locations along the breathiness continuum, perceptual boundary widths, or both would be related to the breathiness of the speakers' own voice productions. Figure 1 illustrates the study hypotheses. Individuals with higher boundary locations (schematized in Figure 1A by increased perceptual distance near the breathy, i.e., the right, end) would have typical voices that are more breathy, as estimated with acoustic measures. Individuals with narrower boundary widths, which represent more categorical precision between typical and breathy voices (schematized in Figure 1B by two distinct clusters, spaced further apart from one another), would have voices that are less breathy, as estimated with acoustic measures. We also examined the

Figure 1. Schematic of study hypotheses on the effects of boundary location (A) and boundary widths (B) on auditory targets. Acoustic space represents actual acoustic distances of the stimuli represented as white circles on the breathiness continuum. Perceptual space represents how listeners would perceive the stimuli, and shorter distances between stimuli would represent more perceptual similarity. The color green indicates stimuli that would be perceived as typical voices; purple indicates stimuli that would be perceived as breathy voices. Auditory targets represent acoustic regions that the listeners would target when they produce their typical or breathy voices.
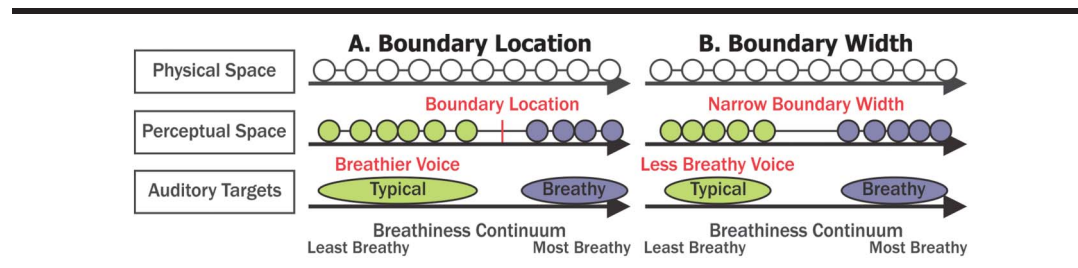
results of identification and discrimination tasks to evaluate whether breathiness is indeed perceived categorically.

Method

Participants

Participants were 20 women aged 19–34 years (M_{age} = 24 years) who reported no history of speech, language, or hearing disorders. The target number of participants was determined by computing the power associated with correlation analysis (assuming α = .05, β = 0.2, r = .5). All participants were native speakers of American English who grew up in the United States and reported that only English was spoken at home. Smokers, singers, and students who had taken courses in speech, language, and hearing sciences were excluded. None of the participants reported any throat discomfort or any illness at the session and scored in the normal range on the Voice Handicap Index (Jacobson et al., 1997) and the Reflux Symptom Index (Belafsky, Postma, & Koufman, 2002). All participants passed a pure-tone hearing screening with 25 dB HL pure tones at 250, 500, 1000, 2000, and 4000 Hz in a sound-treated room (American Speech-Language-Hearing Association, 2005). The participants provided written consent prior to participation, in compliance with the Boston University Institutional Review Board.

Voice Production Experiment

A voice production experiment was performed in order to assess the acoustic correlates of breathiness in participants' own voices. Voice samples were collected in a sound-treated room following the recommended protocols for instrumental voice assessment (R. R. Patel et al., 2018).

Each participant sustained three repetitions of the vowel /ɑ/ for 3–5 s in a comfortable voice. They were not asked to phonate at a specific sound pressure level in order to obtain samples that approximated their typical-use voices. They were asked to produce the /ɑ/s with steady pitch and loudness. During this task, some participants produced unsteady or glottalized voices and were asked to repeat the /ɑ/ again until they had produced three stable /ɑ/s. The participants were recorded in a sound-treated room

using SONAR Artist (Cakewalk) and a Shure headset WH20QTR microphone (Shure), placed 7 cm from the participants' lips at a 45° angle. The microphone signal was amplified by an RME Quadmic II microphone preamplifier (RME) and sampled at 44.1 kHz with a 16-bit resolution by a MOTU UltraLite-MK3 (MOTU).

Estimates of Breathiness

The common features of acoustic signals that are perceived as breathy are diminished periodicity, an enhanced first harmonic amplitude,[1] and a higher level of the energy of the noise component of the signal (vs. harmonics) at high frequencies (Hillenbrand et al., 1994; Klatt & Klatt, 1990). We focused on acoustic measures that reflect these features and have shown correlations with perceived breathiness. Breathy voices present with an overall reduction in the periodicity of the signal due to increased aspiration noise. As measures of the strength of the periodicity, smoothed cepstral peak prominence (CPPS) and harmonics-to-noise ratio (HNR) were chosen.[2] CPPS is a cepstral peak amplitude normalized over the entire background signal amplitude calculated from the smoothed cepstrum; thus, it can represent the strength of the cepstral peak (which reflects the

---

[1]H1–H2, a measure of the spectral energy ratio between the first and second harmonics, can represent the relative strength of the first harmonic. However, H1–H2 is also known to be influenced by nasality, which usually increases the energy in the spectrum around 250 Hz due to a nasal formant (Arai, 2006). As a result, an interspeaker comparison of H1–H2 could be problematic because the nasal formant may enhance the spectral energy differently depending on speakers' fundamental frequencies and harmonics (Simpson, 2009). We were also concerned that the individual differences in the degree of nasality might affect the measure instead of breathiness, so we decided not to include H1–H2 in our study.

[2]An acoustic estimate of pitch strength is also known to capture signal periodicity and predict perceived breathiness (Eddins, Anand, Camacho, & Shrivastav, 2016). Pitch strength refers to the degree to which listeners can perceive pitch in a sound, and this perceptual measure has also shown a strong correlation with perceived breathiness (Shrivastav, Eddins, & Anand, 2012); however, pitch strength is not commonly used as an acoustic correlate of breathiness and was not included in this study.

degree of periodicity of the signal) compared to the cepstral background noise in the voice signal (Hillenbrand et al., 1994). HNR is calculated as the ratio of the periodic energy to the aperiodic energy. We also used a measure that incorporates high-frequency energy due to aspiration noise containing energy primarily at mid and high frequencies (Klich, 1982; Shoji, Regenbogen, Yu, & Blaugrund, 1992). The high-to-low spectral ratio (HL ratio) is a spectral tilt measure calculated as the ratio of high-frequency energy (> 4000 Hz) to low-frequency energy (< 4000 Hz), which was chosen for inclusion. In summary, three measures, namely, CPPS, HNR, and HL ratio, were included as the acoustic correlates of breathiness to assess participants' voices.

To measure these acoustic correlates of breathiness, a 1-s stable segment (the steadiest portion with the most constant amplitude) was extracted from the middle of each /ɑ/ production. Three 1-s segments were collected from three /ɑ/ productions of each participant. Acoustic analysis was performed on these 1-s segments using Praat software (Version 6.0.21). CPPS was obtained with commands and parameters described in Watts, Awan, and Maryn (2017), and HNR was obtained from Praat's command, 'Voice report.' The HL ratio was calculated as the ratio of energy in high frequency (> 4000 Hz) to energy in low frequency (< 4000 Hz) from Praat's spectral analysis of the 1-s segment. All measures were obtained from each 1-s segment, and the three values (from three 1-s segments) of each measure were averaged to characterize each participant's voice.

In addition, in order to verify that these measures actually corresponded to perceived breathiness in the speaker sample, two voice-experienced speech pathologists performed the Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V; American Speech-Language-Hearing Association, 2002). The speech pathologists evaluated the three 1-s segments from each participant using the CAPE-V form, which includes 100-mm visual analog scales for each parameter (Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009), and the breathiness scores from the two raters were averaged for each participant.

Perceptual Experiment

The same 20 participants also completed a perceptual protocol for measuring the categorical boundary location of breathiness and evaluating the extent of categorization of breathy voice quality. Ten stimuli were synthesized along a breathiness continuum, based on a natural production of /ɑ/ by a female speaker (who was not a participant in the experiment), using the University of California, Los Angeles

(UCLA) voice synthesizer (Kreiman, Gerratt, & Antoñanzas-Barroso, 2016). Breathiness in a female voice may sound more natural than in a male voice due to a higher prevalence of breathiness in females who speak American English (Hanson & Chuang, 1999). The breathiness continuum consisted of a series of stimuli that differed by constant physical quantities of two synthesis parameters that are related to breathiness. Noise-to-harmonics ratio (NHR) directly correlates with the amount of aspiration noise in the signal; the UCLA voice synthesizer allows direct modification of NHR during the synthesis. The synthesizer models the noise spectrum from the original voice sample and uses the model parameters as inputs for the synthesis; this has the benefit of modeling the natural voice more accurately (Kreiman et al., 2016). The first harmonic amplitude can be manipulated in the synthesizer by setting the value of the return time constant (t_a) in a four-parameter model of glottal flow, known as the *Liljencrants–Fant model* (Fant, Liljencrants, & Lin, 1985). The return time constant (t_a) in the Liljencrants–Fant model represents the closing time of the vocal folds in one glottal cycle; as t_a increases, the duration of the open phase of the glottal cycle increases (Fant et al., 1985). When the open phase duration increases, the resulting glottal waveform has a smoother shape, which increases the relative strength of the first harmonic amplitude compared to higher frequencies; the resulting signal is usually perceived as more breathy (Hanson, 1997).

To create 10 stimuli across the breathiness continuum, both NHR and t_a were modified to generate the least and most breathy stimuli that still sounded natural. The least breathy stimulus had the lowest NHR and t_a, and the most breathy stimulus had the highest NHR and t_a. The rest of the stimuli were generated by linearly interpolating between the least and most breathy stimuli, using even spacing in values of NHR and t_a (see Table 1). Increasing NHR and t_a together not only increases the perception of breathiness but also reduces a potential confusion with perceived nasality, which is also related to a high open phase. Increasing the degree of aspiration noise has been shown to decrease the perception of nasality in speech signals with a high open phase (Arai, 2006; Klatt & Klatt, 1990). After creating stimuli using the UCLA synthesizer, we adjusted the intensity of the stimuli to the same output level (dB) using Praat's command "scale intensity," so that all 10 stimuli had identical intensity (see Supplemental Material S1 for the sound clips for the 10 stimuli). The duration of each stimulus was 1 s, with 10-ms rise-and-fall times. We also performed an acoustic analysis on the stimuli and confirmed that the chosen

Table 1. Stimuli in the breathiness continuum (1 = *least breathy* to 10 = *most breathy*; t_a and noise-to-harmonics ratio [NHR] values in each stimulus).

Stimulus	1	2	3	4	5	6	7	8	9	10
t_a (ms)	0.14	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
NHR (dB)	−29	−27	−25	−23	−21	−19	−17	−15	−13	−11

Table 2. The results of acoustic analysis on the stimuli in the breathiness continuum (1 = *least breathy* to 10 = *most breathy*).

Stimulus	1	2	3	4	5	6	7	8	9	10
HL ratio (dB)	−36.1	−35.2	−33.7	−31.9	−30.0	−28.1	−26.1	−24.2	−22.3	−20.4
CPP (dB)	16.6	15.5	14.5	13.6	12.7	11.7	10.8	10.0	9.1	8.2
HNR (dB)	28.4	26.6	24.8	22.8	20.9	18.9	17.0	15.0	13.0	11.1

Note. HL ratio = high-to-low spectral ratio; CPP = cepstral peak prominence; HNR = harmonics-to-noise ratio.

acoustic correlates of breathiness (CPPS, HNR, and HL ratio) varied as predicted along the continuum of the stimuli (see Table 2).

Tasks

The same participants completed two tasks for evaluating categorical perception: (a) an identification task in which they were instructed to classify each stimulus as "breathy" or "not breathy" and (b) a discrimination task in which they differentiated between two adjacent stimuli along the breathiness continuum (Liberman et al., 1957). The participants were tested individually in a sound-treated room, and the stimuli were presented at a comfortable loudness through a pair of Sennheiser HD-290 headphones. During the identification task, the participants determined if the stimulus was "breathy" or "not breathy" in a forced-choice button press task after hearing each stimulus. They were not given a definition of a "breathy" voice; rather, they were asked to determine whether a voice was "breathy" based on their individual judgments and experiences in daily conversations. We did not supply a definition of breathiness in order to avoid higher cognitive processes from affecting the lower level perception of breathiness. Each of the 10 stimuli in the breathiness continuum was presented three times in a randomized order; the task was composed of 30 trials and took approximately 4 min to complete. The participants then discriminated between pairs of stimuli in an ABX discrimination task in which stimuli A, B, and X were presented sequentially with a 1-s interval between each stimulus (Liberman et al., 1957). Stimulus X was the same as either stimulus A or B, and the participants determined whether X corresponded to either A or B. Stimuli A and B were always adjacent on the breathiness continuum, so nine pairs of A and B were possible. For each pair, the four possible orders of presentation (ABA, ABB, BAA, and BAB) were each presented three times, resulting in 36 trials (9 pairs × 4 orders). The trials were presented in a randomized order, and the task took approximately 8 min to complete.

Measurement of the Categorical Boundary Location and the Boundary Width for Breathiness

A categorical boundary location for breathiness was measured using the data from the identification task. The number of "breathy" responses at each step of the breathiness continuum was converted to an identification percentage, and the identification percentage curve of each participant was fitted with a sigmoid curve. The point in the breathiness

continuum at which the sigmoid curve reached 50% identification was collected as the boundary location (de Gelder, Teunisse, & Benson, 1997; McAllister Byun & Tiede, 2017). The 50% location along the continuum has been regarded as the location for a categorical boundary—that is, where the perception of stimuli changes sharply from one category to another (de Gelder et al., 1997; Liberman et al., 1957). The boundary width was defined as the distance in the breathiness continuum between where the identification curve reached 25% and 75% (McAllister Byun & Tiede, 2017).

Statistical Analysis

Perception–Production Relation

To test our primary hypothesis that the perception of breathy voice quality is related to the production of breathy voice, we performed Spearman's correlation analysis to estimate associations between the boundary location and width and the acoustic correlates of breathiness in the participants' voices (one-tailed). Spearman's correlation was chosen because the boundary width data showed a skewed distribution (skewness = 1.14, kurtosis = 1.34). We also performed an additional Spearman correlation analysis with boundary location, boundary width, acoustic correlates of breathiness, and perceived breathiness from CAPE-V ratings to confirm that the observed correlations between boundary measures and the acoustic correlates were also reflected in the perceived breathiness and that the acoustic correlates were appropriate representatives of perceived breathiness. Spearman's correlations range from −1.0 to +1.0; the magnitudes of the correlations were interpreted as follows: 0–0.1 as negligible, 0.1–0.39 as weak, 0.40–0.69 as moderate, 0.70–0.89 as strong, and 0.90–1.00 as very strong (Schober, Boer, & Schwarte, 2018). For all statistical analyses, a significance level of α = .05 was used. Bonferroni correction was not performed because of the exploratory nature of the study.

Reliabilities of the Perceived Breathiness

Intra- and interrater reliabilities of the two raters on the breathiness parameter of the CAPE-V were assessed. The raters rerated a randomly selected 20% of the participants' voices, and intrarater reliability was assessed using Pearson's correlations. The two raters' r values were .81 and .91 (mean r = .86). Interrater reliability was assessed using intraclass correlation coefficient (two-way mixed effects, consistency, single measure) and was found to be .50, which was slightly lower than the previous reported interrater

reliability for breathiness using the CAPE-V (intraclass correlation coefficient = .6; Zraick et al., 2011). Although interrater reliability was somewhat lower than previously reported, this may be the case because ratings were performed using sustained /ɑ/s instead of the full CAPE-V protocol, which includes sustained vowels, sentences, and running speech. Sustained /ɑ/s were elicited because they were also used to extract the acoustic correlates.

Categorization of Breathiness

We evaluated whether breathy voice would be perceived categorically using data from the discrimination task. The discrimination performance at the categorical boundary was evaluated statistically relative to the non-boundary performance values, calculated as the mean discrimination performances averaged across remaining stimulus pairs. A paired (α = .05) t test was performed to determine if the boundary discrimination percentage was significantly different than nonboundary percentages (de Gelder et al., 1997).

Results

Participants' mean perceived breathiness from CAPE-V ratings was 17.5 (SD = 16.5). The standard deviation of the perceived breathiness suggested that there was a moderate range of breathiness among the participants. This variation in the degree of perceived breathiness was considered to be sufficient for the correlation analyses used to evaluate the link between perception and production of breathy voice quality.

Acoustic Correlates of Breathiness Versus Perceived Breathiness of the Same Tokens

Significant correlations were found among the acoustic correlates of the produced tokens and their perceived breathiness (see Table 3). CPPS showed the strongest correlation (r = −.83, p < .01), and HNR also showed a strong correlation (r = −.77, p < .01). HL ratio had a significant, but weaker, correlation with perceived breathiness than CPPS and HNR (r = .45, p = .02). Based on this Spearman's

Table 3. Spearman correlations between the acoustic correlates of breathiness and perceived breathiness from Consensus Auditory–Perceptual Evaluation of Voice ratings.

Measures	CPPS	HNR	HL Ratio	Breathiness
CPPS	1.00	.80*	−.61*	−.83*
HNR		1.00	−.76*	−.77*
HL ratio			1.00	.45
Breathiness				1.00

Note. CPPS = smoothed cepstral peak prominence; HL ratio = high-to-low spectral ratio; HNR = harmonics-to-noise ratio.

*p < .01

analysis, we confirmed that CPPS and HNR were appropriate acoustic measures that corresponded to perceived breathiness in our study sample.

Perception–Production Relation

Figure 2 presents plots of boundary location versus acoustic correlates of breathiness, and Figure 3 presents plots of boundary width versus acoustic correlates of breathiness. From Spearman's analyses, no significant correlations were found between the boundary location and acoustic correlates. However, significant, moderate correlations were found between boundary width and the acoustic correlates, CPPS (r = −.53, p = .008) and HNR (r = −.60, p = .003). The directions of the correlations were as hypothesized; as the boundary width decreased, there were increases in CPPS and HNR, which reflect more periodic and less breathy voices. A significant, weak correlation was found between boundary width and HL ratio (r = .39, p = .04), and the direction of the correlation was also as hypothesized: As the boundary width decreased, there was a decrease in HL ratio, which reflects having less high-frequency energy in the spectrum and a less breathy voice.

From the plots of boundary width (see Figure 3), it was apparent that the distribution of boundary width was skewed, such that half of the participants (n = 10) had a boundary width near 0. Thus, "narrow boundary" and "wide boundary" groups were assigned based on the boundary

Figure 2. Scatter plots and best linear fits of boundary location versus smoothed cepstral peak prominence (CPPS), harmonics-to-noise ratio (HNR), and high-to-low spectral ratio (HL ratio). Spearman's correlations are represented as r values. No significant correlations were found between boundary location and the acoustic measures.
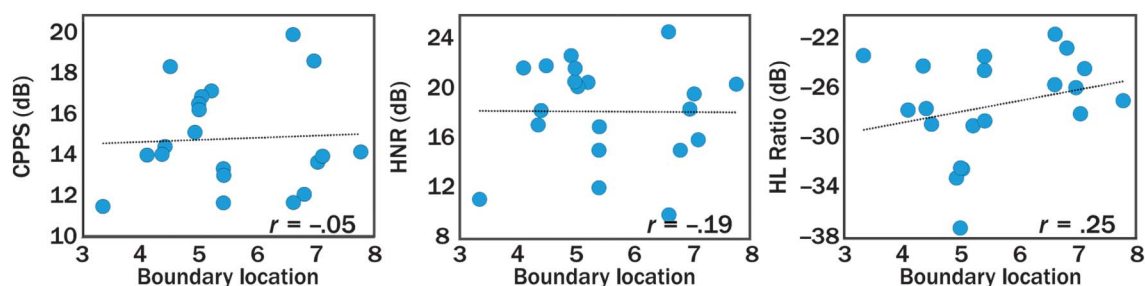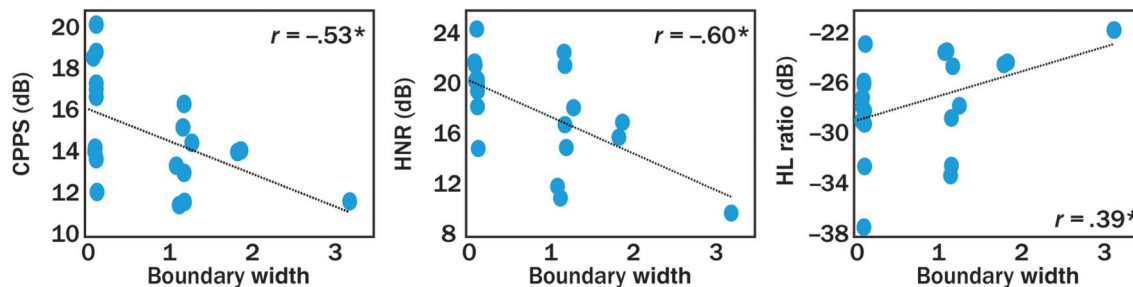
Figure 3. Scatter plots and best linear fits of boundary width versus smoothed cepstral peak prominence (CPPS), harmonics-to-noise ratio (HNR), and high-to-low spectral ratio (HL ratio). Spearman's correlations are represented as *r* values. *p < .05.

width results, and the means were compared. This classification was similar to Perkell et al. (2004), who assigned participants as "high" and "low" discriminators of vowels based on their discrimination performance. The narrow boundary group consisted of participants who had boundary width values near 0 (n = 10, range: 0.05–0.09), and the wide boundary group consisted of the rest of the participants, whose boundary width values ranged from 1.1 to 3.2 (n = 10). We hypothesized that the narrow boundary group would have typical voices that would be less breathy as estimated with the acoustic correlates; independent-samples t tests were performed to compare the means of the acoustic measures between the groups (one-tailed; α = .05), and Cohen's d effect sizes were calculated. We confirmed that the narrow boundary group had significantly higher CPPS and HNR values compared to the wide boundary group (t = 2.8 and 2.7, p = .006 and 0.007, d = 1.2 and 1.2, respectively; see Figure 4). No significant difference was found between group mean values of HL ratio (t = −1.2, p = .12).

Boundary width also showed a significant correlation with perceived breathiness from CAPE-V ratings (see Figure 5; r = .54, p = .007). Perceived breathiness also showed a significant group difference (see Figure 5; t = −2.3, p = .017, d = 1.0), with the wide boundary group having significantly breathier voices.

Both the narrow boundary and wide boundary groups exhibited identification data that were well fit by sigmoidal curves, which indicates a sharp perceptual change of one

category to another, a hallmark of categorical perception (see Figure 6). However, the identification data from the wide boundary groups contained more data points that deviated from the fit curves, and the slopes of the identification curves were shallower than in the narrow boundary group. This trend is consistent with the other findings, suggesting that the wide boundary group was less consistent in their categorizations of typical versus breathy voices.

Categorization of Breathiness

Boundary and nonboundary discrimination percentages are presented in Figure 7. Boundary discrimination percentage was not significantly greater than nonboundary discrimination percentages (t = 0.143, p = .9). These data do not support the categorical perception of breathy voice quality, which is usually characterized by boundary discrimination greater than nonboundary discrimination.

Discussion
The Relationship Between Perception and Production of Breathy Voice Quality

A relationship between voice perception and production was demonstrated by the statistically significant correlations between the perceptual boundary width and acoustic correlates of breathiness from participants' voices. People who had smaller boundary widths, which reflect higher

Figure 4. Mean smoothed cepstral peak prominence (CPPS), harmonics-to-noise ratio (HNR), and high-to-low spectral ratio (HL ratio) of each group. Error bars indicate 95% confidence intervals. *Speakers with narrow boundaries had significantly higher CPPS and HNR than speakers with wide boundaries.
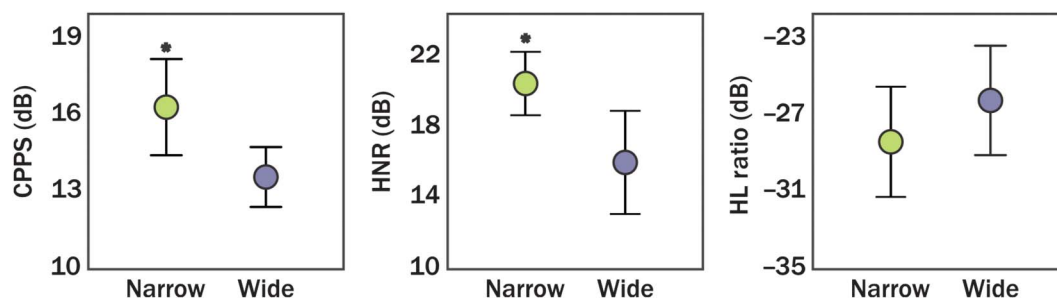
Figure 5. (Left) A scatter plot and a best linear fit of boundary width versus breathiness rating from the Consensus Auditory–Perceptual Evaluation of Voice with Spearman's correlation represented as an *r* value. (Right) Mean breathiness ratings of each group. Error bars indicate 95% confidence intervals. *Speakers with narrow boundaries had significantly lower breathiness ratings than speakers with wide boundaries.
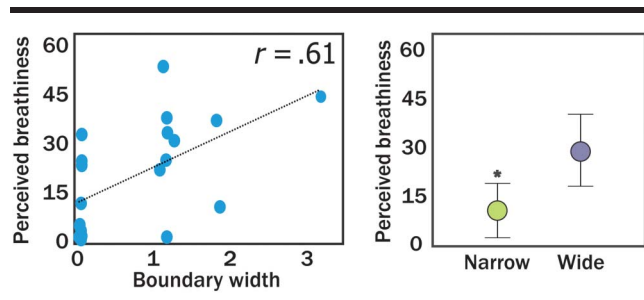
precision in differentiating between typical and breathy voices, produced their voices with less breathiness as estimated by the acoustic correlates. Having typical voices with less breathiness may imply that the speakers made greater distinction between typical and breathy voices. Our results are in general agreement with previous studies, which found that variation in the ability to differentiate speech sounds is positively correlated with the size of contrasts speakers produce for those sounds (Franken et al., 2017; Ghosh et al., 2010; McAllister Byun & Tiede, 2017; Perkell et al., 2004). Researchers have suggested that this link could be explained by the presence of auditory targets, influenced by perception, and feedforward and feedback control mechanisms, which guide productions to stay within the boundaries of auditory targets (Franken et al., 2017; Perkell et al., 2004). According to this explanation, people with more acute perception would have auditory targets for speech sounds that are smaller and further apart from one another than people with less acute perception (Tourville & Guenther, 2011). The results of this study suggest that voice quality may be controlled with a similar mechanism: Individuals with narrow boundaries in their categorical perception of breathiness may have auditory target regions for "typical voice" that are smaller and further apart from those auditory target regions associated with breathy voices, thus resulting in voice production with acoustic features consistent with less breathiness in their typical speech.

Boundary Width Versus Boundary Location

We observed significant correlations between boundary widths and acoustic correlates of breathiness, but not between boundary locations and those acoustic measures. Based on these results, we infer that the location of category boundaries may play a less important role in control of breathiness compared to perceptual precision. We included the boundary location as a perceptual measure in an attempt to characterize individual perceptual prototypes of typical and breathy voice qualities and expected that

individuals who had perceptual prototypes for typical voices located in a less breathy area would be likely to have less breathy voices. However, we did not find statistically significant correlations between boundary location and the acoustic correlates. One possible reason for this finding may be due to the compactness of the prototype in the auditory target region. Although individuals may have a perceptual prototype for typical voice located in the less breathy auditory space, if they have a low perceptual precision, they are likely to have less compact, large auditory targets for typical voices, which may result in breathier voice production. The greater importance of boundary width versus boundary location has also been shown in the link between speech perception and production, such as the case of American rhotic perception and production. Children who misarticulated between /w/ and /r/ showed shallow and inconsistent identification curves, whereas children with normal articulation showed the classic sharp transition between the two phonemes (Hoffman, Daniloff, Bengoa, & Schuckers, 1985). The relationship between the steepness of the identification curve and more contrastive production was also shown in a study in which Japanese speakers showed shallower categorical boundaries between American English /r/ and /l/ than American speakers, although their boundary locations did not significantly differ (Best & Strange, 1992).

Another reason that boundary location may be less important in voice control is the fact that breathiness is not considered to be phonologically distinctive in American English. In phoneme categories, boundary locations of two phonemes are known to be language specific, as observed in a study in which French, American, and Japanese speakers showed different boundary locations for the same stimuli continuum (Hallé, Best, & Levitt, 1999). For this reason, Liberman et al. (1957) suggested that learning a new language requires learning to perceive different categorical boundary locations between two phonemes, which is specific to the language of interest. However, voice quality, in general, is thought to instead be used to convey emotion, mood, or affect in conversational settings (Gobl & Ní Chasaide, 2003), in which the precision of the acoustic cue of breathy voice or other voice qualities is expected to be less important than precision of the phonemic production.

### The Role of Auditory Feedback in Perception–Production Relation

Our results supporting the link between perception and production of voice quality extend the understanding of voice motor control with auditory feedback. The use of auditory feedback for the control of voicing has been evidenced by observations from the Lombard effect, which demonstrated that people increase sound intensity in noisy environments (Lombard, 1911). Support has also come from f_o perturbation experiments, which demonstrated that people modulate their f_o to compensate for perturbed auditory feedback (e.g., Houde & Jordan, 1998; Larson et al., 2001; Purcell & Munhall, 2006). Our findings may support

Figure 6. Identification curves for all participants, divided into narrow and wide boundary groups (*n* = 10 each). The *x*-axes represent the stimuli number. The *y*-axes are omitted. Each curve starts with 0 identification percentage at Stimuli 1 and ends with 100 identification percentage at Stimuli 10.
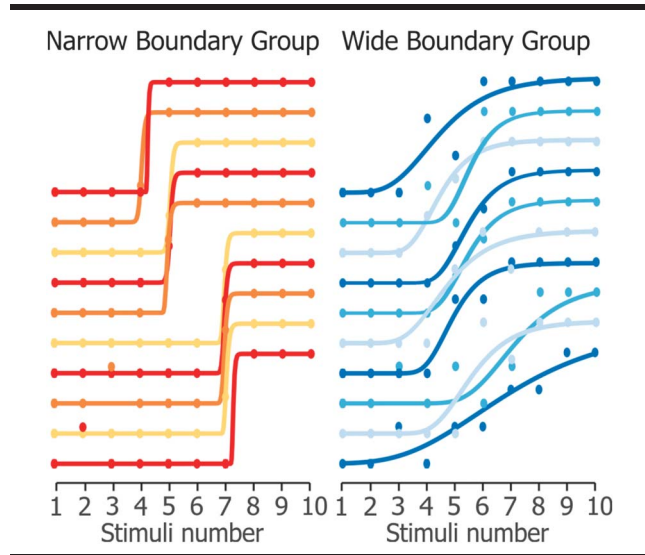
the role of auditory feedback in controlling breathiness, as we found that people who perceive breathy and non-breathy voices more distinctively had acoustic correlates that represent less breathiness in their typical voices. These people may have developed more refined perceptual space for nonbreathy and breathy voices and thus have smaller and more distinct auditory targets for typical and breathy voice. We speculate that auditory feedback of breathiness enables the realization of the auditory targets and also helps to update feedforward commands so that the typical voice can stay within the individual's auditory target, as predicted by the DIVA model for speech control (Guenther, 2016).

Figure 7. Boundary discrimination and averaged nonboundary discrimination percentages of individual participants (red: narrow boundary group, blue: wide boundary group), connected by lines.
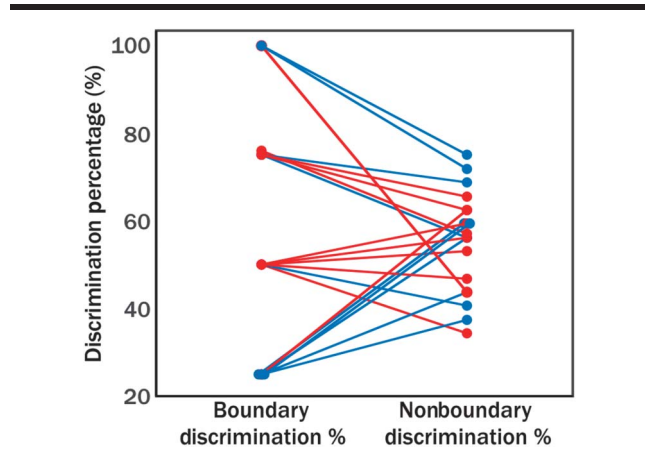
Auditory Targets for Breathiness

From the results of this study, we can also examine which aspects of breathy voice quality are more likely to be monitored in sensorimotor control of the voice. We used acoustic correlates of breathiness that reflect two major aspects of breathy voice: periodicity and spectral tilt. We found that the periodicity measures, CPPS and HNR, showed stronger correlations with boundary width and perceived breathiness than HL ratio. This result possibly suggests that the periodicity of the signal may be a key aspect in the sensorimotor control of breathiness. This finding is also consistent with previous perceptual studies in which signal periodicity was found to be the most important factor in predicting perceived breathiness (Hillenbrand et al., 1994; Klatt & Klatt, 1990). HL ratio showed somewhat weaker correlations with boundary width and perceived breathiness than did the other two acoustic measures. This finding also corresponds to results of previous perceptual studies (Hillenbrand et al., 1994; Klatt & Klatt, 1990), which showed a moderate correlation between HL ratio and a perceptual rating of breathiness. HL ratio may show weaker correlations with perceived breathiness because it is measured with spectral energies above and below 4000 Hz; both frequency ranges can contain both periodic (i.e., harmonic) and aperiodic energy. Although the high-frequency band may contain more aperiodic components, the percentage of periodic and aperiodic elements can vary between individuals, so the HL ratio may be less accurate in representing breathy voice quality.

Categorization and Methodology

In order to test whether perception of breathy voice quality would exhibit a feature of categorical perception similar to other speech parameters, we performed a discrimination task and statistically compared the boundary discrimination percentage with the nonboundary discrimination rate. The heightened boundary discrimination percentage at the categorical boundary has been considered as a hallmark of categorical perception (Liberman et al., 1957). We did not observe a significantly higher boundary discrimination rate than average nonboundary discrimination rate, but this does not preclude the possibility that some degree of categorization occurs in perceiving breathy voice. As support for the hypothesis of categorization of breathy voice quality, half of the participants showed steep identification curves (see Figure 6), another hallmark of categorical perception. The discrimination task has received some criticism for its task dependency when it is used to examine categorical perception (Pisoni & Lazarus, 1974). Different factors such as stimulus pair steps (e.g., one or two steps) and types of task (e.g., ABX and 4IAX) have shown to affect the results (Gerrits & Schouten, 2004; Perkell et al., 2004). When participants were instructed to pay attention to small details of acoustic cues rather than phonemes, people also showed heightened discrimination scores within the categorical boundary (Pisoni & Lazarus, 1974). These

findings suggest that categorical and continuous perceptions are not dichotomous and different degrees of categorization are possible. For example, vowels are known to be perceived less categorically compared to consonants, but phoneme categories exist in vowels from frequent exposure to prototypes for different vowel categories when learning a native language, as described in relation to the perceptual magnet effect (Pisoni, 1973). Thus, a similar kind of categorization may exist in the perception of voice quality from speakers being exposed frequently to specific voice qualities.

Comparison to Values of the Acoustic Measures in the Literature

We compared the results of our acoustic analysis to values from the literature for young female adults with typical voices. The published normative values are 25.3 dB ($SD = 3.1$; Goy, Fernandes, Pichora-Fuller, & van Lieshout, 2013) for HNR and −31.3 dB ($SD = 3.6$; Garrett, 2013) for HL ratio (it was reported as LH ratio, which is the same as the –HL ratio). The mean values for our speakers were 18.3 dB ($SD = 4.0$) and −27.3 dB ($SD = 4.1$) for HNR and HL ratio, respectively. There is no normative value published for CPPS as estimated from sustained vowels in Praat, since obtaining CPPS using Praat is a relatively new method. The mean CPPS value reported previously for a sample that included both males and females with both healthy and disordered voices was 22.9 dB ($SD = 4.1$; Watts et al., 2017). Our mean CPPS value was 15.0 dB ($SD = 2.6$).

Comparison of these values appears to indicate that our participants had breathier voices (lower CPPS and HNR and higher HL ratio) than the individuals examined in previous studies. However, this apparent difference might actually be due to a difference between recording signal-to-noise ratios (SNRs) in the studies. Acoustic measurements can be affected by recording environment; thus, it has been recommended that ambient SNR be higher than 30 dB for general acoustic analysis and 42 dB for acoustic analysis associated with voice quality (Deliyski, Evans, & Shaw, 2005). Although all samples were collected in a sound-treated room, the mean background SNR in our study was 27.6 dB ($SD = 6.3$). Goy et al. (2013) reported an average SNR of 42 dB, whereas Garrett (2013) and Watts et al. (2017) did not report their SNRs. Based on this observation, we suspect that our recordings may have included more background noise and that the acoustic analysis may indicate the participants' voices as breathier than they actually are. If this speculation were true, the observed correlations between the acoustic correlates and boundary width would then likely underestimate the strengths of the relationships. The strong correlation results, despite the potential for recording noise in the acoustic signals, suggest that the correlations between the boundary width and the acoustic correlates of breathiness may be even stronger than those reported here.

Limitations and Future Directions

In this study, we used synthesized stimuli, which may have diminished the naturalness and thus the ecological validity of the study; however, using synthesized stimuli allowed us to manipulate the signals systematically, which is a study strength. Another potential limitation of this study was that only two speech-language pathologist perceptually evaluated the breathiness of speaker productions. Perceptual evaluations have received some criticism over their low interrater reliability (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993); therefore, we attempted to mitigate concerns about the reliability of perceptual measures of voice quality by incorporating acoustic correlates. The acoustic correlates still need to be interpreted with caution because the SNR for the acoustic recordings was lower than the recommended SNR for acoustic analysis associated with voice quality (Deliyski et al., 2005). Another limitation of the current study is that it only explored the perception and production relationship of voice quality using breathiness, which was chosen because of the relatively strong understanding of its physiological basis and convincing acoustic correlates. It remains unclear how this relationship between perception and production may apply to other types of voice quality, such as strain and roughness. Finally, an additional weakness of this study is that the results cannot be generalized to include individuals with voice disorders. The demonstrated link between voice perception and production is likely to be limited to individuals with healthy voices, since numerous individuals with voice disorders are breathy due to glottal incompetence resulting from structural pathology, rather than from voice motor control.

Some voice disorders, such as vocal hyperfunction, may not involve structural pathology. Individuals with hyperfunction may provide a population in which to study the link between voice perception and production. Vocal hyperfunction is a common symptom of many voice disorders, which is characterized by abuse and/or misuse of laryngeal or extralaryngeal muscles (Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). Researchers have suggested that the etiology of vocal hyperfunction may include different factors, such as psychological stress, vocal abuse/misuse, or compensatory mechanisms for sudden changes in laryngeal structures (Van Houtte, Van Lierde, & Claeys, 2011). One additional factor in development and persistence of vocal hyperfunction may be disordered sensorimotor integration (Stepp et al., 2017). Individuals with vocal hyperfunction have been shown to respond differently in an auditory–motor perturbation experiment compared to control participants, suggesting that people with vocal hyperfunction may have altered sensorimotor integration. Similarly, Tam, Carding, Heard, and Madhill (2018) found that individuals with vocal hyperfunction have reduced pitch discrimination abilities compared to individuals with healthy voices. Further examination of the relationship between perception and production in individuals with vocal hyperfunction may shed light on the pathophysiology of this prevalent voice disorder.

Conclusion

The link between perception and production in breathy voice quality was supported with perceptual and production experiments. Participants who showed greater precision in categorizing typical and breathy voice qualities had typical voices that were less breathy, as estimated with acoustic measures, compared to participants with lower precision. In line with previous findings on the link between speech perception and production (McAllister Byun & Tiede, 2017; Perkell et al., 2004), we assert that this finding may be attributed to auditory feedback mechanisms and the presence of auditory targets for breathy and typical voices that are smaller and further apart in people with high categorical precision. This link between perception and production of breathy voice quality can be explored further to offer future insight into sensorimotor integration and control in hyperfunctional voice disorders.

Acknowledgment

References

American Speech-Language-Hearing Association. (2002). *Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V): ASHA Special Interest Division 3, Voice and Voice Disorders.* Retrieved from https://www.asha.org/uploadedFiles/members/divs/D3CAPEVprocedures.pdf

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry* [Guidelines]. Retrieved from http://www.asha.org/policy

Arai, T. (2006). Cue parsing between nasality and breathiness in speech perception. *Acoustical Science and Technology, 27,* 298–301.

Belafsky, P. C., Postma, G. N., & Koufman, J. A. (2002). Validity and reliability of the Reflux Symptom Index (RSI). *Journal of Voice, 16*(2), 274–277.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305–330.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America, 103*(6), 3153–3161.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11*(1), 1–23.

Deliyski, D. D., Evans, M. K., & Shaw, H. S. (2005). Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice, 19*(2), 176–186.

Donath, T. M., Natke, U., & Kalveram, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *The Journal of the Acoustical Society of America, 111*(1), 357–366.

Eddins, D. A., Anand, S., Camacho, A., & Shrivastav, R. (2016). Modeling of breathy voice quality using pitch-strength estimates. *Journal of Voice, 30*(6), 774.e1–774.e7.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *STL-QPSR, 26*(4), 1–13.

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception–production link. *Phonetica, 39*(1), 1–22.

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. (2017). Individual variability as a window on production–perception interactions in speech motor control. *The Journal of the Acoustical Society of America, 142*(4), 2007–2018.

Garrett, R. (2013). *Cepstral- and spectral-based acoustic measures of normal voices* (Master's thesis). The University of Wisconsin-Milwaukee, Milwaukee, WI.

Gerrits, E., & Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*(3), 363–376.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., . . . Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America, 128*(5), 3079–3087.

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40*(1–2), 189–212.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69–78.

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice, 27*(5), 545–555.

Guenther, F. H. (2016). *Neural control of speech.* Cambridge, MA: MIT Press.

Guenther, F. H., & Bohland, J. W. (2002). Learning sound categories: A neural model and supporting experiments. *Acoustical Science and Technology, 23*(4), 213–220.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *The Journal of the Acoustical Society of America, 106*(5), 2900–2912.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics, 27*(3), 281–306.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America, 101*(1), 466–481.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America, 106*(2), 1064–1077.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769–778.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32*(2), 373–392.

Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders, 50*(1), 46–53.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213–1216.

Ishikawa, K., Boyce, S., Kelchner, L., Golla Powell, M., Schieve, H., de Alarcon, A., & Khosla, S. (2017). The effect of background

noise on intelligibility of dysphonic speech. *Journal of Speech, Language, and Hearing Research, 60*(7), 1919–1929.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6,* 66–70.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America, 108*(3), 1246–1251.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory–Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857.

Klich, R. J. (1982). Relationships of vowel characteristics to listener ratings of breathiness. *Journal of Speech and Hearing Research, 25*(4), 574–580.

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2016). *UCLA voice synthesizer*. Los Angeles: UCLA School of Medicine.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*(1), 21–40.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93–107.

Labuschagne, I., & Ciocca, V. (2016). The perception of breathiness: Acoustic correlates and the influence of methodological factors. *Acoustical Science and Technology, 37*(5), 191–201.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America, 110*(6), 2845–2848.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368.

Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of voice raising]. *Annals of Ear and Larynx Diseases, 37,* 101–119.

McAllister Byun, T., & Tiede, M. (2017). Perception–production relations in later development of American English rhotics. *PLOS ONE, 12*(2), e0172022.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America, 113*(5), 2850–2860.

Patel, R., Niziolek, C., Reilly, K., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research, 54*(4), 1051–1059.

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., . . . Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887–905.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America, 116*(4), 2338–2344.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253–260.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America, 55*(2), 328–333.

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America, 119*(4), 2288–2297.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763–1768.

Shoji, K., Regenbogen, E., Yu, J. D., & Blaugrund, S. M. (1992). High-frequency power ratio of breathy voice. *The Laryngoscope, 102*(3), 267–271.

Shrivastav, R., Eddins, D. A., & Anand, S. (2012). Pitch strength of normal and dysphonic voices. *The Journal of the Acoustical Society of America, 131*(3), 2261–2269.

Simpson, A. P. (2009). *Breathiness differences in male and female speech. Is H1–H2 an appropriate measure?* Paper presented at FONETIK 2009, Stockholm University, Stockholm, Sweden.

Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter Noordzij, J., & Lupiani, A. A. (2017). Evidence for auditory–motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research, 60*(6), 1545–1550.

Tam, K. H., Carding, P., Heard, R., & Madhill, C. J. (2018). *The relationship between voice quality and pitch discrimination ability in a population with features of mild vocal hyperfunction [Abstract].* Paper presented at the the Voice Foundation's 47th Annual Symposium, Philadelphia, PA.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981.

Van Houtte, E., Van Lierde, K., & Claeys, S. (2011). Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge. *Journal of Voice, 25*(2), 202–207.

Watts, C. R., Awan, S. N., & Maryn, Y. (2017). A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice, 31*(3), 387.e1–387.e10.

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., . . . Glaze, L. E. (2011). Establishing validity of the Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22.