

Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech

James Hillenbrand
Western Michigan University
Kalamazoo

Robert A. Houde
RIT Research Corporation
Rochester, NY

In an earlier study, we evaluated the effectiveness of several acoustic measures in predicting breathiness ratings for sustained vowels spoken by nonpathological talkers who were asked to produce nonbreathy, moderately breathy, and very breathy phonation (Hillenbrand, Cleveland, & Erickson, 1994). The purpose of the present study was to extend these results to speakers with laryngeal pathologies and to conduct tests using connected speech in addition to sustained vowels. Breathiness ratings were obtained from a sustained vowel and a 12-word sentence spoken by 20 pathological and 5 nonpathological talkers. Acoustic measures were made of (a) signal periodicity, (b) first harmonic amplitude, and (c) spectral tilt. For the sustained vowels, a frequency domain measure of periodicity provided the most accurate predictions of perceived breathiness, accounting for 92% of the variance in breathiness ratings. The relative amplitude of the first harmonic and two measures of spectral tilt correlated moderately with breathiness ratings. For the sentences, both signal periodicity and spectral tilt provided accurate predictions of breathiness ratings, accounting for 70%–85% of the variance.

KEY WORDS: *breathy voice, voice disorders, acoustic analysis, dysphonia, periodicity*

Interest in the physiology, acoustics, and perception of breathy vocal quality arises from several sources. For example, breathy voice serves a contrastive function in the phonological systems of several languages (e.g., Fischer-Jorgensen, 1967; Huffman, 1987; Ladefoged, 1975, 1983). Breathiness is also a very common symptom of both organic and functional voice disorders (Aronson, 1971, 1990; Boone & McFarlane, 1988; Colton & Casper, 1990). There is also some evidence that breathiness is associated with aging (Hollien, 1987; Ryan & Burk, 1974), and among English speakers there appears to be some tendency for women to produce somewhat breathier voices than men (Klatt & Klatt, 1990; McKay, 1987).

Previous work on the acoustic correlates of breathy vocal quality has identified three major features that predict perceptual ratings of breathiness, with varying degrees of success: (a) the amplitude of the first harmonic (H1), (b) the degree of signal periodicity, and (c) spectral tilt (see Klatt & Klatt, 1990, for a very thorough review).

First harmonic amplitude. Several investigators have noted that glottal source functions associated with the production of breathy voice are more nearly sinusoidal than those associated with nonbreathy phonation (e.g., Bickley, 1982; Fischer-Jorgensen, 1967; Hillenbrand, Metz, Colton, & Whitehead, 1990; Huffman, 1987). The rounding of the glottal source function has been attributed to nonsimultaneous closure along the length of the vocal folds (Klatt & Klatt, 1990). As shown in Figure 1, the rounded glottal source function results in an acoustic signal with a relatively high amplitude first harmonic and relatively weak upper harmonics (Bickley, 1982; Fischer-Jorgensen, 1967; Hillenbrand et al., 1994; Huffman, 1987; Klatt & Klatt, 1990; Ladefoged, 1983; Ladefoged & Antonanzas-Barroso, 1985). The amplitude of

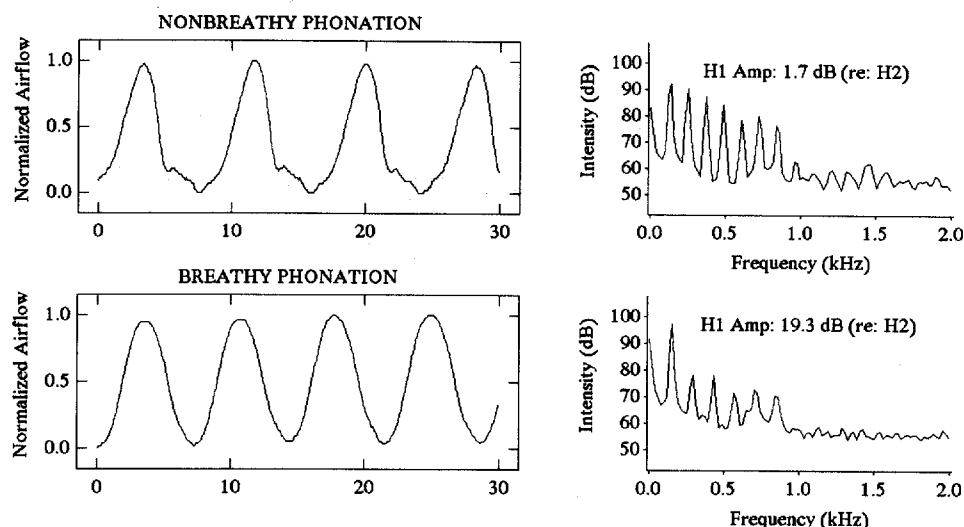


FIGURE 1. Airflow functions measured with a Rothenberg airflow mask and inverse filter (Rothenberg, 1973) for nonbreathy and breathy phonation (from Hillenbrand et al., 1990). Airflow measurements have been normalized between 0.0 (minimum flow) and 1.0 (maximum flow).

the first harmonic has been measured relative to the amplitude of the first formant (Fischer-Jorgensen, 1967) and relative to the amplitude of the second harmonic (e.g., Fischer-Jorgensen, 1967; Klatt & Klatt, 1990; Ladefoged & Antonanzas-Barroso, 1985). Evidence from some perceptual studies suggests that H1 amplitude is the primary cue to the perception of breathiness (Bickley, 1982; Ladefoged, 1983; Ladefoged & Antonanzas-Barroso, 1985), whereas evidence from other studies suggests that breathiness judgments are more strongly correlated with signal periodicity than H1 amplitude (Klatt & Klatt, 1990; Hillenbrand et al., 1994).

Signal periodicity. The turbulent airflow pattern that is associated with breathy voice results in an acoustic signal that tends to be less periodic than a nonbreathy voice. Because the periodic component of the voice source is inherently weak in the mid and high frequencies, and aspiration noise is inherently weak in the low frequencies, noise tends to be especially noticeable above about 2–3 kHz (Klatt & Klatt, 1990). The degree of signal periodicity can be measured directly from time-domain signals (e.g., Hillenbrand et al., 1994; Klatt & Klatt, 1990), or indirectly from frequency-domain representations by measuring the degree of harmonic organization (e.g., de Krom, 1993; Fischer-Jorgensen, 1967; Hillenbrand et al., 1994; Kasuya, Ogawa, Kazuhiko, & Satoshi, 1986). Some investigators have reported strong correlations between breathiness ratings and the degree of signal periodicity (Klatt & Klatt, 1990; Hillenbrand, 1988a; Hillenbrand et al., 1994), whereas others have reported negative results (Bickley, 1982; Fischer-Jorgensen, 1967).

Spectral tilt. Although the harmonic component of breathy signals tends to be relatively weak in high-frequency energy, the presence of aspiration noise, which is stronger in the mid and high frequencies than in the lows, can result in a voice signal that is richer in high-frequency energy than nonbreathy signals. Some investigators have reported strong correlations between breathiness ratings and spec-

tral tilt (Fukazawa, El-Assuootty, & Honjo, 1988; Klich, 1982), whereas other investigators have reported weak or nonsignificant correlations (Hillenbrand et al., 1994; Klatt & Klatt, 1990). Further, a study by Hillenbrand (1988a) using synthesized vowels found that breathiness ratings were affected only by the level of aspiration noise, with no independent effect for spectral tilt.

In a recent study (Hillenbrand et al., 1994) we measured correlations between breathiness ratings and 12 acoustic measures. Recordings were made from eight men and seven women with no evidence or history of voice disorders who were asked to produce nonbreathy, moderately breathy, and very breathy sustained vowels (/a/, /i/, /æ/, and /o/). Twenty listeners rated the degree of breathiness using a direct magnitude estimation procedure. The acoustic measures were designed to capture either H1 amplitude, signal periodicity, or spectral tilt, using methods described below. The results showed that measures of signal periodicity, based on either the time waveform or its spectrum, 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. However, there were two important limitations to that study. First, the test utterances were spoken by nonpathological talkers who were purposely producing nonbreathy, moderately breathy, and very breathy signals. It is not clear whether the results obtained with these signals apply to the breathy voices produced by pathological speakers. Second, as with many of the studies in this area, the signals consisted of monotone, sustained vowels. The relatively static nature of these signals greatly simplifies the measurement problems, but it is not clear that the results obtained with sustained vowels apply to more complicated connected speech samples. The purpose of the present study was to extend our previous work with the acoustic measures to both isolated vowels and connected

speech samples produced by pathological and nonpathological talkers.

Method

Voice Samples

Voice samples were taken from the Voice Disorders Database recorded at the Massachusetts Eye and Ear Infirmary and distributed by Kay Elemetrics Corporation. The database consists of sustained /a/ and Rainbow Passage (Fairbanks, 1940) recordings from approximately 700 subjects with a variety of organic and functional voice disorders, and from 55 nonpathological talkers. From this database, voice samples were selected from 20 talkers in the disordered group and 5 talkers in the nondisordered group. Because the goal of the study was the prediction of breathiness ratings, voice samples were selected which appeared to depart from normal voice quality primarily in the direction of breathiness. Utterances from the database were auditioned by a graduate student, who was instructed to select voice samples that represented a range of breathiness percepts from mild to severe, while avoiding samples that were largely or exclusively aphonic.¹ Summary information on the speakers in the pathological group is given in Table 1. The nonpathological talkers consisted of three men and two women between the ages of 28 and 55 who showed no

evidence or history of voice disorders and who were native speakers of English.

Voice samples from the database are available as digitized files on CD-ROM. The recordings were made in a sound-treated booth with a Sennheiser MKE2 microphone at approximately 15 cm from the lips and recorded on a Tascom DA-30 MK2 digital recorder. The signals were subsequently digitized using a Kay CSL system. Recordings from the disordered talkers were digitized at 25 kHz, whereas recordings from the nonpathological talkers were digitized at 50 kHz. To maintain a constant sample frequency, the 50 kHz recordings were digitally lowpass filtered at 10 kHz and downsampled to 25 kHz. The sustained /a/ samples are 1 s in duration for the disordered group and 3 s in duration for the nonpathological group. The 3-s samples were edited by hand to include only the first 1 s of phonation. To prevent onset and offset transients, all signals were ramped on and off with an inverted and lifted half-cosine function. The Rainbow Passage signals were edited by hand to include only the second sentence ("The rainbow is a division of white light into many beautiful colors.")

Breathiness Ratings

Twenty listeners (14 women, 6 men) were recruited from among graduate students and faculty in the Speech Pathology and Audiology Department at Western Michigan University. All listeners were native speakers of English. Signals were lowpass filtered at 10 kHz, amplified, and presented free field at approximately 80 dBA over a single loudspeaker (Boston Acoustics A60). Listeners were tested in a single session consisting of two blocks of trials. One block consisted of 10 randomly ordered presentations of the 25

¹The signals that were selected were subsequently auditioned by the first author, with the intent of making appropriate changes to meet the criteria listed above. As it turned out, the list submitted by the graduate student was judged to be appropriate, and no changes were made.

TABLE 1. Summary of subject information for speakers in the voice disordered group.

Subject number	Age	Sex	Diagnosis/Description
1	57	M	Hyperfunction; A-P compression; adduction/compression
2	48	M	Hyperfunction; A-P compression
3	66	F	Unilateral vocal fold paralysis; unilateral contact granuloma
4	35	M	Submucosal granuloma; over injection of teflon
5	38	F	Vocal nodules; hyperfunction; A-P compression
6	27	F	Hyperfunction; A-P compression
7	55	M	Unilateral vocal fold paralysis
8	80	M	Hyperfunction; ventricular compression; gastric reflux; bowing
9	64	F	Hemilaryngectomy
10	81	M	Unilateral vocal fold paralysis; laryngeal hyperfunction; A-P compression
11	42	F	Blunt trauma; hematoma; scarring; paralysis
12	49	F	Hyperfunction; A-P compression; scarring; vocal fold edema
13	69	M	Laryngeal hyperfunction; suspected unilateral T1 lesion; A-P compression
14	34	F	Hyperfunction; bilateral mild vocal fold edema; conversion dysphonia; A-P compression
15	23	M	Hyperfunction; unilateral scarring; laryngeal trauma
16	67	F	Unilateral vocal fold paralysis
17	46	F	Laryngeal unilateral left paralysis; post-teflon injection; A-P compression
18	72	F	Hyperfunction; vocal tremor; scarring; post-surgery for amalyoid disease; complete ventricular compression
19	52	M	Hyperfunction; malignant tumor
20	44	M	Hyperfunction; unknown lesion of outer third of left vocal fold; A-P compression

sustained vowels, and the other block consisted of 10 randomly ordered presentations of the 25 Rainbow Passage signals. The order of presentation of the two blocks was counterbalanced across subjects.

Listeners were asked to rate the amount of breathiness for each signal using a direct magnitude estimation procedure. Subjects were asked to "enter a large number if the signal is very breathy and a small number if the signal shows little or no breathiness." Subjects were allowed to repeat each signal as often as they wished before entering their responses on a computer terminal. Ratings were later linearly rescaled so that each listener's ratings for each 25-stimulus set ranged from 0.0 to 10.0.

Acoustic Analysis

The six acoustic measures are described briefly below. A more complete description can be found in Hillenbrand et al. (1994).

Cepstral peak prominence (CPP). A cepstrum is a log power spectrum of a log power spectrum. For periodic signals, the first power spectrum will show energy at harmonically related frequencies, and the second spectrum (i.e., the spectrum of the spectrum) will show a strong component corresponding to the regularity of the harmonic peaks. The time ("quefrency") at the cepstral peak corresponds to the fundamental period of the signal. A signal whose spectrum shows a well-defined harmonic structure will show a very prominent cepstral peak. CPP is a measure of the amplitude of the cepstral peak corresponding to the fundamental period, normalized for overall signal amplitude (see de Krom, 1993, for a related measure). A cepstral peak is first located between the minimum and maximum expected fundamental period (3.3–16.7 ms, or 60–300 Hz, for the present study). The quefrency at the cepstral peak generally corresponds to the fundamental period, and the amplitude of the cepstral peak reflects both the level of harmonic organization and the overall amplitude of the signal. To normalize for overall amplitude, a linear regression line is calculated relating quefrency to cepstral magnitude. The CPP measure is the difference in amplitude (in dB) between the cepstral peak and the corresponding value on the regression line that is directly below the peak (see Figure 2). The CPP measure was computed from unfiltered signals, from signals that were bandpass filtered between 2.5 and 3.5 kHz, and from signals that were highpass filtered at 2.5 kHz. The highpass and bandpass filtering conditions were included because of the common observation that aspiration noise tends to be more prominent in the mid and high frequencies (e.g., Klatt & Klatt, 1990). Filtering was accomplished with a fourth-order Butterworth digital filter. The measure was made every 10 ms using a 1024-point (41 ms) analysis window. The CPP measure for a signal was averaged over all analysis frames. Because CPP is a measure of the degree of the degree of harmonic organization, it was expected that breathy signals would tend to show smaller values of CPP than nonbreathy signals. No attempt was made to correct pitch tracking errors, which were quite common for breathy signals. Also, no attempt was made to separate voiced regions from unvoiced regions in the Rain-

bow Passage signals. The goal of this work is the development of signal processing methods that are fully automatic, and the techniques that are normally used to differentiate voiced and unvoiced regions (e.g., Parsons, 1986; Witten, 1982) are unlikely to work reliably for many dysphonic voices.

CPP-smoothed (CPPS). After a good deal of experimentation with the measures described in our earlier paper, it was discovered that a rather simple modification to the CPP algorithm produces a noticeable improvement in prediction accuracy. The additional processing step involves smoothing the individual cepstra before extracting the cepstral peak and calculating the peak prominence. With the CPPS method, the cepstra are calculated every 2 ms instead of every 10 ms, and the smoothing is carried out in two steps (see Figure 3). First, as shown in the middle panel of Figure 3, the cepstra are averaged across time; that is, each unsmoothed cepstrum is replaced by the average of some number of cepstral frames to the left of the current frame and some number of cepstral frames to the right of the current frame. For example, for a smoothing window of five frames, the smoothed output for a given frame would consist of the average of the current frame with the two previous frames and the two subsequent frames. In the next smoothing step, shown in the bottom panel of Figure 3, a running average of cepstral magnitude is calculated across quefrency. For example, for a five-bin averaging window, each cepstral magnitude is replaced by the average of the current bin with the two adjacent bins of lower quefrency and the two adjacent bins of higher quefrency. Experimentation with a variety of settings of smoothing windows showed that very large smoothing windows across time and relatively small smoothing windows across quefrency worked best for sustained vowels. Although much smaller time-smoothing windows worked nearly as well, the tests reported here used a 150-frame (300 ms) time-smoothing window followed by 3-bin quefrency smoothing. Because the fundamental frequency (F_0) varies over time in the Rainbow Passage sentences, much smaller time-smoothing windows are needed. For the present study, a 10-frame (20 ms) time-smoothing window was used, followed by 10-bin quefrency smoothing.

Pearson r at autocorrelation peak (RPK). A conventional autocorrelation pitch tracker (Hillenbrand, 1988b) is used to compute 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, again using period limits of 3.3 and 16.7 ms. The delay corresponding to the peak in the autocorrelation function generally corresponds to the fundamental period. As a measure of the degree of periodicity, a Pearson moment-product correlation coefficient is calculated between the signal and itself at a delay corresponding to the peak in the autocorrelation function. For an accurately tracked signal, this is simply an amplitude-normalized measure of the degree of correlation between adjacent pitch pulses. A perfectly periodic signal would show a correlation of 1.0, and the correlation should decrease systematically with increasing departures from perfect periodicity. RPK was calculated every 10 ms using a 30-ms analysis window,

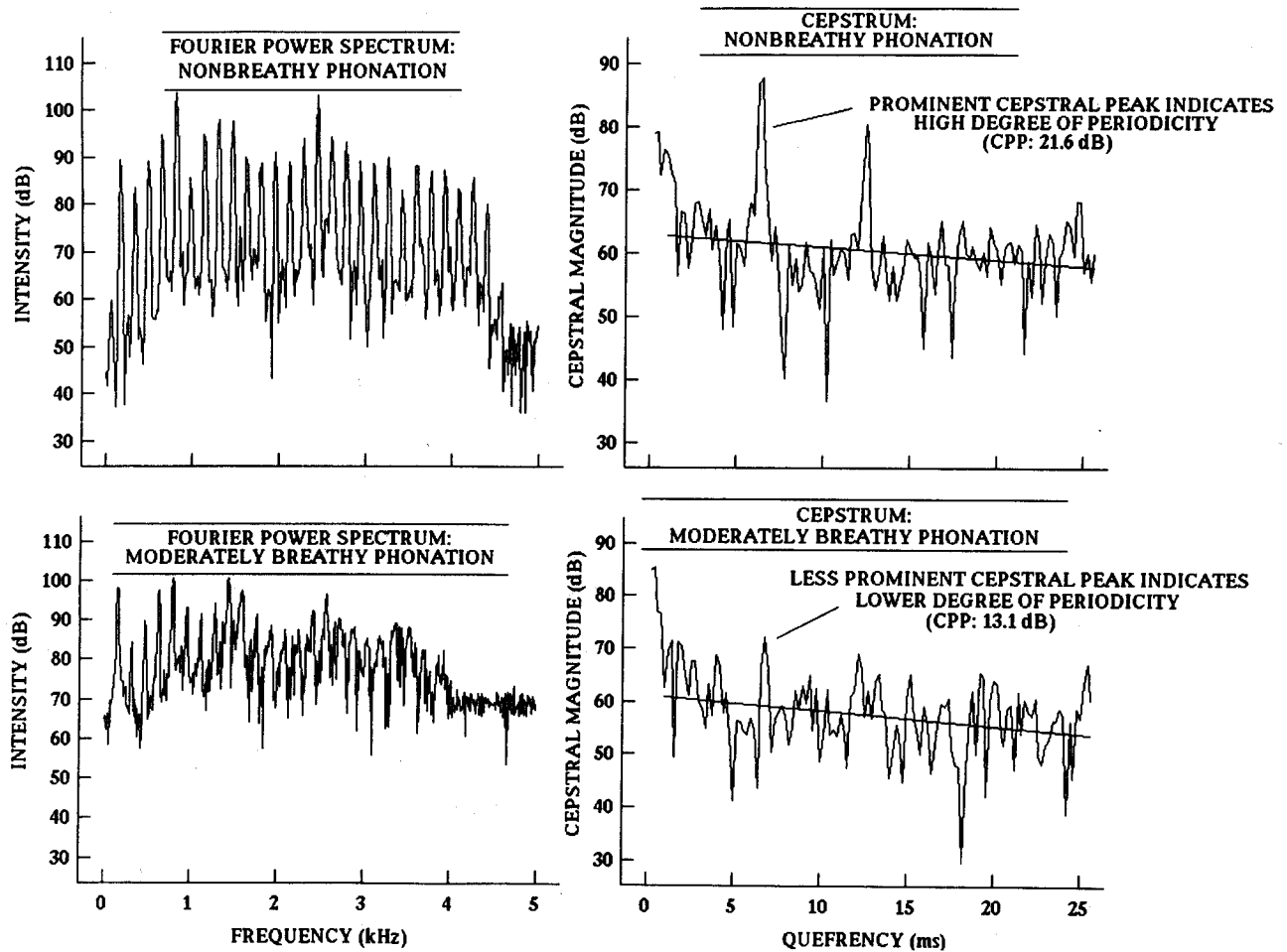


FIGURE 2. Fourier power spectra and cepstral representations for nonbreathy and moderately breathy signals. The nonbreathy signal received a rating of 0.8 on a 0–10 rating scale, and the moderately breathy signal received a rating of 4.0. The linear regression line relating quefrency to cepstral magnitude is used in the CPP measure to normalize the cepstral peak for overall amplitude. CPP is the difference in amplitude between the cepstral peak and the value on the regression line that is directly below the peak.

and the RPK measure for a signal was averaged over all analysis frames. RPK was calculated from bandpass, high-pass, and unfiltered signals. As with CPP, no attempt was made to correct pitch-tracking errors, and no attempt was made to differentiate voiced and unvoiced regions of the Rainbow Passage. Since RPK is a measure of the degree of signal periodicity, it was expected that breathy signals would tend to show smaller values of RPK than nonbreathy signals.

Peak-to-average ratio (P/A). As suggested by Klatt & Klatt (1990), P/A is the ratio of peak amplitude to average amplitude from full-wave rectified time-domain signals (see Davis, 1981, and Koike & Markel, 1975, for related measures). On the basis of Klatt & Klatt's observations, it was expected that highly periodic signals would show larger peak-to-average ratios than breathy signals (see Figure 4). P/A measures were averaged over successive nonoverlapping 10-ms segments of bandpass, highpass, and unfiltered signals.

Breathiness index (BRI). This is a slightly modified version of the overall spectral tilt measure described by Fukazawa et al. (1988). BRI is defined as the ratio of the energy in

the second derivative of a signal to the energy in a non-derived signal. The index is conceptually similar to a simple ratio of energy in high- and low-frequency bands in the spectrum (see the H/L measure below), except that the Fukazawa et al. technique produces a global measure of spectral tilt which does not require arbitrary cutoff frequencies. Large values of BRI indicate a relatively large amount of high frequency energy, so breathy signals should tend to show large BRI values. The measure was calculated every 10 ms using a 512-point (20.5 ms) 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 measure was calculated from unfiltered signals only.

Ratio of high- to mid/low-frequency energy (H/L). This is a measure of the average spectral energy at or above 4 kHz to the average energy below 4 kHz. The expectation was that breathy signals would tend to show larger values of H/L. The energies were calculated from 128-point (5.1 ms) Fourier spectra computed every 2.56 ms. The H/L measure for a signal was averaged over all analysis frames, using unfiltered signals.

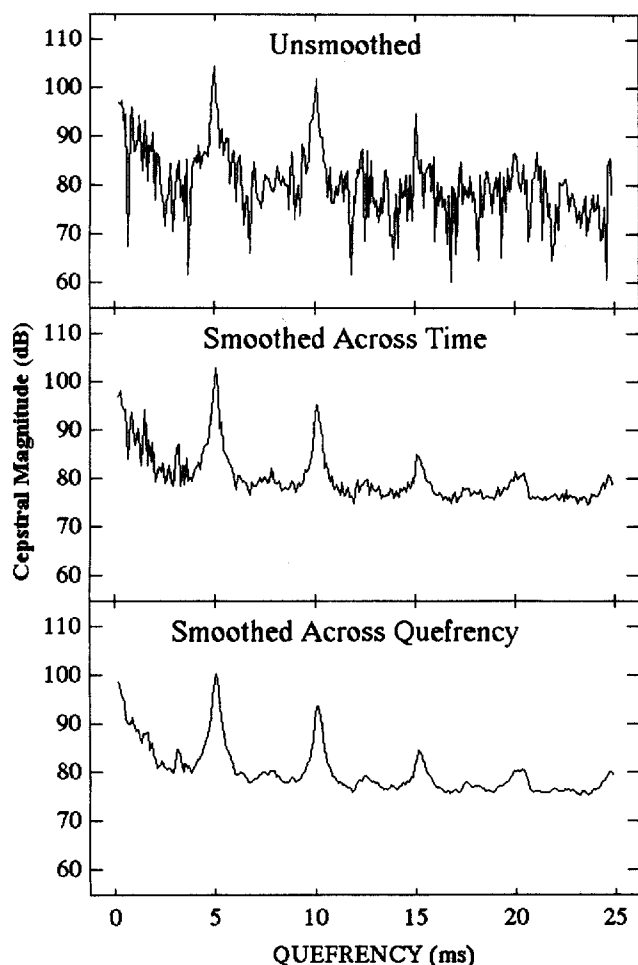


FIGURE 3. Illustration of the smoothing method used in CPPS. The top panel shows the original, unsmoothed cepstrum; the middle panel shows the cepstrum after smoothing across time; and the bottom panel shows the cepstrum after smoothing across quefrency.

First harmonic amplitude (H1A). This measure is simply the amplitude of the first harmonic (in dB) relative to the second harmonic. For the isolated vowels, an automatic method was used that is different from the manual method described in Hillenbrand et al. (1994). The median fundamental frequency is first measured with an autocorrelation pitch tracker. Fourier spectra are then calculated every 10 ms using a 1024-point analysis window. For each spectrum analysis frame, the algorithm attempts to locate (a) a harmonic peak corresponding to H1 that is within 5% of the median fundamental frequency, and (b) a harmonic peak corresponding to H2 that is within 5% of $2 \times \text{H1}$. If either of these harmonic peaks cannot be located within the specified frequency bands, the analysis frame is simply skipped. The H1A measure for a signal is the average H1 amplitude in dB (re: H2) averaged over all frames with valid measures of H1 and H2. Because this automatic method works well only with a relatively steady F_0 , the manual method used in our previous study was adopted for the Rainbow Passage signals. We calculated 1024-point Fourier spectra at approximately the center of the vocalic nuclei from four syllables: /e/ in *rainbow*, /l/ in *division*, /a/ in *white*, and /ʌ/ in *colors*. H1 and H2 were located visually and their amplitudes were measured using a custom editing tool. The H1A measure for these signals was the average first harmonic amplitude sampled at these four points. The expectation was that breathy signals would show large values of H1A.

bles: /e/ in *rainbow*, /l/ in *division*, /a/ in *white*, and /ʌ/ in *colors*. H1 and H2 were located visually and their amplitudes were measured using a custom editing tool. The H1A measure for these signals was the average first harmonic amplitude sampled at these four points. The expectation was that breathy signals would show large values of H1A.

Results

Listener Reliability

Interjudge reliability was computed using Cronbach's Coefficient Alpha (Cronbach, 1951), which involves measuring the correlation between each individual listener's mean rating for each stimulus and the group mean of all the other listeners. As can be seen in Table 2, interjudge reliability was quite high for both sets of test signals. Although the difference is rather small, the reliability coefficients were significantly higher for the sustained vowels than for the Rainbow Passage signals ($t = 2.48$, $df = 19$, $p < 0.05$). The average Pearson r relating each listener's ratings with those of each other listener was 0.92 for the sustained vowels, and 0.87 for the Rainbow Passage Sentences.

Prediction of Breathiness Ratings

Sustained vowels. Table 3 shows the correlation of each measure with breathiness ratings and with all other acoustic measures for the sustained vowels. For comparison, Table 4 shows correlations between breathiness ratings and acoustic measures from the present study and from our earlier study of nondisordered talkers. Squared correlations between the acoustic measures and the breathiness ratings are displayed in Figure 5. The overall pattern of results obtained with the sustained vowels is quite similar to our previous study using nondisordered talkers. The best predictors of breathiness ratings were the three measures of signal periodicity: RPK, CPP, and CPPS. The periodicity measures worked somewhat better with unfiltered signals than with bandpass and highpass filtered signals. In our previous study the prediction accuracy of CPP was affected only slightly by filtering, but there was a large improvement in the accuracy of RPK predictions when filtered signals were used. It can also be seen that the version of CPP with smoothing provided more accurate predictions of breathiness ratings than the original implementation, with the unfiltered version accounting for 92% of the variance. Consistent with our previous study, the peak-to-average measure correlated weakly with breathiness ratings, and moderate correlations were seen for H1 amplitude and the two measures of spectral tilt.

Rainbow Passage. Table 5 shows the correlation of each measure with breathiness ratings and with all other acoustic measures for the Rainbow Passage signals. Squared correlations between the acoustic measures and the breathiness ratings are shown in Figure 6. As with the sustained vowels, the measures of periodicity accounted for approximately 70%–85% of the variance in breathiness ratings. Three of the measures, however, produced substantially more accu-

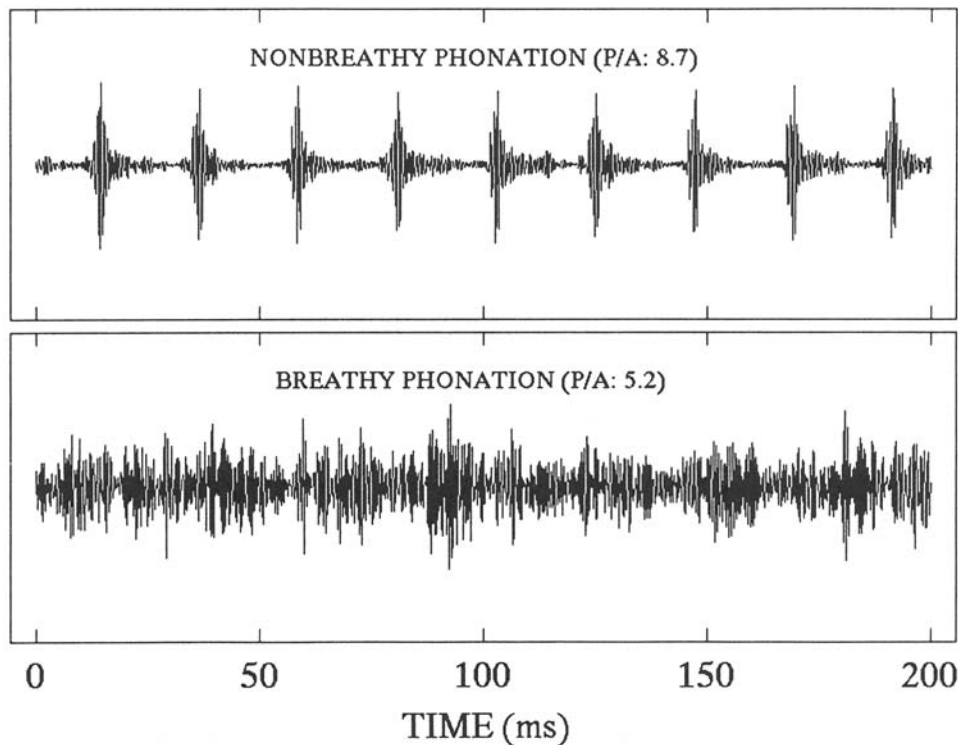


FIGURE 4. Bandpass filtered signals (2.5–3.5 kHz) from nonbreathy and breathy vowels.

rate predictions of breathiness ratings for the Rainbow Passage signals than for the sustained vowels. The peak-to-average measure accounted for just under 50% of the variance for the bandpass-filtered Rainbow Passage signals, but only 18%–24% of the variance for the filtered sustained vowels. Similarly, the two measures of spectral tilt accounted for approximately 70% of the variance in breathi-

ness ratings for the Rainbow Passage signals, compared to approximately 40% of the variance for the sustained vowels.

Discussion

To summarize briefly, this study was a direct extension of an earlier experiment in which we evaluated the effectiveness of several acoustic measures in predicting breathiness ratings for sustained vowels spoken by nonpathological talkers who were asked to produce nonbreathy, moderately breathy, and very breathy phonation (Hillenbrand, Cleveland, & Erickson, 1994). The purpose of the present study was to determine if our earlier results could be extended to dysphonic speakers and to connected speech. Breathiness ratings were obtained from a sustained vowel and a 12-word sentence spoken by 20 pathological and 5 nonpathological talkers. Acoustic measures were made of (a) signal periodicity, (b) first harmonic amplitude, and (c) spectral tilt. For the sustained vowels, the best predictor of breathiness ratings was CPPS, a smoothed version of a cepstrum-based measure of signal periodicity. Two other periodicity measures, CPP and RPK, also provided accurate predictions of breathiness ratings. The relative amplitude of the first harmonic amplitude (H1A) and two measures of spectral tilt (BRI and H/L) correlated moderately with breathiness ratings. For the sentences, the signal periodicity measures also provided accurate predictions of breathiness ratings, with the most accurate predictions coming again from CPPS. The two spectral tilt measures provided substantially more

TABLE 2. Interjudge reliability measured with Cronbach's coefficient alpha (/a/ = sustained vowel, RB = second sentence of the Rainbow Passage).

Subject	/a/	RB
1	0.96	0.79
2	0.94	0.89
3	0.97	0.94
4	0.95	0.87
5	0.98	0.95
6	0.92	0.95
7	0.95	0.93
8	0.97	0.96
9	0.90	0.84
10	0.97	0.96
11	0.98	0.97
12	0.96	0.96
13	0.96	0.97
14	0.95	0.93
15	0.94	0.89
16	0.98	0.90
17	0.97	0.92
18	0.93	0.96
19	0.92	0.96
20	0.88	0.90
Mean	0.95	0.92

TABLE 3. Intercorrelation matrix for breathiness ratings and acoustic measures for the sustained vowels. Correlations with an absolute value greater than 0.34 are significant at 0.05 or better.

	CPP	CPPb	CPPh	CPPS	CPPSb	CPPSh	RPK	RPKb	RPKh	P/A	P/Ab	P/Ah	BRI	H/L	H1A
BR	-0.89	-0.86	-0.83	-0.96	-0.93	-0.90	-0.84	-0.82	-0.76	-0.25	-0.45	-0.49	0.62	0.64	0.70
CPP	—	0.99	0.98	0.96	0.98	0.99	0.66	0.90	0.92	0.15	0.61	0.63	-0.63	-0.62	-0.69
CPPb	—	—	0.99	0.93	0.98	0.99	0.59	0.92	0.94	0.13	0.57	0.59	-0.62	-0.61	-0.66
CPPh	—	—	—	0.90	0.96	0.98	0.55	0.91	0.95	0.12	0.58	0.61	-0.63	-0.59	-0.65
CPPS	—	—	—	—	0.97	0.95	0.80	0.86	0.82	0.18	0.54	0.56	-0.61	-0.63	-0.68
CPPSb	—	—	—	—	—	0.99	0.69	0.92	0.90	0.17	0.53	0.53	-0.62	-0.66	-0.68
CPPSh	—	—	—	—	—	—	0.66	0.92	0.93	0.12	0.53	0.55	-0.69	-0.69	-0.65
RPK	—	—	—	—	—	—	—	0.55	0.47	0.03	0.36	0.42	-0.61	-0.57	-0.43
RPKb	—	—	—	—	—	—	—	—	0.91	-0.02	0.38	0.44	-0.61	-0.63	-0.55
RPKh	—	—	—	—	—	—	—	—	—	0.03	0.45	0.44	-0.68	-0.69	-0.54
P/A	—	—	—	—	—	—	—	—	—	—	0.25	0.16	0.15	0.12	-0.41
P/Ab	—	—	—	—	—	—	—	—	—	—	—	0.89	-0.27	-0.18	-0.40
P/Ah	—	—	—	—	—	—	—	—	—	—	—	—	-0.29	-0.16	-0.45
BRI	—	—	—	—	—	—	—	—	—	—	—	—	—	0.83	0.11
H/L	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.21

Note. BR = breathiness rating; CPP = cepstral peak prominence; CPPS = CPP with smoothing; 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

accurate predictions of breathiness ratings for the sentences than the sustained vowels.

The present results from speakers with laryngeal pathologies are surprisingly similar to our previous study using nonpathological speakers who were asked to produce non-breathy, moderately breathy, and very breathy sustained vowels. As with our previous study, we found that the periodicity measures provided rather accurate predictions of breathiness ratings in spite of the fact that pitch-tracking errors were very common for the more dysphonic voices. This finding is in contrast to "single event" methods such as

TABLE 4. Correlations between breathiness ratings and acoustic measures for (a) the sustained vowels recorded from 20 dysphonic and 5 nondysphonic speakers in the present study, (b) the Rainbow Passage (RB) signals recorded from 20 dysphonic and 5 nondysphonic speakers in the present study, and (c) the sustained vowels recorded from nondisordered speakers in our previous study (Hillenbrand et al., 1994).

Acoustic measure	Dysphonic speakers		Nondysphonic speakers
	Vowel	RB	Vowel
CPP	-0.89	-0.88	-0.92
CPPb	-0.85	-0.81	-0.90
CPPh	-0.83	-0.81	-0.89
CPPS	-0.96	-0.92	—
CPPSb	-0.93	-0.82	—
CPPSh	-0.90	-0.85	—
RPK	-0.84	-0.84	-0.54
RPKb	-0.82	-0.77	-0.91
RPKh	-0.75	-0.76	-0.89
P/A	-0.25	-0.58	-0.30
P/Ab	-0.45	-0.72	-0.54
P/Ah	-0.50	-0.63	-0.58
BRI	0.62	0.83	0.41
H/L	0.64	0.84	0.51
H1A	0.70	0.52	0.66

Note. CPP = cepstral peak prominence; CPP-S = CPP with smoothing; 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

pitch perturbation and harmonics-to-noise ratio (HNR), which can be strongly affected by errors in locating the onsets of individual pitch pulses (Hillenbrand, 1987). Pitch tracking errors are probably responsible to a large degree for the poor reliability of perturbation measures (e.g., Karnell, Scherer, & Fischer, 1991; Karnell, Hall, & Landahl, in press; Rabinov, Kreiman, Gerratt, & Bielamowicz, 1995; Gerratt & Kreiman, 1994).

Although CPP, CPPS, and RPK are based on pitch trackers, it is important to note that, unlike pitch perturbation and HNR, the F_0 measurements themselves do not enter into the periodicity calculations. The reason that CPP and CPPS continue to correspond to perceived breathiness in spite of errors in locating the fundamental period seems to

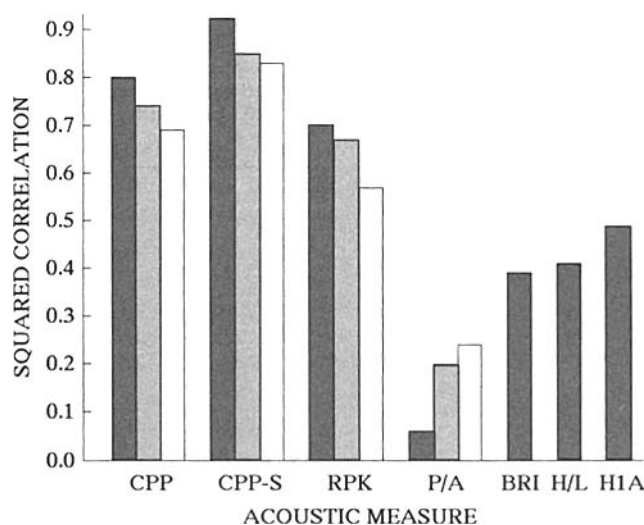
**FIGURE 5.** Squared correlation coefficients between acoustic measures and breathiness ratings for the sustained vowels (darker bars = unfiltered signals, lighter bars = bandpass filtered signals, open bars = highpass filtered signals; CPP = cepstral peak prominence; CPPS = CPP with smoothing; RPK = 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).

TABLE 5. Intercorrelation matrix for breathiness ratings and acoustic measures for the Rainbow Passage signals. Correlations with an absolute value greater than 0.34 are significant at 0.05 or better.

	CPP	CPPb	CPPh	CPPS	CPPSb	CPPSh	RPK	RPKb	RPKh	P/A	P/Ab	P/Ah	BRI	H/L	H1A
BR	-0.88	-0.81	-0.81	-0.92	-0.82	-0.85	-0.85	-0.77	-0.76	0.58	-0.73	-0.64	0.83	0.84	0.52
CPP	—	0.98	—	0.95	0.98	0.99	0.66	0.92	0.94	-0.36	0.81	0.68	-0.81	-0.76	-0.45
CPPb	—	—	0.99	0.90	0.98	0.99	0.56	0.94	0.96	-0.25	0.79	0.65	-0.76	-0.68	-0.41
CPPh	—	—	—	0.89	0.97	0.98	0.56	0.94	0.97	-0.27	0.81	0.65	-0.78	-0.70	-0.40
CPPS	—	—	—	—	0.94	0.94	0.82	0.81	0.80	-0.52	0.77	0.69	-0.82	-0.80	-0.45
CPPSb	—	—	—	—	—	0.99	0.62	0.90	0.91	-0.31	0.80	0.67	-0.78	-0.72	-0.41
CPPSh	—	—	—	—	—	—	0.62	0.91	0.93	-0.34	0.83	0.70	-0.82	-0.75	-0.43
RPK	—	—	—	—	—	—	—	0.49	0.46	-0.74	0.54	0.58	-0.66	-0.72	-0.26
RPKb	—	—	—	—	—	—	—	—	0.96	-0.19	0.63	0.47	-0.68	-0.63	-0.52
RPKh	—	—	—	—	—	—	—	—	—	-0.22	0.71	0.51	-0.74	-0.66	-0.39
P/A	—	—	—	—	—	—	—	—	—	—	-0.32	-0.41	0.58	0.68	-0.11
P/Ab	—	—	—	—	—	—	—	—	—	—	—	0.89	-0.70	-0.65	-0.26
P/Ah	—	—	—	—	—	—	—	—	—	—	—	—	-0.57	-0.55	-0.15
BRI	—	—	—	—	—	—	—	—	—	—	—	—	—	0.96	0.29
H/L	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.31

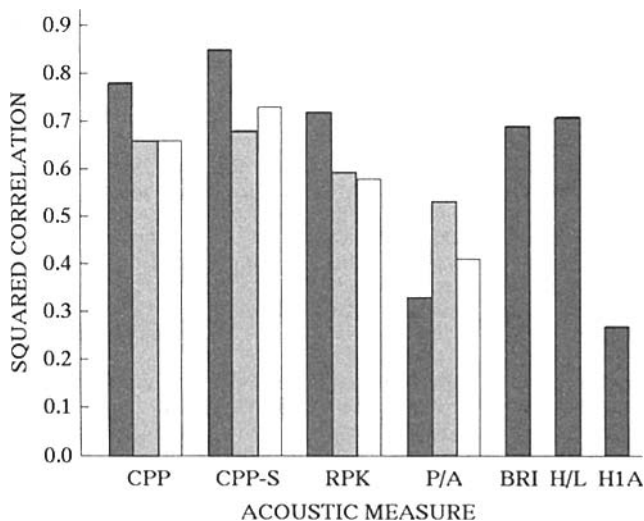
Note. BR = breathiness rating; CPP = cepstral peak prominence; CPPS = CPP with smoothing; RPK = Pearson r at autocorrelation peak; P/A = peak-to-average ratio; BRI = breathiness index, H/L ratio of high- to mid/low-frequency energy; H1A = first harmonic amplitude; b = bandpass filtered between 2.5 and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

be related to the fact that the cepstrum of a breathy voice tends to be rather flat. Consequently, the peak prominence measure, which is essentially a peak-to-average calculation, tends to be rather small for breathy voices even when the cepstral peak does not correspond to the fundamental period. Likewise, RPK measures the similarity of adjacent segments of the time waveform, and this similarity measure tends to be low for breathy signals and high for nonbreathy signals even when the adjacent waveform segments do not correspond to separate pitch pulses. In the case of pitch-halving errors, which are rather common, the measure simply reflects the overall similarity of a pair of pitch pulses with an adjacent pair, so it is not surprising that RPK

provides a reasonable estimate of periodicity in spite of the octave error.

In contrast to the cepstrum and autocorrelation periodicity measures reported here, even a single major pitch-tracking error can result in an extremely large error for such measures as jitter and HNR (e.g., Hillenbrand, 1987). The brittle nature of traditional periodicity measures was summarized nicely by Gerratt & Kreiman (1995), who described the "... inherent difficulty in measuring perturbation in signals that deviate significantly from periodicity. This presents something of a paradox: As the phenomenon of interest (departure from periodicity) increases, confidence in determining periodicity (the essence of the measure) decreases. ..." In our view, accurate cycle-by-cycle determination of the fundamental period for marginally periodic signals is quite possibly an unsolvable problem. As Gerratt and Kreiman (1995) point out, even manual measurement of the fundamental period is problematic for many dysphonic voices. The cepstrum and autocorrelation measures described here represent an attempt to measure periodicity that does not directly depend on accurate determination of the fundamental period.

It was expected that the performance of the periodicity measures (RPK, CPP, and CPPS) would improve with bandpass or highpass filtering, but these measures provided the most accurate predictions of breathiness ratings with full-band signals. Given the strong first harmonic that is often associated with breathy voice, and the common observation that aspiration noise tends to be most prominent in the mid and high frequencies, it was anticipated that filtering to remove the low frequency components would improve prediction accuracy. This was true only for RPK in our previous study, and we have no explanation for the advantage seen in the present study for full band signals. It is possible that the problem lies in the use of fixed filter settings for all signals (2.5–3.5 kHz bandpass, 2.5 kHz highpass). It may be that a method is needed that specifically isolates the third formant from each signal, as was done by Klatt and Klatt (1990). However, tracking F3 in dysphonic talkers is likely to prove

**FIGURE 6.** Squared correlation coefficients between acoustic measures and breathiness ratings for the Rainbow Passage sentences (darker bars = unfiltered signals, lighter bars = bandpass filtered signals, open bars = highpass filtered signals; CPP = cepstral peak prominence; CPPS = CPP with smoothing; 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).

problematic, which would make the technique difficult to automate.

Another unexpected finding was the substantially greater predictive power of the spectral tilt measures for the Rainbow Passage sentences as compared to the sustained vowels. BRI and H/L accounted for approximately 40% of the variance in breathiness ratings for the sustained vowels, compared to approximately 70% of the variance for the sentences. It seems unlikely that this can be attributed to measurement error. The spectral tilt measures themselves are computationally quite simple, and examination of long-term average spectra for the sustained vowels shows rather inconsistent relationships between spectral tilt and breathiness ratings. The peak-to-average measure also accounted for substantially more of the variance in breathiness ratings for the sentences than for the sustained vowels. Although we do not have an explanation for these results, the clear lesson is that findings from sustained vowels do not necessarily apply to more complex utterances.

In contrast to spectral tilt and peak-to-average, H1A was a poorer predictor of breathiness ratings for the sentences than for the sustained vowels. This is not surprising given that H1-H2 amplitude relations will be strongly affected by variations in phonetic content as well as laryngeal posture. It is also quite possible that measurement error played a role. Because a semiautomated procedure was used to measure harmonic amplitudes for the sustained vowels, it was possible to sample H1A many more times for the sustained vowels than for the sentences.

It should be kept in mind that the present results apply only to dysphonic voices that depart from normal in the direction of breathiness. It is not known whether any of these measures will predict subjective ratings for voices that vary in dimensions such as roughness or hoarseness, or whether the measures will predict overall dysphonia ratings from a set of voices showing a broad range of dysphonic vocal qualities. Based on our current state of knowledge, this means that clinical use of these measures to track progress or to evaluate treatment outcomes would require that the clinician first make a judgment that the dysphonia is characterized primarily by breathiness. Our results provide some preliminary indication that measures such as CPPS, CPP, and RPK correlate well with perceived breathiness for voices that are selected in this way. However, the results are based on voices selected by a single listener. It would be useful to demonstrate the generality of these findings for other comparably trained listeners.

It should also be noted that the voices that were used both in the present experiment and in our previous study showed a rather wide range of breathiness percepts. It is not known whether these measures are sufficiently sensitive to predict subtle variations in breathiness among nondysphonic talkers.

Acknowledgments

We are very grateful to Jessica Schneider, who conducted the pilot work for this study, and to Jody Kreiman, Bruce Gerratt, and Michael Clark for helpful comments on an earlier draft. 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* (pp. 71-82). Cambridge, MA: Research Laboratory of Electronics, MIT.
- 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. *Psychometrika*, 16, 297-334.
- Davis, S. B. (1981). Acoustic characteristics of normal and pathological voices (ASHA Reports 11, pp. 97-115). Rockville, MD: American Speech-Language-Hearing Association.
- 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.
- 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.
- Gerratt, B., & Kreiman, J. (in press). The utility of acoustic measures of voice quality. In D. Wong (Ed.), *Proceedings of the Workshop on Acoustic Voice Analysis*.
- 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., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37, 769-778.
- 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.
- Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, 81, 495-504.
- Karnell, M. P., Hall, K. D., Landahl, K. L. (in press). Comparison of fundamental frequency and perturbation measurements among three analysis systems. *Journal of Voice*.
- Karnell, M. P., Scherer, R., & Fischer, L. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. *Journal of Speech and Hearing Research*, 34, 781-790.
- 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, CA: College Hill.
- Ladefoged, P., & Antonanzas-Barroso, N.** (1985). Computer measures of breathy phonation. *Working Papers in Phonetics*, 61, 79-86. University of California at Los Angeles.
- McKay, I. R. A.** (1987). *Phonetics: The science of speech production* (2nd ed). Boston: College Hill.
- Parsons, T. W.** (1986). *Voice and speech processing*. New York: McGraw-Hill.
- Rabinov, C. R., Kreiman, J., Gerratt, B., & Bielamowicz, S.** (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research*, 38, 26-32.
- Rothenberg, M.** (1973). A new inverse filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America*, 53, 1632-1645.
- 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.
- Witten, I. H.** (1982). *Principles of computer speech*. London: Academic Press.

Received May 24, 1995

Accepted September 6, 1995

Contact author: James Hillenbrand, PhD, Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI 49008.
E-mail: james.hillenbrand@wmich.edu