Journal of Speech and Hearing Research, Volume 37, 769–778, August 1994

Acoustic Correlates of Breathy Vocal Quality

James Hillenbrand
Ronald A. Cleveland
Robert L. Erickson
Department of Speech Pathology and Audiology
Western Michigan University, Kalamazoo

The purpose of this study was to evaluate the effectiveness of several acoustic measures in predicting breathiness ratings. Recordings were made of eight normal men and seven normal women producing normally phonated, moderately breathy, and very breathy sustained vowels. Twenty listeners rated the degree of breathiness using a direct magnitude estimation procedure. Acoustic measures were made of: (a) signal periodicity, (b) first harmonic amplitude, and (c) spectral tilt. Periodicity measures provided the most accurate predictions of perceived breathiness, accounting for approximately 80% of the variance in breathiness ratings. The relative amplitude of the first harmonic correlated moderately with breathiness ratings, and two measures of spectral tilt correlated weakly with perceived breathiness.

KEY WORDS: acoustic analysis, breathy voice, voice quality, cepstrum, autocorrelation

This study was designed to improve our understanding of the acoustic correlates of breathy vocal quality. Breathy voice plays an important role in both normal and disordered speech production. Breathiness is a prominent feature of numerous pathological conditions affecting the laryngeal mechanism, including neoplasms, abductor spasmodic dysphonia, vocal cord paralysis, and laryngeal trauma (Aronson, 1971, 1990; Colton & Casper, 1990). Breathiness can also be associated with vocal misuse and other functional conditions (Aronson, 1990; Boone & McFarlane, 1988), and there is evidence that the physiological effects of aging may include breathy voice (Hollien, 1987; Ryan & Burk, 1974). Breathy (murmured) voice quality also plays an important role in the phonological systems of many of the world's languages (Fischer-Jorgensen, 1967; Huffman, 1987; Ladefoged, 1975, 1983). Finally, there is some evidence that breathy quality may be a social marker of female gender among English speakers (Klatt & Klatt, 1990; McKay, 1987). Despite the widespread occurrence of breathy voice, a good deal remains to be learned about the acoustic features underlying the perception of breathiness.

Acoustic Correlates of Breathy Voice

First harmonic amplitude. Breathiness is thought to be due to incomplete glottal closure during the "closed" phase of the phonatory cycle (Fairbanks, 1940; Hillenbrand et al., 1990; Zemlin, 1968). Breathy glottal source signals obtained through inverse filtering typically show more symmetrical opening and closing phases with little or no complete closed phase (Bickley, 1982; Fischer-Jorgensen, 1967; Huffman, 1987). The rounded, more nearly sinusoidal shape of breathy glottal waveforms is responsible for increases in the relative amplitude of the first harmonic.

Enhanced H1 amplitude in the spectra of breathy speech signals has been observed by a number of investigators (Bickley, 1982; Fischer-Jorgensen, 1967; Huffman, 1987; Klatt & Klatt, 1990; Ladefoged, 1983). Huffman (1987) used inverse filtering to derive glottal waveforms from samples of four phonation types used in

0022-4685/94/3704-0769

Hmong. Breathy (murmured) samples showed stronger first harmonics than nonbreathy samples.

Fischer-Jorgensen (1967) used a variety of techniques to study the acoustic characteristics of Gujarati murmured vowels. She considered the high intensity of H1 to be the most salient spectral feature of murmured vowels and conducted a listening experiment to measure the effects of filtering on the perception of breathy vocal quality. Highpass filtering at 230 Hz was used to reduce H1 amplitude by about 25 dB. Contrary to the expected outcome, correct identification of murmured vowels was not significantly decreased. Despite her belief that the relative amplitude of H1 was "the most obvious and constant feature" (p. 133–134), Fischer-Jorgensen concluded that no single acoustic feature was sufficient to produce the sensation of breathiness.

A perceptual study by Bickley (1982) used synthetic Gujarati breathy and clear word pairs varying in aspiration noise and H1 amplitude. Identification of the stimuli as breathy or clear by native Gujarats was affected by H1 amplitude only, with no effect for aspiration noise. However, as was noted by Klatt & Klatt (1990), the H1 enhancement that was needed to effect a decisive shift from clear to breathy greatly exceeded measured H1 amplitude differences between naturally produced breathy and clear word pairs.

Ladefoged (1983) reported enhanced H1 amplitudes for breathy vowels in !Xóõ. A follow-up study by Ladefoged & Antonanzas-Barroso (1985) found that the breathiness judgments of American listeners were more strongly correlated with H1 amplitude than aspiration noise, as measured by a waveform variability index.

Klatt & Klatt (1990) used a variety of acoustic and perceptual techniques to investigate male-female differences in voice quality. In general, women were judged to be breathier than men. H1 amplitude measures were also generally greater for women than men. A variety of acoustic measures were evaluated by measuring correlations with breathiness ratings. Only two were found to be significantly correlated with breathiness ratings: H1 amplitude and the degree of aspiration noise present in band-limited signals. In a second listening experiment using synthetic speech, a number of synthesis parameters were manipulated in order to measure their effect on the perception of vocal quality. Increases in H1 alone were heard as breathy by some listeners and nasal by many others. Signals were never judged to be nasal when H1 amplitude increases were accompanied by aspiration noise. Contrary to previous investigators, Klatt & Klatt concluded that the sensation of breathiness is controlled primarily by aspiration noise in the middle and upper portion of the spectrum.

Additive noise. When a portion of the breathstream passes through a persistent and relatively narrow glottal chink it results in the generation of noise. The higher frequency harmonics are reduced in amplitude and the upper portion of the spectrum becomes dominated by dense aspiration noise (Klatt & Klatt, 1990).

Based on spectrographic observations, Fischer-Jorgensen (1967) believed differences in spectral noise between murmured and clear vowels were small and inconsistent. Similarly, Bickley (1982) found no correlation between breathiness ratings and increases in additive noise in synthetic vowels. In contrast, a synthesis study by Hillenbrand (1988a) found a strong relationship between breathiness ratings and additive noise.

Klatt & Klatt (1990) used a bandpass filter centered at F3 to isolate the third formant of [ha] samples. Unfiltered signals, which tend to be dominated by low-frequency periodic components, were judged to be unsuitable for noise estimation. (A similar observation was made by Kasuya, et al., 1986, who noted that even pathological voices sometimes exhibit well defined harmonic structure in the lower frequencies.) The degree of periodicity in the band-limited signals was judged by visual inspection of time-domain waveforms using a five-point rating scale. The noise ratings accounted for approximately 60% of the variance in listener ratings of breathiness. A follow-up synthesis study found that increases in spectral noise were the single most important cue to perceived breathiness.

Spectral tilt. Several investigators have noted that breathy signals tend to have more high-frequency energy than normally phonated signals. For example, Klich (1982) reported strong correlations between perceived breathiness and several measures of spectral tilt calculated as energy ratios of low-, mid-, and high-frequency bands. A similar energy ratio method was described by Frokjaer-Jensen & Prytz (1976), who showed a reduction in high-frequency energy in the long-term average spectra of voice patients following treatment.

A closely related spectral tilt measure described by Fukazawa, El-Assuooty, and Honjo (1988) was defined as the ratio of the energy in the second derivative of a signal to the energy in a nonderived signal. The index is conceptually similar to the ratios described by Klich (1982) and Frokjaer-Jensen & Prytz (1976), except that the Fukazawa et al. technique produces a global measure of spectral tilt that does not require arbitrary cutoffs for low-, mid-, or high-frequency bands. Fukazawa et al. reported that the spectral tilt measure accounted for approximately half of the variance in breathiness ratings obtained from the sustained vowels of speakers with various laryngeal pathologies.

In contrast to these findings, Klatt & Klatt (1990) reported no significant correlations between breathiness ratings and two measures of spectral tilt. Further, a synthesis study by Hillenbrand (1988a) found that breathiness ratings were affected only by the level of aspiration noise, with no effect for spectral tilt.

In summary, previous work on breathy voice suggests that the perception of breathiness is affected by first harmonic amplitude and the presence of aspiration noise. There are conflicting findings on the relative importance of these two cues. Some investigators (e.g., Bickley, 1982; Fischer-Jorgenson, 1967; Ladefoged & Antonanzas-Barroso, 1985) have concluded that H1 amplitude is the primary cue to breathiness, with aspiration noise playing little or no role. More recently, Klatt & Klatt (1990) have suggested that the presence of aspiration noise is the primary cue to breathiness, with H1 amplitude playing a secondary role. There are also conflicting findings on the relationship between spectral tilt and breathiness, with some studies (e.g., Frokjaer-Jensen & Prytz, 1976; Fukazawa et al., 1988; Klich, 1982) suggesting that breathiness is associated with greater amounts of

high frequency energy, and other studies (e.g., Hillenbrand, 1988a; Klatt & Klatt, 1990) suggesting that spectral tilt plays little or no role in the perception of breathy voice.

The primary purpose of the present study was to measure the relationship between breathiness ratings and a relatively large set of acoustic measurements. The measurements fell into three categories: (a) measures of signal periodicity, (b) measures of first harmonic amplitude, and (c) measures of spectral tilt. A second purpose was to introduce two fully automatic measures of signal periodicity. These automated periodicity measures were intended to supplement or replace the subjective, visual rating method used by Klatt & Klatt (1990).

Methods

Recording of Voice Samples

Subjects and training. Voice samples were provided by eight men and seven women between the ages of 22 and 37 with no reported history of voice, speech, or hearing problems. Nine additional subjects were recorded who proved to be unable to perform the speaking task.

A sample recording of a naturally produced normal, moderately breathy, and very breathy [u] was used to familiarize talkers with the speaking task. Each talker was trained to produce three voicing variations (normal, moderately breathy, and very breathy) at estimated average fundamental frequency. Average fundamental frequency was estimated by asking each subject to read a portion of the Rainbow Passage (Fairbanks, 1940). Average fundamental frequency was calculated from this reading with a Kay Elemetrics Visi-Pitch/IBM PC Interface (Horii, 1983).

Recordings. Recordings were made in a sound-treated chamber using a unidirectional microphone (Audio-Technica 250XL) at a distance of 7–10 cm in front of the lips and 3 cm above the breath stream. The signals were recorded with a Sony PCM-F1 digital audio processor. A portable electronic keyboard was used to provide talkers with the target pitch determined previously through the Rainbow Passage analysis. The Visi-Pitch was used again to provide talkers with visual feedback regarding fundamental frequency and duration during the recording of vowel tokens. Speech materials included the four vowels [a], [i], [æ] and [o]. A randomized list of these vowels was provided for each talker. Talkers were asked to sustain each vowel for approximately 3 seconds. Each utterance was recorded at least twice during a 45–60 minute recording session. A total of 12 vowel samples (4 vowels × 3 voicing variations) was selected from each talker's recordings, yielding 180 vowel stimuli.

The signals were low-pass filtered at 7.2 kHz and digitized at 20 kHz using a 12-bit A/D board (Data Translation DT3382). The 3 sec sustained vowels were then digitally edited to the most stable 1 sec segment. A preliminary estimate of the maximally stable 1 sec interval was derived by computing a non-pitch-synchronous amplitude variability measure over consecutive 1 sec segments of the signal at intervals of 100 msec (e.g., 0–1000 msec, 100–1100 msec, etc.). The amplitude variability measure was essentially a shimmer calculation (Horii, 1980), except that it was computed non-pitch-synchronously using adjacent 10 msec intervals. The 1 sec segment of the signal showing the lowest amplitude variability was taken as a preliminary estimate of the maximally stable segment. The signal was then edited at the zero-crossings so that it included an integer number of pitch periods. In a few cases, the 1 sec segment extracted in this way did not appear by ear to be the most steady in terms of pitch. In these cases, a more appropriate interval was chosen by listening to different 1 sec segments of the signal. After editing, the signals were ramped on and off with a 20 msec inverted and lifted cosine function to eliminate onset and offset transients.

Breathiness Ratings

Twenty listeners (19 women, 1 man) were recruited from among graduate students and faculty in the Speech-Language Pathology Department at Western Michigan University. Listeners were audiometrically screened at 20 dB HL at five frequencies (0.5, 1, 2, 4, 6 kHz). Listening experiments were conducted in a sound-treated room. The stimuli consisted of 180 vowels (15 talkers × 3 breathiness levels × 4 vowels). Signals were low-pass filtered at 7.2 kHz, amplified and presented free field at approximately 80 dBA over a single loudspeaker (Boston Acoustics A60).

Listeners were asked to rate each signal according to the amount of perceived breathiness. Subjects were allowed to repeat each signal as often as they wished before entering their responses on a computer terminal. The direct magnitude estimation procedure required the listeners to determine their own numerical rating scales. Subjects were simply told to "enter a large number if the signal is very breathy and a small number if the signal shows little or no breathiness." Ratings were later linearly rescaled so that each listener's ratings for the 180-stimulus set ranged from 0.0 to 1.0.

Listeners participated in two 30–45 minute listening sessions at least 24 hours apart. In each listening session signals were presented in random order within three blocks including all 180 signals. A practice session of 60 trials began all listening sessions to familiarize subjects with the task and the range of breathiness percepts. The practice trials were identical to the listening session except that ratings from these trials were disregarded.

Acoustic analysis

The six acoustic measures that were used are described below. The measures were based on three assumptions about breathy voice: (a) the more rounded glottal waveforms for breathy signals should produce stronger first harmonics, (b) breathy signals should be less periodic, especially in the mid and high frequencies where aspiration noise is most prominent, and (c) breathy signals should show more high frequency energy than normally phonated signals. With the exception of first harmonic amplitude, the measures were fully automatic. Explanations of some of the signal processing techniques can be found in Parsons (1986) and Witten (1982).

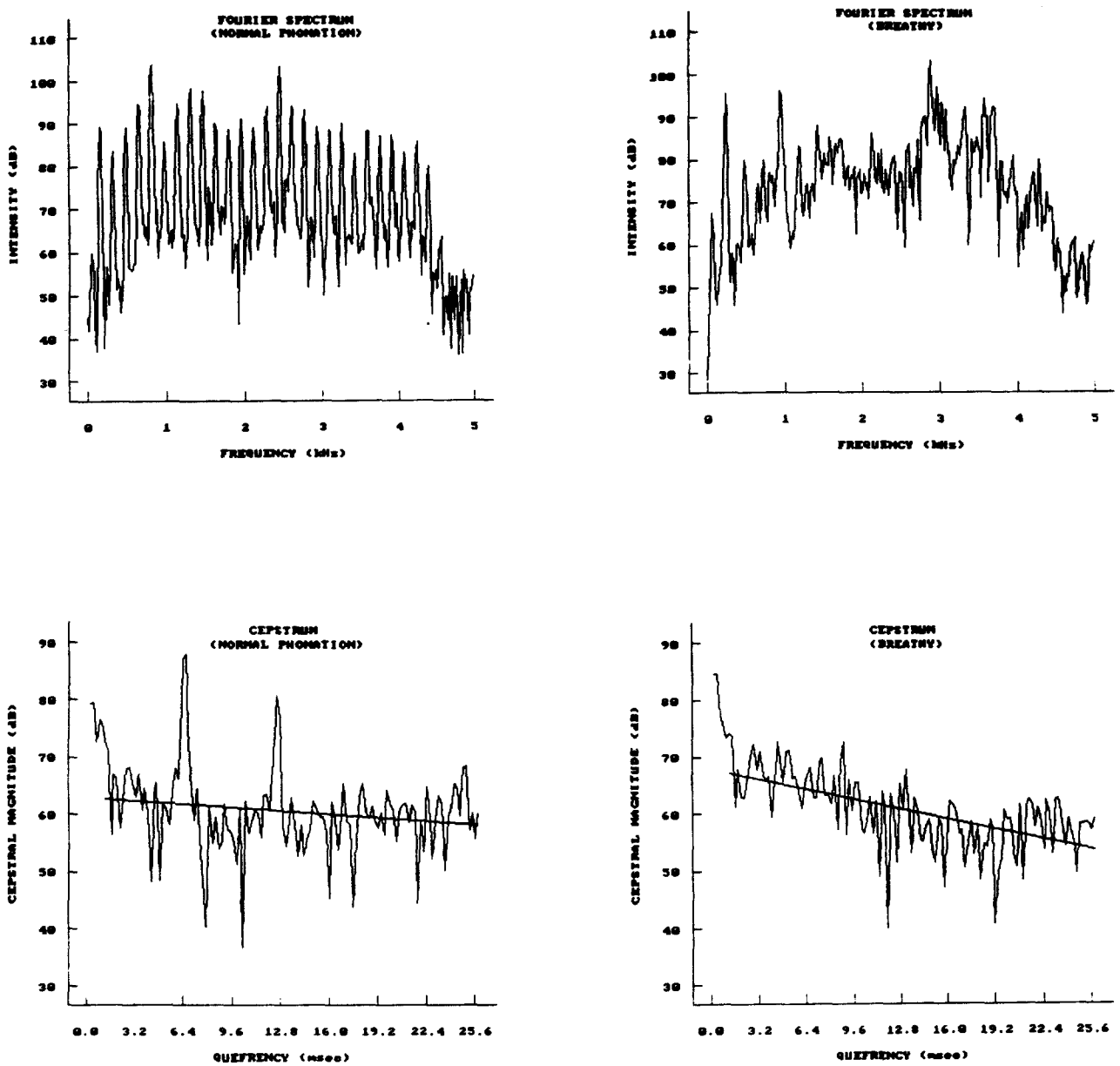

FIGURE 1. Fourier power spectra and cepstral representations for normally phonated and breathy signals. The linear regression line relating quefrency to cepstral magnitude is used in the CPP measure to normalize the cepstral peak for overall amplitude (see text).

CPP (Cepstral Peak Prominence): This is a measure of cepstral peak amplitude normalized for overall amplitude. Figure 1 shows examples of Fourier spectra and cepstral analyses from normally phonated and breathy signals. The cepstral peak in both cases corresponds to the fundamental period. The idea behind the CPP measure is that a highly periodic signal should show a well defined harmonic structure and, consequently, a more prominent cepstral peak than a less periodic signal. What is needed is a measure of the *prominence* of the peak rather than its absolute amplitude. This is because the amplitude of the cepstral peak is affected not only by the degree of periodicity but also by overall energy and the window size of the cepstrum analysis. Any number of methods might be used to normalize for overall amplitude. The method that we adopted involved fitting a linear regression line relating quefrency to cepstral magnitude. The line is computed between 1 msec and the maximum quefrency. (Quefrencies below 1 msec are ignored since these components are primarily sensitive to the envelope of the spectrum shape rather than signal periodicity.) The CPP measure is the difference in amplitude between the cepstral peak and the corresponding value on the regression line that is directly below the peak (i.e., the predicted cepstral magnitude for the quefrency at the cepstral peak). In other words, the CPP measure represents how far the cepstral peak emerges from the cepstral "background noise." Since breathy signals often retain a good deal of periodicity in the low frequencies, the CPP measure was computed not only

from unfiltered signals but also from signals that were: (a) bandpass filtered between 2.5 and 3.5 kHz, and (b) highpass filtered at 2.5 kHz. The measure was made every 10 msec using a 25.6 msec analysis window. The CPP measure for a signal was averaged over all analysis frames.

It is important to note that this method is fully automatic, and no attempt was made to verify that the peak in the cepstrum actually corresponded to the fundamental period. Comparison of the fundamental period extracted using this method with hand measurements of the same signals showed that the technique frequently produced pitch tracking errors, especially for the breathy signals. As will be noted below, this method provided accurate predictions of breathiness ratings in spite of these errors.

RPK (Pearson r at Autocorrelation Peak): This measure is based on a standard autocorrelation pitch tracker (Hillenbrand, 1988b). The tracker computes the correlation between the signal and a delayed version of itself at delays between the minimum expected fundamental period and the maximum expected fundamental period (period limits of 3.3 and 16.7 msec were used in the present study). For periodic signals, a peak occurs in the autocorrelation function at a delay corresponding to the fundamental period.

The idea behind the RPK measure is that highly periodic signals should show more prominent autocorrelation peaks than breathy signals. Since the correlation at each delay is calculated as the sum of the cross-product between the signal and the delayed copy, the peak amplitude will be strongly affected by overall signal amplitude and autocorrelation window length. Consequently, as with the cepstral measure, a method is needed to normalize for overall amplitude. A variety of amplitude normalization schemes were tested, including the regression-line method used in CPP. The method that was the most successful in predicting breathiness ratings involved computing a Pearson product-moment correlation coefficient between the signal and a delayed copy at a delay corresponding to the peak in the autocorrelation function. For an accurately tracked signal, this is simply a normalized measure of the degree of correlation between adjacent pitch pulses. The expectation was that adjacent periods would be less highly correlated for breathy signals. The RPK measure was calculated every 10 msec using a 30 msec analysis window. The RPK measure for a signal was averaged over all analysis frames. RPK was calculated from bandpass, highpass, and unfiltered signals. As with CPP, no attempt was made to correct tracking errors, which were quite common for breathy signals.

P/A (Peak-to-Average ratio): This very simple measure is based on a suggestion made by Klatt & Klatt (1990; see also Koike & Markel, 1975, and Davis, 1981, for related measures). P/A is the ratio of peak amplitude to average amplitude from full-wave rectified time-domain signals (see Figure 2). The expectation was that highly periodic signals would show relatively large peak-to-average ratios. P/A measures were averaged over successive nonoverlapping 10 msec segments of bandpass, highpass, and unfiltered signals.

BRI (Breathiness Index): This is a slightly modified version of the spectral tilt measure described by Fukazawa et al. (1988). The measure was calculated every 10 msec using a 25.6 msec analysis window. The BRI measure for a signal was averaged over all analysis frames. (The Fukazawa et al. implementation made a single spectral tilt measure at approximately the center of the signal.) The expectation was that breathy signals would show more high frequency energy and, consequently, larger BRI values. The measure was calculated from unfiltered signals only.

H/L (ratio of High- to Mid/Low-Frequency energy): This a measure of the average spectral energy at or above 4 kHz to the average energy below 4 kHz. The energies were calculated from 128-point (6.4 msec) Fourier spectra computed every 3.2 msec. The H/L measure for a signal was averaged over all analysis frames, using unfiltered signals.

H1A (First Harmonic Amplitude): This measure is simply the dB amplitude of the first harmonic relative to the second harmonic. The measure was made by visual inspection of 512-point (25.6 msec) Fourier spectra taken at approximately the center of the signal (see Figure 3).

Conspicuously absent from the list above are the extensively studied measures of jitter (Lieberman, 1963), shimmer (e.g., Horii, 1980), and harmonics-to-noise ratio (Yumoto, Gould, & Baer, 1982). There were several reasons for this decision. First, as was pointed out by Klatt and Klatt (1990), measures such as these are primarily influenced by low-frequency signal components that tend to dominate waveform characteristics because of their relatively high energy. Aspiration noise, on the other hand, is most prominent in the mid and high frequencies. There is also evidence from synthesis studies that variations in jitter produce a sensation of roughness rather than breathiness (Hillenbrand, 1988; Wendahl, 1966a, 1966b). Finally, the problem of reliably and automatically identifying the oscillographic landmarks used by these measures in noisy voice signals is not trivial, and it is known that these measures can be quite sensitive to relatively small errors in detecting these landmarks (Hillenbrand, 1987).

Listener Reliability

Intra-judge reliability was assessed by measuring the correlation between ratings in the first listening session and ratings for the same signals in the second listening session. As can be seen in Table 1, correlations between the two sessions were relatively high, with an average session-to-session correlation of 0.91.

Between-subject reliability was computed using Cronbach's Coefficient Alpha (Cronbach, 1951). This method involves measuring the correlation between each individual listener's mean rating for each stimulus with the group mean of all the other listeners. Table 2 summarizes the results. The correlations ranged from 0.88 to 0.98 with a mean of 0.95.

Effects of Phonation Type, Vowel, and Gender

An analysis of variance was performed in order to investigate the effects of talker gender, phonation type, and vowel ([a], [æ], [i], or [o]) on breathiness ratings. There was a highly significant effect for phonation type [F (2,156) = 330.94, p = 0.0001], no effect for vowel [F (3,156) = 1.31, p NS], a significant effect for gender [F (2, 156) = 4.83, p 0.05], and

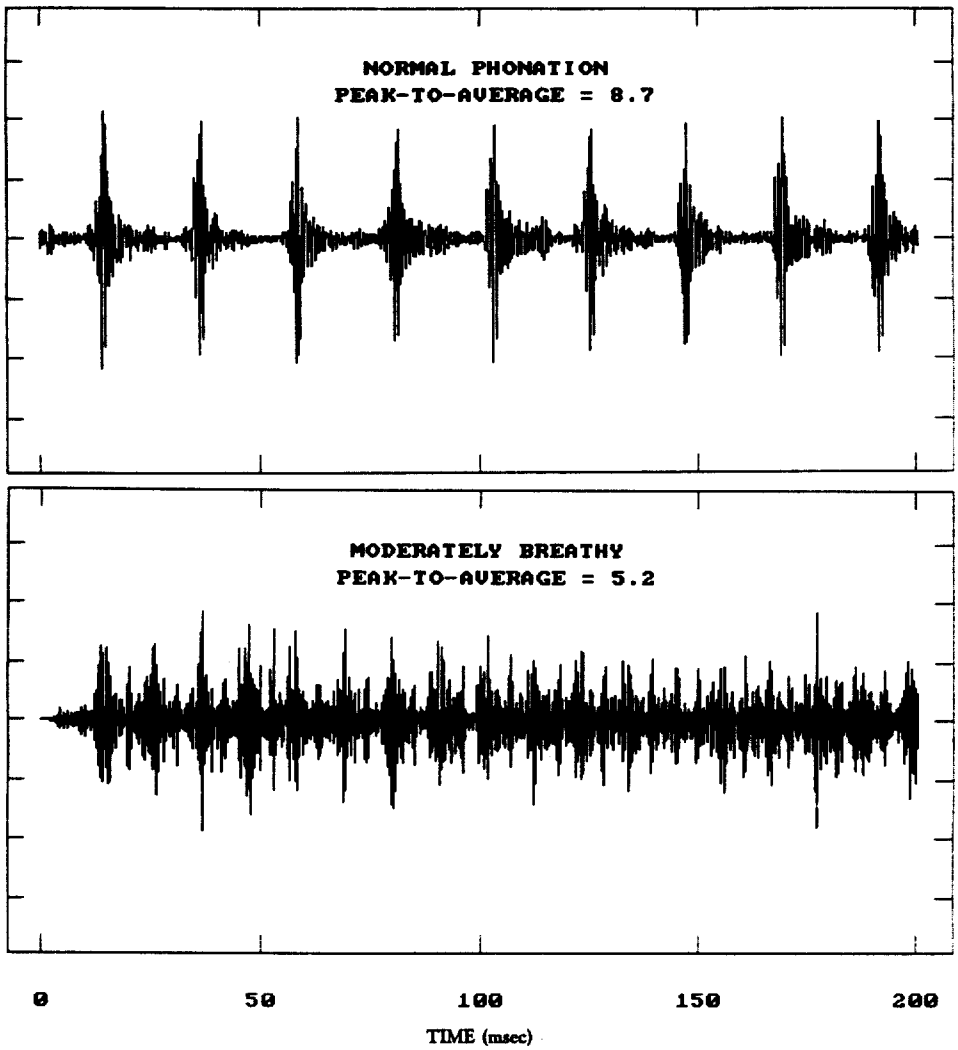

FIGURE 2. Bandpass filtered signals (2.5–3.5 kHz) from normally phonated and breathy vowels.

no significant interactions. The gender effect was due to higher breathiness ratings for men than women. However, additional analyses found no differences between men and women in breathiness ratings for normally phonated or moderately breathy signals (cf. Klatt & Klatt, 1990). Consequently, the gender effect simply indicates that men tended to produce higher levels of simulated breathiness than women for the very breathy condition. It might also be the case that listeners allowed for greater breathiness in female talkers, whereas the breathy male voice was perceived as being more abnormal.

Prediction of Breathiness Ratings

Table 3 shows the correlation of each measure with breathiness ratings and with all other acoustic measures. *Squared* correlations between the acoustic measures and the breathiness ratings are displayed in Figure 4. All correlations with breathiness ratings were significant at the 0.05 level or better. It can be seen that the cepstrum- and autocorrelation-based measures of signal periodicity accounted for a large proportion of the variance in breathiness ratings. The autocorrelation measure worked well only with filtered signals, while the CPP measure produced accurate predictions of breathiness ratings for unfiltered signals as well. For both of these measures it appears not to matter whether a bandpass or highpass filter is used. Breathiness predictions with CPP and RPK were somewhat more accurate than the predictions reported by Klatt & Klatt (1990) for a subjective visual method of rating the degree of periodicity in bandpass filtered signals.

The very simple peak-to-average measurement suggested by Klatt & Klatt (1990) correlated weakly with breathiness ratings. Examination of the scatter plot showed some indication of a nonlinear relationship between P/A and breathiness ratings. A log transform of the P/A measures resulted in slight improvements in prediction accuracy for all three types of signals, but prediction accuracy remained considerably lower than that for CPP and RPK.

Contrary to the findings of Fukazawa et al. (1988), spectral tilt correlated weakly with breathiness ratings. This was true both for BRI and H/L. This finding is in agreement with Klatt

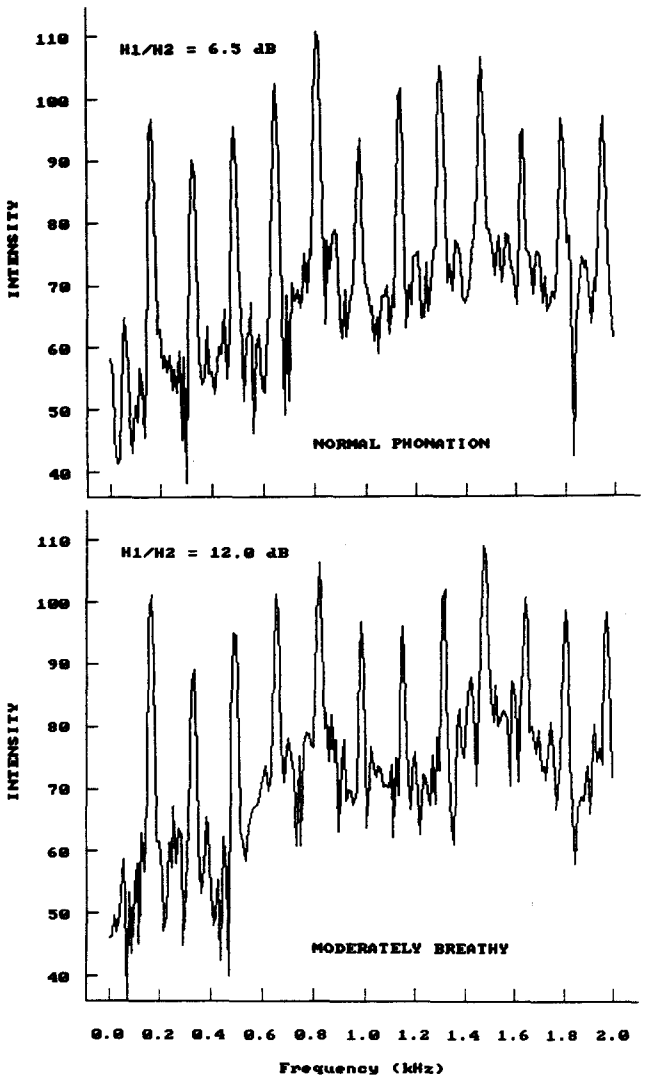

FIGURE 3. Sample spectra showing H1/H2 ratios for normal and moderately breathy phonation.

& Klatt (1990) who reported weak correlations between breathiness ratings and two measures of spectral tilt. We experimented with several variations of both BRI and H/L with limited success. For example, using a highpass filter to remove the fundamental frequency resulted in a slight improvement in prediction accuracy for BRI. We also experimented with a variety of frequency bands for the H/L calculation (including several bands that excluded the low frequency components of the spectrum) but did not find anything that worked better than the bands described previously. For the set of signals used in this study we consider it unlikely that additional tinkering with the measures would result in significant improvements in prediction accuracy. Examination of long-term average spectra for these signals simply did not show consistent breathiness-related differences in spectral tilt.

TABLE 1. Within-subject reliability measured as the correlation of average breathiness ratings between two listening sessions.

Subject	Pearson r	Subject	Pearson r
1	0.91	11	0.91
2	0.94	12	0.95
3	0.85	13	0.92
4	0.87	14	0.93
5	0.88	15	0.91
6	0.95	16	0.86
7	0.94	17	0.87
8	0.89	18	0.97
9	0.94	19	0.81
10	0.91	20	0.87

TABLE 2. Between-subject reliability measured with Cronbach coefficient alpha.

Subject	Correlation with set of all others	Subject	Correlation with set of all others
1	0.95	11	0.92
2	0.96	12	0.96
3	0.94	13	0.96
4	0.96	14	0.96
5	0.94	15	0.95
6	0.93	16	0.92
7	0.96	17	0.95
8	0.95	18	0.98
9	0.97	19	0.88
10	0.96	20	0.95

The measure of first harmonic amplitude correlated moderately with breathiness ratings. The 0.66 correlation is somewhat lower than the 0.83 correlation reported by Klatt & Klatt (1990). This might be related to the fact that Klatt & Klatt's talker group contained more women (n = 10) than men (n = 6). Further analysis of our measurements showed that H1A produced somewhat better predictions of breathiness ratings for women than men. We experimented with several methods for representing H1 amplitude (e.g., H1 amplitude relative to F1 amplitude and H1 amplitude relative to overall amplitude) but did not find anything that worked better than H1 relative to H2.

A stepwise multiple regression (SPSS-X, Version 4.00) was performed to determine if some linear combination of acoustic variables would improve prediction accuracy. Results are shown in Table 4. It can be seen that combining the acoustic measures results in relatively modest improvements in prediction accuracy. This is not surprising given the large proportion of the variance accounted for by CPP and RPK and the strong correlations between these two measures.

Discussion

Listening Experiment

Our results are in close agreement with those of Klatt & Klatt (1990), who reported that the best predictors of breathiness ratings were measures of: (a) periodicity in band-limited signals, and (b) first harmonic amplitude. The CPP and RPK measures are essentially automated versions of the subjective visual rating method used by Klatt & Klatt to represent periodicity. The CPP measure is similar in principle to a cepstrum-based signal-to-noise ratio calculation described

TABLE 3. Intercorrelation matrix for acoustic measures.* All correlations with breathiness ratings are significant at 0.05 or better.

	BR	CPP	CPPb	CPPh	RPK	RPKb	RPKh	P/A	P/Ab	P/Ah	BRI	H/L	H1A
BR	—	−0.92	−0.90	−0.89	−0.54	−0.91	−0.89	−0.30	−0.54	−0.58	0.41	0.51	0.66
CPP		—	0.98	0.98	0.42	0.92	0.91	0.32	0.51	0.56	−0.23	−0.37	−0.58
CPPb			—	0.98	0.38	0.93	0.91	0.30	0.46	0.52	−0.24	−0.40	−0.52
CPPh				—	0.36	0.93	0.92	0.37	0.53	0.59	−0.22	−0.39	−0.53
RPK					—	0.39	0.36	−0.21	0.22	0.23	−0.28	−0.24	−0.31
RPKb						—	0.97	0.36	0.55	0.59	−0.25	−0.42	−0.57
RPKh							—	0.40	0.57	0.60	−0.28	−0.48	−0.58
P/A								—	0.48	0.48	−0.04	−0.19	−0.33
P/Ab									—	0.95	−0.19	−0.32	−0.43
P/Ah										—	−0.14	−0.27	−0.42
BRI											—	0.81	0.20
H/L												—	0.25
H1A													—

*BR = breathiness rating, CPP = cepstral peak prominence, RPK = Pearson r at autocorrelation peak, P/A = peak-to-average ratio, BRI = breathiness index, H/L = ratio of high- to mid/low-frequency energy, H1A = first harmonic amplitude; b = bandpass filtered between 2.5 and 3.5 kHz; h = highpass filtered at 2.5 kHz.

recently by de Krom (1993), but is considerably simpler. It is somewhat surprising that CPP and RPK performed as well as they did since, as noted earlier, the very simple pitch trackers upon which they are based are not particularly accurate in identifying the fundamental period for breathy signals that are band-limited well above the fundamental frequency. In pilot work not reported here we attempted to improve the accuracy of the tracker in locating the fundamental period by constraining the search for the cepstral or autocorrelation peaks to relatively narrow limits around the average fundamental period determined from hand measurements. For reasons that are not clear, CPP and RPK measures obtained in this way were somewhat more weakly correlated with breathiness ratings than were measures obtained by the unconstrained method. Although we do not have an explanation for this finding, it is encouraging that these techniques do not depend on accurate tracking of the fundamental period. This is quite important since reliable measurement of fundamental frequency for marginally periodic signals is an extremely difficult and quite possibly unresolvable problem.

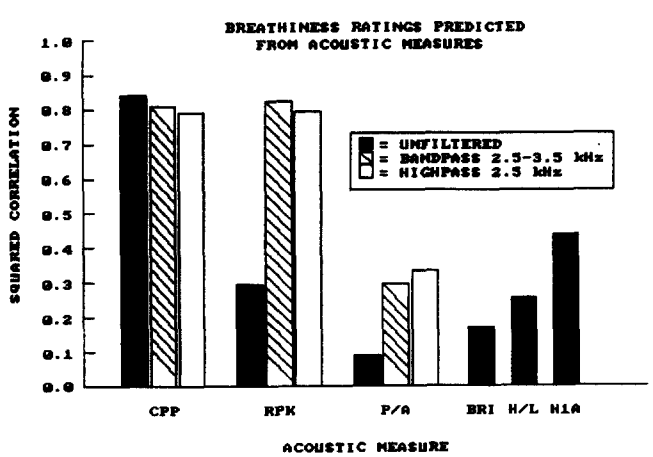

FIGURE 4. Squared correlation coefficients between acoustic measures and breathiness ratings.

It should be noted that the present results are purely descriptive and do not provide any direct evidence about the cues used by listeners in making breathiness judgments. However, our findings are quite consistent with the synthesis study of Klatt & Klatt (1990), which demonstrated that listener judgments of breathiness were more strongly influenced by aspiration noise than first harmonic amplitude. This conclusion is at odds with that reached by Bickley (1982) and Fischer-Jorgensen (1967). A detailed discussion of these conflicting findings can be found in Klatt & Klatt (1990).

Potential clinical applications of these techniques include the use of such acoustic measures as: (a) visual feedback in the treatment of vocal pathologies involving breathy voice, (b) a method of tracking progress throughout the course of a treatment program, and (c) indices to evaluate the relative effectiveness of alternative treatment approaches. Viewed in relation to these kinds of applications the CPP and RPK measures appear to be the most promising, in part because they provide accurate predictions of perceived breathiness, but also because they operate automatically and appear to be insensitive to pitch tracking errors. Although H1A correlated moderately with breathiness ratings, this measure appears to be much less promising as a clinical tool because of the difficulty of producing a reliable automatic algorithm that is capable of handling signals that are marginally periodic. We recently developed a semiautomatic algorithm that produces accurate measurements of H1A when given an estimate of fundamental frequency. However, the measure-

TABLE 4. Stepwise multiple regression analysis showing the prediction of breathiness ratings from various combinations of acoustic measures.

Step	Acoustic measure	Variance explained
1	CPP	0.84
2	BRI	0.88
3	H1A	0.90
4	RPK	0.92
5	RPK-BP	0.94

ments become unreliable if the estimated fundamental frequency is off by more than about 5%.

Development of an acoustically based feedback device would obviously require that the signal processing algorithms operate in real time. While the software used in this study operates in many times real time, none of the techniques are particularly computationally intensive. Real-time implementations of these algorithms are well within the capabilities of current digital signal processing hardware.

An important limitation of the present findings concerns the fact that the signals were produced by normal talkers who were simulating various levels of breathiness. The advantage of this method is that it ensured that the signals varied primarily in a single voice quality dimension. It was almost certainly the nearly univariate nature of the stimulus variations, in addition to the broad range of breathiness percepts, that accounted for the high degree of reliability in the listener ratings of breathiness. However, it remains to be determined whether the acoustic methods that were used in this study will provide accurate predictions of perceived breathiness in naturally occurring rather than simulated breathiness. Naturally occurring breathy voices will almost certainly present a more complicated picture, particularly in the voices of dysphonic speakers, which are seldom simply breathy and often show rather complex laryngeal vibratory patterns (e.g., von Leden, Moore, & Timcke, 1960).

A second limitation, common to many studies in this area, is the exclusive reliance on monotone, sustained vowels. The use of these simple signals greatly simplifies the measurement of phenomena such as aspiration noise since it is not necessary to separate waveform and spectral changes that are due to aspiration from those that might be due to intonation or supraglottal articulatory dynamics. However, since breathiness is quite easy to detect auditorily in continuous speech, it is clear that perceptual mechanisms exist to extract this quality from more complex speech signals. A challenge for future research will be to extend acoustic methods developed for sustained vowels to continuous speech.

Acknowledgments

We are grateful to William Dawson for his technical help with the instrumentation used to record the voice samples. This work was supported by a research grant from the National Institutes of Health (1-R01-DC01661).

References

Aronson, A. E. (1971). Early motor unit disease masquerading as psychogenic breathy dysphonia: A clinical case presentation. *Journal of Speech and Hearing Disorders, 36,* 115–124.

Aronson, A. E. (1990). *Clinical voice disorders* (3rd ed.). New York: Thieme.

Bickley, C. (1982). *Acoustic analysis and perception of breathy vowels.* (Speech Communication Group Working Papers I). Cambridge, MA.: Research Laboratory of Electronics.

Boone, D. R., & McFarlane, S. C. (1988). *The voice and voice therapy.* 4th ed. Englewood Cliffs, NJ: Prentice Hall.

Colton, R. A., & Casper, J. K. (1990). *Understanding voice problems: A physiological perspective for diagnosis and treatment.* Baltimore: Williams and Wilkins.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrica, 16,* 297–334.

Davis, S. B. (1981). Acoustic characteristics of normal and pathological voices. *ASHA Reports, 11,* 97–115.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. *Journal of Speech and Hearing Research, 36,* 254–266.

Fairbanks, G. (1940). *Voice and articulation drillbook.* New York: Harper and Brothers.

Fischer-Jorgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics, 28,* 71–139.

Frokjaer-Jensen, B., & Prytz, S. (1976). Registration of voice quality. *Bruel and Kjaer Technical Review, 3,* 3–17.

Fukazawa, T., El-Assuooty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustical Society of America, 83,* 1189–1192.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research, 30,* 448–461.

Hillenbrand, J. (1988a). Perception of aperiodicities in synthetically generated voices. *Journal of the Acoustical Society of America, 83,* 2361–2371.

Hillenbrand, J. (1988b). *MPITCH: An autocorrelation fundamental-frequency tracker.* [Computer Program]. Kalamazoo, MI.: Western Michigan University.

Hillenbrand, J., Metz, D. E., Colton, R. A., & Whitehead, R. L. (1990). *A high-speed film and acoustic study of breathy voice.* Paper presented at the meeting of the American Speech-Language-Hearing Association, Seattle, WA.

Hollien, H. (1987). "Old voices": What do we really know about them? *Journal of Voice, 1,* 2–17.

Horii, Y. (1980). Vocal shimmer in sustained phonation. *Journal of Speech and Hearing Research, 23,* 202–209.

Horii, Y. (1983). Automatic analysis of voice fundamental frequency and intensity using a Visi-Pitch. *Journal of Speech and Hearing Research, 26,* 467–471.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America, 81,* 495–504.

Kasuya, H., Ogawa, S., Kazuhiko, M., & Satoshi, E. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. *Journal of the Acoustical Society of America, 80,* 1329–1334.

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

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

Koike, Y., & Markel, J. D. (1975). Application of inverse filtering for detecting laryngeal pathology. *Annals of Otology, Rhinology and Laryngology, 84,* 117–124.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351–360). San Diego: College Hill.

Ladefoged, P., & Antonanzas-Barroso, N. (1985). *Computer measures of breathy phonation.* (Working Papers in Phonetics, 61). University of California, Los Angeles, 79–86.

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America, 35,* 344–353.

McKay, I. R. A. (1987). *Phonetics: The science of speech production* (2nd ed.). Boston: College Hill.

Parsons, T. W. (1987). *Voice and speech processing.* New York: McGraw-Hill.

Ryan, W. J., & Burk, K. W. (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders, 7,* 181–192.

von Leden, H., Moore, P., & Timcke, R. (1960). Laryngeal vibra-

tions: Measurements of the glottic wave. *AMA Archives of Otolaryngology, 71,* 26–45.

Wendahl, R. (1966a). Some parameters of auditory roughness. *Folia Phoniatrica, 18,* 26–32.

Wendahl, R. (1966b). Laryngeal analog synthesis of jitter and shimmer: Auditory parameters of harshness. *Folia Phoniatrica, 18,* 98–108.

Witten, I. H. (1982). *Principles of Computer Speech.* London: Academic Press.

Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America, 71,* 1544–1550.

Zemlin, W. (1968). *Speech and hearing science: Anatomy and physiology.* Englewood Cliffs, NJ: Prentice-Hall.

Received September 27, 1993
Accepted February 28, 1994

Contact author: James Hillenbrand, Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI 49008.