

Acoustic Prediction of Voice Type in Women with Functional Dysphonia

*Shaheen N. Awan and †Nelson Roy

*Bloomsburg, Pennsylvania, and †Salt Lake City, Utah

Summary: The categorization of voice into quality type (ie, normal, breathy, hoarse, rough) is often a traditional part of the voice diagnostic. The goal of this study was to assess the contributions of various time and spectral-based acoustic measures to the categorization of voice type for a diverse sample of voices collected from both functionally dysphonic (breathy, hoarse, and rough) ($n = 83$) and normal women ($n = 51$). Before acoustic analyses, 12 judges rated all voice samples for voice quality type. Discriminant analysis, using the modal rating of voice type as the dependent variable, produced a 5-variable model (comprising time and spectral-based measures) that correctly classified voice type with 79.9% accuracy (74.6% classification accuracy on cross-validation). Voice type classification was achieved based on two significant discriminant functions, interpreted as reflecting measures related to “Phonatory Instability” and “ F_0 Characteristics.” A cepstrum-based measure (CPP/EXP ratio) consistently emerged as a significant factor in predicting voice type; however, variables such as shimmer (RMS dB) and a measure of low- vs. high-frequency spectral energy (the Discrete Fourier Transformation ratio) also added substantially to the accurate profiling and prediction of voice type. The results are interpreted and discussed with respect to the key acoustic characteristics that contributed to the identification of specific voice types, and the value of identifying a subset of time and spectral-based acoustic measures that appear sensitive to a perceptually diverse set of dysphonic voices.

Key Words: Voice—Dysphonia—Cepstral analysis—Spectral analysis—Shimmer.

Accepted for publication March 22, 2004.

From *Bloomsburg University, Bloomsburg, Pennsylvania; and †The University of Utah, Salt Lake City, Utah.

Portions of this paper were presented at the Voice Foundation’s 32nd Annual Symposium: Care of the Professional Voice, Philadelphia, PA, June 2003.

Address correspondence and reprint requests to Shaheen N. Awan, Department of Audiology and Speech Pathology, Centennial Hall, 400 East Second St., Bloomsburg, PA 17815-1301.

E-mail: sawan@bloomu.edu

Journal of Voice, Vol. 19, No. 2, pp. 268–282

0892-1997/\$30.00

© 2005 The Voice Foundation

doi:10.1016/j.jvoice.2004.03.005

INTRODUCTION

The categorization of disordered voice into type (ie, breathy, hoarse, rough) is an essential part of the conventional voice diagnostic. The accurate categorization of voice quality can provide key insight regarding the underlying pathophysiology of the individual patient and, thus, is an important guide to the direction of treatment. In addition, changes in the categorization of voice type (particularly from dysphonic toward normal) can be an effective means of tracking changes in the voice after treatment (behavioral and/or medical-surgical). The

categorization of voice type has been traditionally accomplished via perceptual evaluation alone, and to date, many still consider perceptual assessment of the voice the key method by which dysphonias are identified and progress in therapy is tracked. Although perceptual categorization of voice quality type may seem obvious in certain cases, auditory-perceptual categorization can be difficult in several situations:

When the patient has a relatively mild dysphonia
When dysphonic type may be mixed or inconsistent

When the examiner has limited experience in categorizing voice quality type

When attempting to objectively track relatively subtle changes in voice quality type over time

To aid in the discrimination of commonly observed voice types and to gain further insight into their characteristics, voice clinicians and researchers have tried to augment their perceptual assessment of voice quality with more “objective” and quantitative methods of voice analysis. In particular, acoustic methods of voice evaluation have received attention, as they are noninvasive; readily available at relatively low cost compared with other methods of voice analysis; and relatively easy to perform.¹ In addition, because the acoustic signal is determined, in part, by movements of the vocal folds, it can be argued that, “there is a great deal of correspondence between the physiology and acoustics, and much can be inferred about the physiology based on acoustic analysis” (p. 21).² In general, acoustic methods used to categorize type of dysphonia have often focused on time-based measures. These measures have included vocal frequency (F_0) and F_0 variability, as well as methods used to quantify voice signal perturbations such as jitter, shimmer, and harmonic-to-noise ratio (HNR). Although several investigations have revealed reasonable associations between acoustic perturbation measures and voice quality categories,^{3–7} some researchers have questioned the appropriateness, validity, and clinical usefulness of specific perturbation measures, especially when applied to moderate or severely disordered voices. Cycle-to-cycle perturbation measures depend

on accurate identification of cycle boundaries (ie, where a cycle of vibration begins/ends); however, it has become increasingly evident that the presence of significant noise in the voice signal makes it more difficult to accurately locate these cycle onsets/offsets.^{8,9}

The controversy surrounding the validity of traditional methods of perturbation analysis has prompted researchers to consider other methods of quantifying the noise components in the voice signal that may be associated with particular voice types. Specifically, several investigators have reported that measures derived from spectral analysis of the voice signal may be strong predictors of factors such as presence of additive noise in the voice signal, perceived severity of dysphonia, and type of voice disorder.^{10–14} In particular, measures of spectral tilt,^{10,15} amplitude of the first spectral harmonic,^{10,16} and reductions in spectral harmonic-to-noise ratios^{11,17,18} have been reported as effective indices of dysphonic type and severity. In addition to the measures of the spectrum, derivation of the cepstrum has also been investigated as a useful method for describing the dysphonic voice. As originally described by Noll,¹⁹ the cepstrum is derived via a Fourier transform of the power spectrum of the voice signal, and it graphically displays the extent to which the spectral harmonics and, in particular, the vocal fundamental frequency, are individualized and emerge out of the background noise level. It is the degree to which the cepstral peak relates to extraneous vocal frequencies that theoretically provides an effective method of quantification for the disordered voice.¹⁰

Several investigators have demonstrated the effectiveness of measures derived from cepstral analysis to quantify dysphonic voice characteristics such as voice type. For instance, in studies dealing with the acoustic correlates of breathy voice, Hillenbrand et al¹⁰ and Hillenbrand and Houde¹² observed that measures of signal periodicity derived from cepstral analysis were among the measures most strongly correlated with ratings of breathiness from sustained vowels. Research by Dejonckere and Wieneke¹¹ corroborated the work of Hillenbrand and colleagues, and they observed that the magnitude of the dominant cepstral peak was significantly larger in normal voice samples than those peaks measured from pathological voices, such as breathy or rough voices.

Wolfe and Martin⁷ also explored the ability of various acoustic measures to classify dysphonic patients. Using discriminant function analysis, 45 dysphonic subjects were classified with 92% accuracy into breathy, hoarse, and strained voice types using a four-parameter model consisting of jitter standard deviation, fundamental frequency, SNR standard deviation, and cepstral peak prominence (CPP). In a similar finding to Dejonckere and Wieneke,¹¹ CPP was observed to be lower in both breathy and hoarse voices, with no significant difference observed between the groups on this parameter. Finally, Heman-Ackah et al²⁰ reported that measures derived from the cepstral peak (both in continuous speech and in sustained vowel samples) were the strongest individual correlates of overall dysphonia and ratings of breathiness. Cepstral measures were also significantly related to ratings of roughness, although the authors felt that too little variance was accounted for in the prediction of ratings of roughness to make them clinically applicable.

The aforementioned investigations have demonstrated that acoustic measures derived from time-based and spectral/cepstral analyses methods can be used to characterize voice type. However, several limitations with the previous studies warrant further research in the area of acoustic correlates of dysphonia. First, several of the studies that used spectral/cepstral methods to describe dysphonia have focused only on single-quality dimensions such as breathiness^{10,12} or hoarseness,¹¹ and they have ignored other possible voice types. Second, a number of studies^{7,20} did not include normal voice samples among the voices to be classified. The inclusion of normal samples is important because (1) it has been observed that certain voice types such as breathiness may have many similarities to normal voices,²¹ thus limiting the possible effectiveness of acoustic categorization in some cases; and (2) if acoustic methods are to be used to track change in voice characteristics over time, it would be useful to have "normal" classification as one of the diagnostic categories. Thus, the goal of this study was to identify a subset of acoustic measures (both time and spectral/cepstral based) that would aid in classification of voice type for a wide range of normal and dysphonic voice samples. It was intended that the results of this study could serve a verification function for previous studies but

also extend those findings to a group of heterogeneous voice types likely to be encountered clinically.

METHODS

Participants

Voice samples from a variety of vocally normal and disordered adult female subjects were selected for inclusion in this study. Female voices were specifically selected because, like many multidisciplinary voice clinics, the majority of our patients seeking help for voice difficulties are women. Coyle et al²² have confirmed the higher prevalence of voice disorders among women. All subjects (Total N = 134) were native speakers of English and were selected from a diverse patient group consisting of both non-voice-disordered otolaryngology patients who attended a university-based otolaryngology clinic for physical complaints unrelated to voice production, as well as otolaryngology patients who were evaluated for specific voice-related complaints. A perceptually diverse set of voice quality types (breathy; hoarse; rough) and severity were prime considerations in the sample construction. All voice samples for the disordered group were acquired from patients who received the diagnosis of functional dysphonia. The diagnosis of functional dysphonia was determined after comprehensive laryngeal examination and medical investigation by both a laryngologist and a speech language pathologist specializing in voice disorders.

Voice Samples

As part of a standard clinical test battery, the subjects were asked to produce a sustained vowel /a/ at a comfortable pitch and loudness for at least 5 seconds in duration. Voice samples were recorded using a research quality microphone and digitized at 25 kHz and 16 bits of resolution using the *Computerized Speech Lab (CSL) Model 4300* (Kay Elemetrics Corporation, Lincoln Park, New Jersey).²³ All recordings were peaked at -6 dB from overload as determined via the LED indicators on the *CSL* external module. After digitization, vowel onsets and offsets were edited to leave the central 1 second of the phonation for further analysis. The 1-second vowel samples were then saved in .wav format for

later analysis using voice analysis software developed by the first author. The first author's own software was used in this study primarily because it provided a single software solution to the various time and spectral-based analysis methods that were to be employed. In addition, the software provided for automatic cepstral computation and extraction of the cepstral peak prominence and associated normalization computations. Validity data for the algorithms employed is provided in Awan and Frenkel²⁴ and Awan.²⁵

Computer Analysis of Voice Samples:

Time-Based Methods

Although issues with the validity of time-based acoustic measures have been previously alluded to in this study, we included time-based measures in our analysis in an effort to determine the extent of their contribution to the categorization of dysphonia type, especially in light of the inclusion of spectral-based measures. These traditional time-based measures are easily communicated to patients and clinicians alike, and they continue to be supported by a vast literature base. We, therefore, felt it appropriate to include them in our battery of acoustic analysis methods.

The F_0 extraction algorithm used was a peak-picking event detector based on the Gold-Rabiner pitch tracker.^{26,27} In the F_0 extraction algorithm, the signal is moving average filtered to remove higher frequency vocal tract information, windowed (13.33 ms window length) and center-clipped to minimize formant information and retain only information related to periodicity.²⁷ The clipping procedure results in a series of pulses that contain the peak amplitude of the cycle as well as all other amplitudes greater than a predetermined clip level (all amplitudes $> 0.70 \times$ the peak amplitude of the cycle). The peak amplitude and corresponding sample number (ie, time index) provide initial cycle markers that are then applied to the original unfiltered signal to identify the "true" peaks within each cycle. This method of using "rough" estimates of cycle boundaries from a filtered speech signal to guide accurate peak extraction in the original unfiltered speech signal has been previously discussed by Titze et al.²⁸ Analysis of the unfiltered speech signal results in period and frequency estimates for each identified

cycle. The F_0 estimates are then submitted to a series of error correction and smoothing routines (removal of F_0 estimates < 75 Hz or > 1000 Hz; median smoothing) that account for possible gross errors in F_0 estimation prior to providing a graphical F_0 contour and statistical results.

From the cycle-boundary markers and frequency estimates, measures of mean fundamental frequency (F_0 in Hz) and F_0 standard deviation were computed. In addition, once cycle boundaries were identified, perturbation methods such as jitter (%), shimmer (RMS dB), and HNR (dB) could be computed. The following time-based acoustic measures were computed for each vowel sample: F_0 (mean F_0); F_0 SD (F_0 standard deviation); SIG (Pitch sigma—the F_0 standard deviation converted to semitones); RANGEHZ (F_0 range in Hertz); RANGEST (F_0 range in semitones); JIT (Jitter %); HNR (Harmonics-to-noise ratio in decibels); and SHIM (Shimmer in decibels). All Hertz to semitone conversions were computed using formulas presented in Baken.²⁹

Computer Analysis of Voice Samples:

Spectral-Based Methods

These methods were derived from the spectrum of the digitized signal as computed via the discrete Fourier transformation (DFT). Spectral and subsequent cepstral analyses were conducted on full-band signals.¹⁰ Spectral analysis incorporated a series of nonoverlapping 1024-pt. DFTs (41 ms window) that were computed and averaged across the entire 1-second sample. Prior to DFT computation, Hamming windows were applied to eliminate abrupt onsets and offsets for each window. From the averaged DFT, a ratio of low- versus high-frequency energy was calculated. For the purpose of this study, energy < 4000 Hz was compared with energy > 4000 Hz¹⁰ and was referred to as the Discrete Fourier Transform Ratio (DFTR). Variants of this type of ratio have been observed to correlate well with severity ratings of breathiness.^{11,30,31}

Following the DFT, the cepstrum of the voice sample was derived by (1) computing the average log power spectrum from the average DFT spectrum, and (2) computation of the DFT of the average log power spectrum. After computation of the cepstrum, the cepstral peak prominence (CPP) was identified

below 500 Hz (ie, at frequencies >2 ms). Although peak-picking the cepstrum at the frequency associated with the fundamental frequency of the voice may identify the CPP in most normal voice signals, the CPP often does not correspond with the F_0 in signals that have been severely perturbed.²⁵ In these cases, the identification of a CPP that does not correspond to the true fundamental frequency of the voice may result in erroneous estimates of noise in the voice signal. With this in mind, the accuracy of the cepstral peak-picking procedure was guided in reference to identification of (1) the first significant amplitude harmonic and (2) harmonic spacing in the original DFT.

The method of quantifying the relative height of the cepstral peak used in this study was a ratio of the amplitude of the cepstral peak prominence (CPP) to the expected (EXP) amplitude of the cepstral peak (CPP/EXP) as derived via linear regression. The CPP/EXP method is similar to that described by Hillenbrand et al¹⁰ and Hillenbrand and Houde,¹² with the exception that these authors describe the *difference* between the cepstral peak and the expected value via linear regression, and the current study uses the ratio between the aforementioned values converted to decibels. The aforementioned ratio only uses cepstral values greater than 2 ms frequencies because frequencies below 2 ms (ie, higher frequencies) are often attributed to vocal tract resonances. The following spectral/cepstral-based measures were computed for each vowel sample: DFTR (in decibels); and CPP/EXP (in decibels). Figure 1 provides an example of spectral and cepstral analysis computed for a normal voice sample.

Description of the Listener Rating Task

Twelve female judges, ages 23 to 50 years, were asked to rate each of the 134 sustained vowel samples. All judges successfully passed a hearing screening of 20 dB at 0.5, 1, 2, and 4 kHz, and they were recent Master's degree graduates of the Department of Audiology & Speech Pathology, Bloomsburg University. All of the judges had (1) completed a graduate course in voice disorders, (2) been exposed to the terminology to be used in the rating task, (3) participated in classroom exercises in the perceptual evaluation of voice, and (4) had clinical experience with voice-disordered patients. Many judges were

used to reduce the potential for interjudge differences to create spurious experimental conclusions.³²

The digitized sustained vowel samples from the 134 subjects were labeled consecutively (1 to 134) and transferred to CD-R using an LG-CED8080B CD-R/RW recorder (LG Electronics USA Inc, Rosemont, Illinois). Software was developed that would allow the user to randomly select samples for playback from the CD-ROM of a Gateway 2000 (Model E-3000, Gateway, Poway, California) Pentium MMX computer. The software allowed for random playback without the necessity of multiple randomized tape construction or possibility of signal degradation in transferring samples from digital-to-analog tape domains. Judges listened to each sample using high-quality headphones (Technics RP-HTZZ stereo headphones, Matsushita Electronics Corp. of America, Secaucus, New Jersey) connected directly to a SoundBlaster Awe 16 soundboard (Creative Labs Inc., Milpitas, California).

The 12 judges were asked to make judgments regarding type of voice quality (normal, breathy, hoarse, rough)¹ as well as severity for each of the 134 sustained vowel samples. Although prediction of severity was not a focus of this study, severity ratings were evaluated to ensure that the voice sample corpus reflected similar overall degrees of severity within each dysphonic group.² Before the judgment task, a 20-minute training period was provided whereby instructions were provided regarding the randomization of the voice samples, the use of the response form, and review of definitions for voice quality types and severity. In addition, each judge listened to representative samples (preselected by the first author) that illustrated the range of voice types and severities included within the 134 voice samples to be judged.

¹ The following definitions for voice quality type were provided for the 12 judges before the voice sample rating task: *breathy* ("breathiness" is associated with hypoadduction of the vocal folds and refers to the audible detection of airflow through the glottis-the breathy voice is often perceived as a whispery or airy voice); *rough* (rough voice is associated with hyperadduction of the vocal folds and refers to the noise produced as a result of irregular vocal fold vibration-rough voice is often perceived as a coarse, low-pitched noise); *hoarse* (the hoarse voice has both breathy and rough qualities simultaneously).

² The following summary statistics are provided for severity ratings within each of the four groups to be discriminated: normal (mean = 0.31; SD = 0.33); breathy (mean = 2.09; SD = 1.01); hoarse (mean = 3.36; SD = 1.11); rough (mean = 2.38; SD = 1.33).

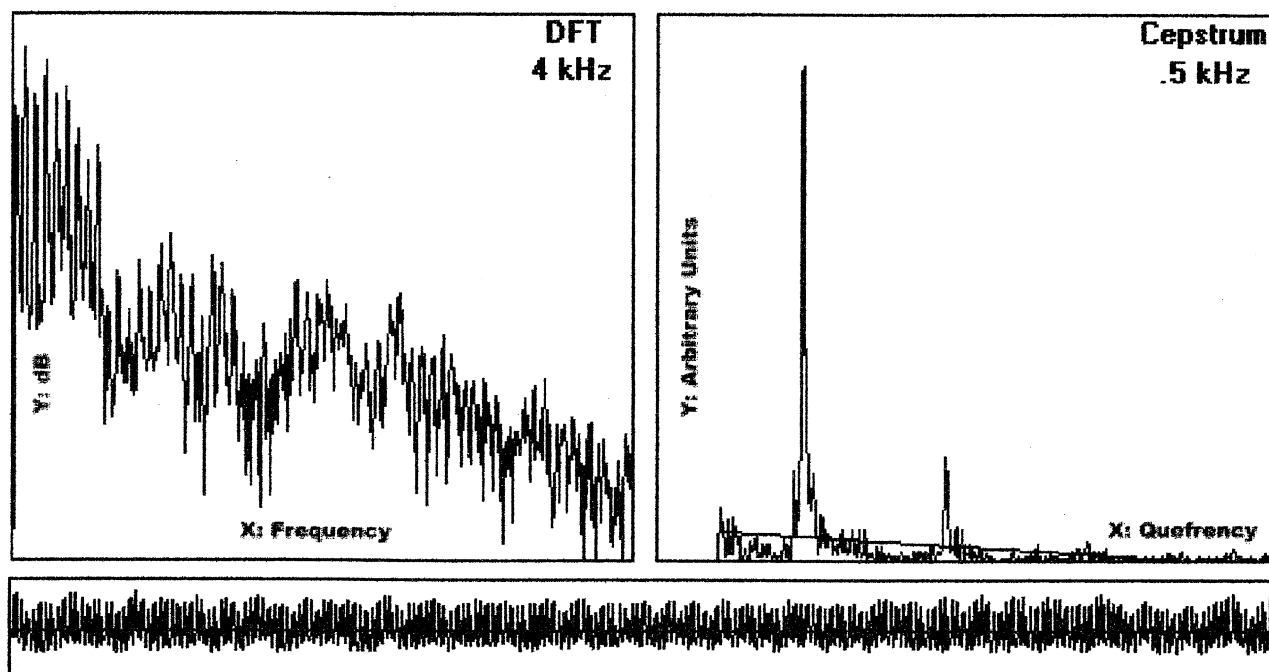


FIGURE 1. Discrete Fourier transformation (DFT) and cepstral analysis for a normal female voice sample. The cepstral peak prominence (CPP) in this sample corresponds to the fundamental period and is substantially greater than the average cepstral amplitude. A regression line used to quantify the relative height of the cepstral peak is shown overlaid on the cepstrum.

Judges were asked to rate all sustained vowel samples within a 2-hour time period (a 15-minute break followed the first 45 minutes of the task). For each voice sample, judges were allowed to replay the sample as many times as necessary during the rating task. Judges also were allowed to compare each voice sample with a preselected external “standard” during each rating. The external “standard” was a voice sample judged to represent “normal” in terms of voice quality and pitch/loudness by an expert listener. The same external “standard” was used for all 134 judgments. The use of referent voice recordings as “anchors” has been discussed as a possible method by which the reliability and validity of rating scales for voice assessment may be improved.³³ By giving all judges a fixed perceptual referent, it was expected that listener-related variability in ratings would be reduced.³⁴ Therefore, all judgments were made in relation to (1) the internal standards of each listener, (2) the verbal definitions provided by the examiner, and (3) the voice characteristics of the external “standard.”

Interjudge and Intrajudge Reliability

The interjudge reliability for the ratings of voice type was assessed by using the proportional reduction in loss (PRL) reliability measure.³⁵ The PRL statistic is analogous to Cronbach’s Coefficient Alpha, but it is applicable to nominal data. The PRL statistic is inversely proportional to the amount of “loss” (ie, error) the researcher would expect from using a measure representative of the consensus of a series of judges. For the current study, a PRL level of 0.99 was achieved, indicating strong interjudge reliability and a low level of expected error when using a consensus measure of the 12 judges. Consensus among the judges for this study was determined via the determination of the modal value (ie, most frequently occurring rating) among the 12 judges for each voice sample. This modal value was then used as the voice quality classification for each particular sample. Using this method, the 134 voice samples were divided into the following classifications: normal ($n = 51$), breathy ($n = 31$), hoarse ($n = 27$), and rough ($n = 25$).

For assessing intrajudge reliability, each judge was asked to rate 40 voices selected at random from the original 134 voice sample corpus within 2 weeks of the original rating. Intrajudge reliability was assessed by computing the percent exact agreement between voice type ratings of the same voice samples from the first vs. second rating sessions. The mean percent exact agreement was 73.5% (range: 62% to 85%). Review of the test-retest data indicated that most of the variability in voice type rating occurred between overlapping categories (ie, breathy vs. hoarse; rough vs. hoarse).

RESULTS

All statistical analyses were conducted with *SPSS 10.0* (SPSS Corporation, Chicago, Illinois).³⁶ A review of results from Kolmogorov–Smirnov tests of normality indicated that data for several acoustic variables were not normally distributed. Log transformations (for measures of jitter, shimmer, and F_0 range) and inverse square root transformations (for measures of mean F_0 and F_0 standard deviation/pitch sigma) produced the best approximations of normality and reduction in outliers for the non-normal variables. These transformations were applied before any parametric statistics. The following acronyms are used to indicate the various transformed-dependent variables: LOGJIT (the logarithm of jitter); LOGSHIM (the logarithm of shimmer); LOGRANGEHZ (the logarithm of F_0 range in Hertz); LOGRANGEST (the logarithm of F_0 range in semitones); INVSQRTF₀ (the inverse square root of the mean F_0); INVSQRTF₀SD (the inverse square root of the F_0 standard deviation); and INVSQRTSIG (the inverse square root of the pitch sigma).

Discrimination of Voice Type

The ability of acoustic variables to accurately discriminate between primary voice types (normal, breathy, hoarse, rough) was evaluated using stepwise discriminant analysis. To control for unnecessary redundancy among variables and to minimize multicollinearity, certain variables were removed before the discriminant analysis if they had particularly high correlations ($r > 0.90$) with other variables. Review of correlation coefficients among all acoustic variables resulted in the removal of LO-

GRANGE (both in Hertz and semitones) and INVSQRTF₀SD from the subsequent discriminant analysis procedure. LOGRANGE (both in Hertz and semitones) and INVSQRTF₀SD were observed to strongly correlate with INVSQRTSIG ($r > 0.93$) and were removed in favor of INVSQRTSIG because (1) measures of range may be particularly affected by gross F_0 extraction errors, and (2) measures of F_0 variability converted to semitones (as in pitch sigma) are scaled in relation to the mean F_0 of the subject.

The remaining acoustic variables were entered into the stepwise discriminant analysis, resulting in a five-variable model consisting of LOGSHIM, CPP/EXP, DFTR, INVSQRTF₀, and INVSQRTSIG. This five-variable model resulted in three significant canonical discriminant functions (all significant at $p < 0.0001$), the first two of which accounted for 93.5% of the total dispersion among the four voice types. The first canonical discriminant function accounted for the greatest degree of spread between group means (79.3%)—a review of the standardized discriminant function coefficients indicated that LOGSHIM, CPP/EXP, and DFTR were all of similar absolute magnitude and were the most important discriminators within the first canonical discriminant function (see Table 1). The second canonical discriminant function accounted for 14.2% of the total dispersion between voice types. The most important discriminators within the second canonical discriminant function were INVSQRTF₀ and CPP/EXP. Group means and standard deviations for each of the five acoustic variables included in the final discriminant analysis model are provided in Table 2.

Based on the five-variable model, classifications were made accurately for 79.9% of the voice samples. Table 3 provides the number of correct and incorrect classifications for each voice type. Because discriminant analysis procedures may provide overly optimistic estimates of the success of the classification, a “leave-one-out” cross-validation procedure was also computed for the 134-sample corpus used in this study. In this leave-one-out procedure (also known as a “jackknife” procedure), each case is reclassified based on the classification functions computed from all of the data *except* for the case being classified.³⁶ This procedure helps to reduce any bias included in the original analysis. For our data,

TABLE 1. *Standardized Discriminant Function Coefficients for the Acoustic Variables Included in the Five-Variable Model Used to Classify Voice Type*

Acoustic Variable	Function 1:	Function 2: F ₀ Characteristics
	Phonatory Instability	
LOGSHIM	−0.473	0.554
INVSQRTF ₀	−0.200	0.683
INVSQRTSIG	0.378	−0.178
DFTR	0.407	0.554
CPP/EXP	0.433	0.612

LOGSHIM, the logarithm of shimmer; INVSQRTF₀, the inverse square root of F₀; INVSQRTSIG, the inverse square root of pitch sigma; DFTR, the Discrete Fourier Transform ratio; CPP/EXP, the ratio of the amplitude of the cepstral peak prominence to the expected amplitude of the cepstrum as determined via linear regression.

the cross-validation procedure resulted in a change of 5.3% accuracy in classification (79.9% original grouped cases correctly classified vs. 74.6% cross-validated grouped cases correctly classified). It is our view that this represents a relatively minor reduction in classification accuracy and, therefore, supports the application of the original five-factor model.

Figure 2 provides a territorial map indicating the boundaries defined for each of the four voice quality types based on the first two canonical discriminant functions. In this map, the first canonical discriminant function has been interpreted and colabeled “Phonatory Instability”; the second canonical discriminant function is interpreted and colabeled “F₀ Characteristics.” The first canonical discriminant function includes a measure of short-term amplitude variability (shimmer), spectral tilt (DFTR), and a global measure that may be affected by high- or low-frequency noise components, vocal fold irregularity, or some combination of these factors (CPP/EXP). The second function (F₀ Characteristics) is affected mostly by the mean F₀ as well as by the amplitude of the F₀ in comparison with surrounding frequencies in the voice spectrum (CPP/EXP). In addition to group boundaries, group centroids (ie, canonical variable means) are also indicated. Pairwise group comparisons indicated that all group centroids were significantly different ($p < 0.0001$).

Hoarse vs. rough voice types showed the most similarity ($F = 9.11$, $p < .0001$), whereas normal vs. hoarse voice types were observed to differ the most ($F = 47.74$, $p < 0.0001$). It is clear that the four voice type categories used in this study were not completely orthogonal. There is obvious overlap between these categories, with normal voice located central on a continuum from breathy to rough voice, and hoarseness also located central to the breathy and rough voice types.

Stepwise discriminant analyses using only the five acoustic variables (LOGSHIM, CPP/EXP, DFTR, INVSQRTF₀, and INVSQRTSIG) that had entered into the initial five-variable model were also computed for all possible normal vs. disordered voice type pairwise comparisons. Inspection of the results revealed the following:

Normal vs. Breathy. A two-variable model consisting of CPP/EXP and DFTR (significant at $p < 0.0001$) was able to correctly classify 87.8% of the original grouped cases (92.2% (47/51) of the normal subjects vs. 80.6% (25/31) of the breathy subjects). Cross-validation resulted in a minor reduction in accuracy to 86.6% correct classification. A review of the standardized canonical discriminant function coefficients indicated that CPP/EXP was the strongest contributor to the two-variable model.

Normal vs. Hoarse. A four-variable model consisting of LOGSHIM, CPP/EXP, DFTR, and INVSQRTSIG (significant at $p < 0.0001$) was able to correctly classify 97.4% of the original grouped cases (100% (51/51) of the normal subjects vs. 92.6% (25/27) of the hoarse subjects). Cross-validation resulted in no change to the classification accuracy. LOGSHIM was observed to be the strongest contributor to the four-variable model.

Normal vs. Rough. A four-variable model consisting of LOGSHIM, INVSQRTF₀, CPP/EXP, and INVSQRTSIG (significant at $p < 0.0001$) was able to correctly classify 93.4% of the original grouped cases (98.0% (50/51) of the normal subjects vs. 84.0% (21/25) of the rough subjects). Cross-validation resulted in a minor reduction in accuracy to 92.1% correct classification. LOGSHIM was again

TABLE 2. Group Means and Standard Deviations for Each of the Five Acoustic Variables Included in the Final Discriminant Analysis Model

Acoustic Variable	Normal	Breathy	Hoarse	Rough
LOGSHIM	-0.549 (0.210)	-0.396 (0.256)	0.138 (0.243)	-0.102 (0.297)
INVSQRTF ₀	0.0697 (0.007)	0.0687 (0.009)	0.0697 (0.011)	0.0795 (0.011)
INVSQRTSIG	1.565 (0.462)	1.113 (0.382)	0.619 (0.278)	0.843 (0.433)
DFTR	70.988 (11.864)	58.89 (13.095)	46.87 (14.983)	64.646 (15.120)
CPP/EXP	25.117 (3.834)	17.15 (4.130)	16.143 (4.486)	18.828 (7.390)

LOGSHIM, the logarithm of shimmer; INVSQRTF₀, the inverse square root of F₀; INVSQRTSIG, the inverse square root of pitch sigma; DFTR, the Discrete Fourier Transform ratio; CPP/EXP, the ratio of the amplitude of the cepstral peak prominence to the expected amplitude of the cepstrum as determined via linear regression.

observed to be the strongest contributor to the four-variable model.

In addition to the normal vs. disordered type comparisons, stepwise discriminant functions were also computed to evaluate the degree of success in classifying one dysphonia type versus another. Three separate discriminant function analyses were computed:

Breathy vs. Hoarse. A one-variable model consisting solely of LOGSHIM (significant at $p < 0.0001$) was able to correctly classify 84.5% of the original grouped cases (83.9% (26/31) of the breathy subjects vs. 85.2% (23/27) of the hoarse subjects). Cross-validation resulted in no change to the classification accuracy.

Breathy vs. Rough. A two-variable model consisting of LOGSHIM and INVSQRTF₀ (significant at $p < 0.0001$) was able to correctly classify 78.6% of the original grouped cases (83.9% (26/31) of the breathy subjects vs. 72.0% (18/25) of the rough subjects). Cross-validation resulted in no change to the classification accuracy. A review of the standardized canonical discriminant function coefficients indicated that the LOGSHIM and INVSQRTF₀ appeared to contribute relatively equally to the two-variable model.

Hoarse vs. Rough. A three-variable model consisting of DFTR, INVSQRTF₀, and LOGSHIM (significant at $p < 0.0001$) was able to correctly classify 80.8% of the original grouped cases (77.8% (21/27) of the hoarse subjects vs. 84.0% (21/25) of the rough subjects). Cross-validation resulted in a minor reduc-

tion in accuracy to 78.8% correct classification. Standardized canonical discriminant function coefficients indicated that the DFTR was the strongest contributor to the three-variable model.

DISCUSSION

Discrimination of Voice Type

A combination of five distinct acoustic measures was observed to successfully classify a wide variety of voice samples into four primary voice types. The five variables included time-based measures derived from fundamental frequency (mean F₀), short-term signal perturbation (shimmer), and long-term signal variability (pitch sigma). In addition, the model incorporated two spectral-based measures—a relative measure of low- vs. high-frequency energy concentration in the spectrum (DFTR) and a measure of the relative strength of the fundamental frequency to strength of the background spectral noise (CPP/EXP). The results of this study indicate that meaningful acoustic models applicable to the description of dysphonic voice may be determined for a diverse set of voice samples encompassing a wide range of types and severities.

The CPP appears to be a general discriminator of dysphonia, most effective in discriminating between normal and various dysphonic types. However, it appears that the relative prominence of harmonic vs. noise components throughout the spectrum may not be a sufficient discriminator of pathological voice type by itself,⁷ and it may not be an effective discriminator between dysphonic voice types. This conclusion is supported by the observation that the

TABLE 3. *Number of Correct and Incorrect Voice Type Classifications Based on the 5-Variable Model*

Type	Percent Correct	Normal	Breathy	Hoarse	Rough
Normal	92.3	47	3	0	1
Breathy	67.7	6	21	2	2
Hoarse	81.5	0	2	22	3
Rough	68.0	3	2	3	17

CPP/EXP ratio was not a significant contributor to any of the discriminant functions separating the dysphonic groups (ie, dysphonic voice types). In contrast, measures derived from shimmer appear to be useful in specifying type of dysphonia in those voices in which irregularity or instability of phonation over time is a key characteristic. In particular, shimmer appears to be related to the aperiodicity of vocal fold vibration associated with rough and hoarse (the rough component) dysphonic types^{7,37} rather than the unmodulated airflow accompanying phonation in the breathy voice type. In addition, shimmer appeared to represent some component of the acoustic signal independent of other time-based measures of perturbation such as jitter and HNR. It may be that the addition of spectral/cepstral methods rendered HNR measures (originally conceived as a measure of spectral noise) redundant. The results of this study suggest that shimmer is, perhaps, the most important of the time-based indices of short-term signal variability. Future studies that attempt to assess the relative strengths of these various measures and their possible associations with underlying vocal physiology will be particularly useful.

Normal vs. Breathy Voice

The accuracy of classification of the normal vs. breathy voice types in isolation was good (87.8% predictive accuracy), with the two groups differing primarily on measures of spectral characteristics. Breathy voice has been said to correspond to turbulent noise originating from the glottis.³⁸ In the current study, the breathy distinction appeared to be made based on two key characteristics. First, it appears that in many breathy voices, there is a significant increase in the upper frequency content of the voice signal resulting in “spectral tilt” (ie, the relative spectral slope dependent on the degree of

energy concentrated in the low- vs. high-frequency areas of the spectrum)^{11,12,15,16} and reduced DFTR. Second, this spectral tilt is reflected in the subsequent cepstral analysis—the increase in high-frequency noise may result in a reduced ratio between the cepstral peak prominence and the expected cepstral amplitude as determined via linear regression.^{10,12}

It is interesting to note that none of the time-based measures were significantly weighted in the discriminant function separating normal vs. breathy groups. It may be that the effects of breathiness, particularly at milder levels of severity, do not substantially affect cycle boundaries and time-based measures of phonatory characteristics. Wolfe and Steinfatt⁵ have indicated that the laryngeal irregularities contributing to the turbulent airflow observed in breathiness may be less complex than those observed in other voice types. Eskanazi et al²¹ have stated that breathy voices are closer to normal voices than other voice types. This view of breathiness as similar in many respects to normal voice productions is consistent with our observation that, in the prediction of breathy voice among all other voice types, a number of subjects were misclassified into the normal group. Many of the breathy voice signals were observed to have relatively strong underlying periodicity combined with the additive noise component of turbulent airflow. It is, therefore, not unreasonable for certain breathy voices to be misclassified as within the realm of normal voice both perceptually and acoustically. In addition, two subjects from the breathy group were misclassified into the hoarse group. As turbulent airflow is a characteristic common to both groups, it is understandable how this type of misclassification can occur.

Normal vs. Rough Voice

Rough voice has been said to correspond to irregular vocal fold vibration, in which vibratory patterns are unstable and sensitive to subglottic pressure, may be diplophonic in nature, amplitude modulated, and characterized by the presence of subharmonics in the spectrum as well as increased perturbation.^{38–40} This description of the possible characteristics of roughness emphasizes the need for both spectral and time-based analysis methods, and it is supported by the results of this study, wherein a four-variable model consisting of time (LOGSHIM and

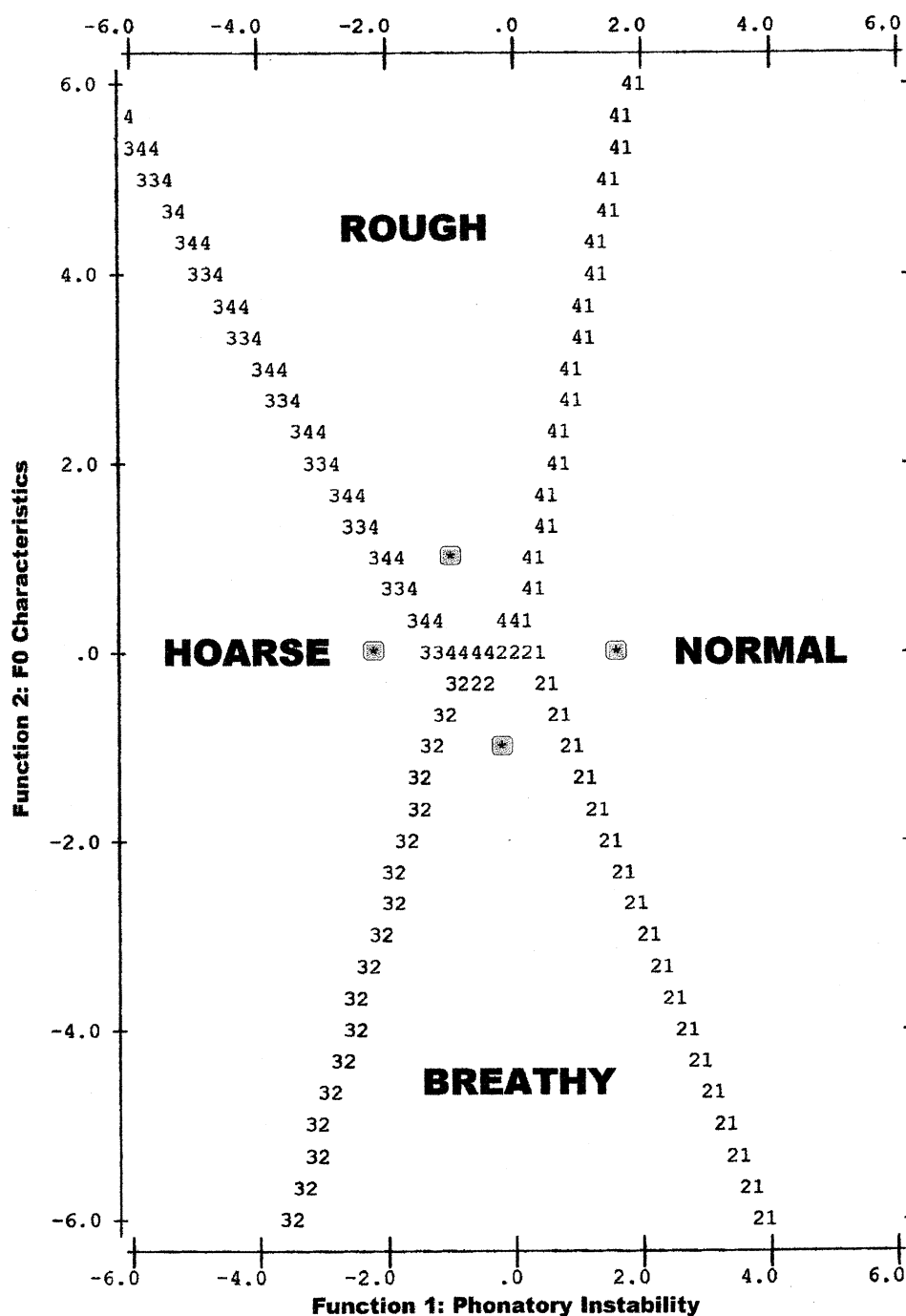


FIGURE 2. Territorial map depicting the separation between voice types (Normal = 1; Breathy = 2; Hoarse = 3; Rough = 4). Group centroids (discriminant function means) for the four groups are indicated by *.

INVSQRTF₀) and spectral-based measures (CPP/EXP and DFTR) produced a 93.4% success rate in classifying normal vs. rough subjects. The logarithm

of shimmer was observed to be the strongest contributor to the four-variable model, consistent with the irregularity of vocal fold vibration that is often

believed to be characteristic of the rough voice type. In addition, the CPP/EXP ratio was also a significant contributor, which may reflect increased amplitude of noise components in relation to the F_0 .

A review of the group means (see Table 2) for each of the key acoustic variables used in the various discriminant analyses provides further insight into the possible characteristics of rough voice. First, a decrease in the vocal F_0 (ie, increase in INVSQRTF_0) was observed, which may reflect the addition of low-frequency noise components and subharmonic tendencies often observed in rough voice. The possible relationship between F_0 and perception of dysphonic voice type (particularly roughness) has been described in several previous reports.^{6,21,38-43} Second, although increased pitch sigma (ie, reduced INVSQRTSIG), increased shimmer (ie, increased LOGSHIM), and a decreased CPP/EXP ratio were observed, the DFTR showed relatively little change from the normal group mean. This may indicate that many of the rough voice samples in this study had noise components concentrated in the low-frequency region of the spectrum versus the high-frequency noise observed in breathy voices.

Review of the territorial map (Figure 2) and group centroids indicated that the normal vs. rough classification was not as distinct as that observed for the normal vs. hoarse voice samples. In addition, three of the “rough” subjects were misclassified as within the normal group during the prediction of rough voice among all other voice types. As in the previous discussion regarding breathy voice, rough voice may share many acoustic characteristics in common with normal voices,³⁸ with similarities occurring particularly in speakers with lower F_0 's. These similarities may make the perceptual and acoustic discrimination of normal vs. rough voice types difficult in certain cases.

In the overall prediction of voice type, three of the rough voices were misclassified as hoarse. As irregularity of vocal fold vibration is common to these two groups, it is understandable that some misclassifications among these two groups may occur. Two of the rough subjects were also misclassified as breathy. It may be that these rough subjects may have actually been better described as harsh, in which high-frequency noise predominates rather than the low-frequency perturbations seen in

roughness.⁴² The presence of high-frequency noise in harsh voice may have some similarity to the high-frequency noise also observed in breathiness. If so, acoustic methods of voice classification may encounter some difficulty separating these two voice types. These subjects may also have had no substantial reduction in their vocal F_0 as compared with those subjects with rough voice who did have substantial emphasis in the low-frequency components of their signal.

Normal vs. Hoarse Voice

The greatest degree of success was observed for the hoarse vs. normal voice type distinction (97.4% predictive accuracy). In addition, the normal vs. hoarse classification showed the largest difference in group centroids as seen in the territorial map (Figure 2). Hoarseness has been said to originate from either (1) a fluctuation in vocal fold vibration or (2) turbulent airflow at the glottis.³⁸ Wolfe and Steinfatt⁵ indicate that laryngeal irregularities and turbulent airflow may spread and intensify noise components within the spectrum, as well as obliterate harmonics. This combination of both vocal fold irregularity and turbulent airflow may be expected to produce a voice type that would be most dissimilar from the normal voice type.

A four-variable model (again composed of both time- and spectral-based measures) was found to be most effective in discriminating between normal vs. hoarse groups. The logarithm of shimmer was again the strongest contributor to the discriminant function. This measure of short-term variability was combined with a long-term variability measure (pitch sigma) to perhaps reflect vocal fold irregularity—however, accepting that hoarseness may represent a hybrid descriptor (breathy and rough voice combined to varying degrees), spectral-based measures (DFTR and CPP/EXP) were also important in accounting for the breathy component of this voice type (ie, accounting for factors such as spectral tilt and cepstral flatness).

As compared with all other voice types, none of the hoarse subjects were misclassified as normal. However, several hoarse subjects were misclassified by our acoustic model into either the breathy or rough categories. Judges were not asked to provide a clear indication of which component (breathiness

or roughness) was most prominent in each hoarse voice sample. Future studies that provide separate categories for “breathy” hoarseness versus “rough” hoarseness may achieve greater accuracy in voice type classification as well as provide further insight into this highly variable and complex voice type.

Interdysphonic Differences

Breathy vs. hoarse voice types were separated by a single variable—shimmer (predictive accuracy of 84.5%). As previously stated, it would appear that increased shimmer may be a characteristic of the irregularity of vocal fold vibration found in the rough voice component of hoarseness. Subjects with hoarseness may be similar to the breathy voice subjects with respect to DFTR, because increased high-frequency emphasis and spectral tilt would be expected to be a common feature between these two groups. Breathy vs. rough groups were accurately discriminated using a two-variable model incorporating shimmer and F_0 . The tendency for a low-frequency F_0 in rough subjects is consistent with characteristics such as the presence of strong low-frequency noise components and/or subharmonics. In examples of subharmonics, it is often observed that periodicity is actually achieved across alternate cycles. Therefore, it seems reasonable that acoustic analyses should result in a reduced estimated F_0 in rough voice types.^{38,40} The hoarse vs. rough groups were discriminated using a three-variable model incorporating time-based (F_0 and shimmer) and spectral-based (DFTR) measures. As hoarseness may be a hybrid classification of breathy and rough voice types, the discriminating DFTR variable is reflective of the spectral tilt often described for the breathy voice type. It is this increase in high-frequency noise and subsequent spectral tilt that appears to be key in separating hoarse from rough groups, because both groups may be expected to show irregularity of vocal fold vibration (as measured via shimmer). As indicated earlier, a lowered F_0 appears to be particularly characteristic of the rough voice type.

To reiterate, it is our view that measures of the CPP provide a general measure of dysphonia sensitive to various dysphonic types. Although this measure may be most effective in discriminating normal from dysphonic states, it may not be particularly useful for separating dysphonic types from each other. This

view is consistent with previous observations by Dejonckere and Wieneke¹¹ and Wolfe and Martin,⁷ and it is confirmed in our own analysis of interdysphonic differences. For all interdysphonic comparisons, other acoustic measures were key to successful classification, aside from the cepstral measure.

Limitations

Several limitations in the methodology of this study should be noted. Revisions in future methodology may provide additional insight into the acoustic prediction of voice type:

1. This study assessed characteristics of the normal and dysphonic *female* voices only. Because male subjects were not included, it is unclear whether the models and acoustic variables identified in this study would be the same when predicting voice type and severity in male voices. Future studies comparing prediction models for men vs. women may provide further insights into possible gender effects on the perception and acoustic analysis of dysphonic voice type.
2. The classification of voice was based on four traditional categories, including three commonly used dysphonic categories (breathy, rough, and hoarse). These traditional categories were selected because they are ubiquitous and familiar to most voice clinicians. However, other voice types and quality deviations such as strain and harshness were not specifically accounted for in this study. Some of the inaccuracy in voice typing may be related to the range of voice types/classifications employed. Future studies that incorporate a larger range of classifications might lead to improved accuracy in voice type classification.
3. We have focused on a particular set of time and spectral-based analysis methods based on their demonstrated effectiveness in numerous past research studies. However, other acoustic measurement methods may also provide important additions to the accuracy of voice type prediction. As an example, Michaelis et al⁴⁴ have reported that a measure referred to as the glottal-to-noise excitation ratio (GNE) may also be effective in characterizing different

voice types and pathological cases. Future studies dealing with the relative effectiveness and contributions of acoustic analysis methods are necessary if we are to eventually isolate a core set of parameters that will be most effective in characterizing both normal and disordered voice.

4. The results of this study come from a particular computer program and set of algorithms. It is well known that interprogram variability may result in variations in analysis results even with similar voice samples. Therefore, it is possible that alternative models and acoustic variables would have been identified in the event that different program/algorithms had been used. However, the results of this study do appear to be consistent with the past findings from various studies, indicating that similarity in theoretical foundation may have precedence over the specifics of how these acoustic analysis algorithms are designed and employed in code.

CONCLUSIONS

Relying on one acoustic measure by itself is clearly not the most effective approach to prediction of voice type. The results of this investigation suggest that a comprehensive analysis of the voice signal incorporating both time- and spectral-based measures is necessary to accurately evaluate the highly variable signals observed both between and within dysphonic voice types. The use of spectral-based measures provide a means to quantify the dysphonia, while reducing possible inaccuracies related to identifying cycle boundaries associated with aperiodic signals. On the other hand, traditional time-based measures are understandable, easily communicated between clinicians and to patients, and do have a literature supporting their use and relationship to perceptual impressions of the voice. The combination of these various measures provides an easily applicable method for voice type prediction that can be applied to the diversity of voices encountered clinically.

Because of the heterogeneity of dysphonia, even within particular voice types, it has often been observed that the variables successful in predicting

voice type in one study are not necessarily found to be as successful in other studies. Future studies need to apply a successful set of variables to different sets of voices in a systematic manner. It is our intention to apply the successful models identified in this study to alternative groups of normal, rough, hoarse, and breathy voices to assess (1) the consistency of these models in successfully predicting voice type, and (2) if necessary, to revise these models in such a way that they are applicable to a wider variety of voices within each of the traditional dysphonic types addressed in this study.

Acknowledgment: We thank Drs. M. Razzaghi (Dept. of Mathematics, Computer Science, and Statistics, Bloomsburg University) and R. Prosek (Dept. of Communication Sciences and Disorders, Penn State University) for their helpful comments and suggestions regarding this study. This study was supported by a Bloomsburg University Sabbatical Leave awarded to the first author (Fall, 2001).

REFERENCES

1. Fröhlich M, Michaelis D, Strube HW. Acoustic "breathiness measures" in the description of pathologic voices. In: *Proceedings ICASSP 98*. Seattle, WA: 1998:937-940.
2. Colton RH, Casper JK. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. 2nd ed. Baltimore, MD: Williams & Wilkins; 1996.
3. Takahashi H, Koike Y. Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Otol (Stock)*. 1975;338:1-24.
4. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*. 1982;71:1544-1550.
5. Wolfe V, Steinfatt TM. Prediction of vocal severity within and across voice types. *J Speech Hear Res*. 1987;30:230-240.
6. Martin D, Fitch J, Wolfe V. Pathologic voice type and the acoustic prediction of severity. *J Speech Hear Res*. 1995;38:765-771.
7. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. *J Comm Dis*. 1997;30:403-416.
8. Rabinov CR, Kreiman J, Gerratt B, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Hear Res*. 1995;38:26-32.
9. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*. 1996;39:126-134.

10. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res.* 1994; 37:769–778.
11. Dejonckere PH, Wieneke GH. Cepstra of normal and pathological voices: Correlation with acoustic, aerodynamic and perceptual data. In: Ball MJ, Duckworth M, editors. *Advances in Clinical Phonetics*. Amsterdam: John Benjamins Publishing Co.; 1996:217–226.
12. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res.* 1996;39:311–321.
13. Callan DE, Kent RD, Roy N, Tasko SM. Self-organizing maps for the classification of normal and disordered female voices. *J Speech Hear Res.* 1999;42:355–366.
14. Wolfe VI, Martin DP, Palmer CI. Perception of dysphonic voice quality by naïve listeners. *J Speech Hear Res.* 2000; 43:697–705.
15. Klich RJ. Relationships of vowel characteristics to listener ratings of breathiness. *J Speech Hear Res.* 1982;25:74–80.
16. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am.* 1990;87:820–857.
17. Kojima H, Gould WJ, Lambiasi A, Isshiki N. Computer analysis of hoarseness. *Acta Otolaryngol.* 1980;89: 547–554.
18. Kitajima K. Quantitative evaluation of the noise level in the pathological voice. *Folia Phoniatr.* 1981;33:115–124.
19. Noll AM. Short-term spectrum and “cepstrum” techniques for vocal pitch detection. *J Acoust Soc Am.* 1964;41: 293–309.
20. Heman-Ackah Y, Michael D, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice.* 2002;16:20–27.
21. Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *J Speech Hear Res.* 1990;33:298–306.
22. Coyle SM, Weinrich BD, Stemple JC. Shifts in relative prevalence of laryngeal pathology in a treatment-seeking population. *J Voice.* 2001;15:424–440.
23. *CSL Computerized Speech Lab*. [Computer program]. Pine Brook, NJ: Kay Elemetrics; 1994.
24. Awan SN, Frenkel ML. Improvements in estimating the harmonics-to-noise ratio of the voice. *J Voice.* 1994;8: 255–262.
25. Awan SN. Automatic estimation of vocal harmonics-to-noise ratio using cepstral analysis. In: Windsor F, Kelly ML, Hewlett N, editors. *Investigations in Clinical Phonetics and Linguistics*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:475–482.
26. Gold B, Rabiner L. Parallel processing techniques for estimating pitch periods in the time domain. *J Acoust Soc Am.* 1969;46:442–448.
27. Papamichalis PE. *Practical Approaches to Speech Coding*. Englewood Cliffs, NJ: Prentice-Hall; 1987.
28. Titze IR, Horii Y, Scherer RC. Some technical considerations in voice perturbation measurements. *J Speech Hear Res.* 1987;30:252–260.
29. Baken RJ. *Clinical Measurement of Speech and Voice*. Boston, MA: Little, Brown and Co.; 1987.
30. Frokjaer-Jensen B, Prytz S. Registration of voice quality. *Bruel Kjaer Tech Rev.* 1976;3:3–17.
31. de Krom G. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res.* 1995;38:794–811.
32. Cordes AK. The reliability of observational data: I. Theories and methods for speech-language pathology. *J Speech Hear Res.* 1994;37:264–278.
33. American Speech-Language-Hearing Association Division 3, Voice and Voice Disorders. (2003). *Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) – Instructions and Rating Form*. [Electronic Version]. Available at: http://www.asha.org/about/membership-certification/divs/div_3.htm.
34. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J Speech Hear Res.* 1993;36:14–20.
35. Rust RT, Cooil B. Reliability measures for qualitative data: theory and implications. *J Marketing Res.* 1994;31:1–14.
36. *SPSS 10.0* [Computer program]. Chicago, IL: SPSS, Inc.; 1999.
37. Kreiman J, Gerratt BR, Percoda K. Listener experience and perception of voice quality. *J Speech Hear Res.* 1990;33: 103–115.
38. Omori K, Kojima H, Kakani R, Slavik DH, Blaugrund SM. Acoustic characteristics of rough voice: Subharmonics. *J Voice.* 1997;11:40–47.
39. Imaizumi S. Acoustic measures of roughness in pathological voice. *J Phonet.* 1986;14:457–462.
40. Titze I. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall; 1994.
41. Murry T, Singh S, Sargent M. Multidimensional classification of abnormal voice qualities. *J Acoust Soc Am.* 1977;61: 1630–1635.
42. Askenfelt A, Hammarberg B. Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *J Speech Hear Res.* 1986;29:50–64.
43. Wolfe VI, Ratusnik DL. Acoustic and perceptual measurements of roughness influencing judgments of pitch. *J Speech Hear Res.* 1988;53:15–22.
44. Michaelis D, Gramss T, Strube HW. Glottal-to-noise excitation ratio: A new measure for describing pathological voices. *Acustica/Acta Acustica.* 1997;83:700–706.