# Acoustic and Perceptual Correlates of Breathy Vocal Quality

Ronald Allen Cleveland

# ACOUSTIC AND PERCEPTUAL CORRELATES OF BREATHY VOCAL QUALITY

by

Ronald Allen Cleveland

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Speech Pathology and Audiology

Western Michigan University
Kalamazoo, Michigan
December 1991

ACKNOWLEDGMENTS

My thanks go to the members of my committee, Dr. Robert Erickson, Dr. Gary Lawson, and Dr. James Hillenbrand for their tutelage and encouragement on this project and throughout my graduate career. Thank you, Dr. Erickson, for your guidance as committee chairman. I would especially like to express my appreciation to Dr. Hillenbrand for the answers he has given me and for showing me, through example, ways in which to ask questions. I am grateful to Mr. William Dawson for his extensive technical help in collecting the recorded samples. My deepest "thanks" to my parents, Guy and Pauline Cleveland, for their love and support. This work is dedicated to my friend, Casey Cole.

Ronald Allen Cleveland

ACOUSTIC AND PERCEPTUAL CORRELATES OF BREATHY VOCAL QUALITY

Ronald Allen Cleveland, M.A.

Western Michigan University, 1991

Recordings were made of seven normal female and eight normal male talkers producing sustained vowels under normal, moderately breathy, and very breathy conditions. Twenty listeners judged the recorded vowel tokens for degree of breathiness using a direct magnitude estimation procedure. A Cronbach coefficient alpha revealed strong intrasubject agreement.

Several acoustic analyses were evaluated by measuring their correlations with the mean of the listeners' breathiness ratings. Measures of cepstral peak prominence in band limited signals were most strongly correlated with perceived breathiness. The height of the autocorrelation peak in highpass filtered signals and the relative amplitude of the first harmonic were also found to correlate with perceived breathiness.

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# INTRODUCTION

The purpose of this study was to investigate the psychoacoustic nature of the human voice quality which is commonly referred to as breathy, murmured, or asthenic. The need for such investigation is clearly justifiable given three current conditions: (1) the paucity of knowledge regarding the physio-acoustic correlates of normal and diseased voices in general (Ludlow, 1981), and breathy voices in particular (Zemlin, 1968); (2) the frequency with which breathy dysphonia is encountered by the voice therapist (Zemlin, 1968); and (3) the subjective nature of classifying abnormal voice characteristics which has resulted in poor reliability and a lack of consensus among therapists (Aronson, 1990; Boone & McFarlane, 1988; Jensen, 1965; Ladefoged, 1983; Weiss & Lillywhite, 1976; Zemlin, 1968). The essentially unidimensional qualities of pitch and loudness perception have been quantified with the *mel* and *sone* scales respectively (Stevens, 1975). Measuring perceived breathiness, hoarseness, harshness, throatiness, shrillness, or any of the more than three dozen qualitative descriptors that have been used (Perkins, 1971) presents a still greater challenge.

Breathy vocal quality is one indicator (indeed, at times the earliest indicator [Aronson, 1971]) of numerous pathological conditions affecting vocal fold physiology (Aronson, 1990; Colton & Casper 1990). Such conditions could include hyperthyroidism; laryngeal trauma; neoplasms such as polyps, nodules, webbing and carcinomas; neurologic disorders such as lesions of the vagus nerve, myasthenia gravis, parkinsonism, spastic dysphonia and a host of others leading to laryngeal dysarthria (Darley, Aronson, & Brown, 1969a, 1969b).

Breathy (murmured) voice is a normal aspect of the running speech of many of the world's languages, including English (Fisher-Jorgensen, 1976; Huffman, 1987; Klatt & Klatt, 1990; Ladefoged, 1975, 1983). For that reason, a greater understanding of breathy phonation can be used in the study of phonetics and the improvement of voice synthesis techniques (Klatt & Klatt, 1990).

Although evidence appears to be limited, some researchers believe that there are several variables which correlate with the perception of the aged voice (Hollien, 1987; Ryan & Burk, 1974). Ryan and Burk (1974) found breathiness (air loss) to be one of the major variables in the prediction of perceived vocal age. Breathiness may well be the result of the physiological effects of aging on laryngeal muscular control (Hollien, 1987). Breathiness can also be a concomitant of vocal misuse and other functional disorders (Aronson, 1990; Boone & McFarlane 1988).

Breathiness (along with harshness and hoarseness) has been thought to be symptomatic of insufficient glottal closure "caused by poor coordination of vocal cord tension and breath supply" (Fairbanks, 1940, p. 214) during phonation. High speed photography has revealed a posterior glottal opening ("glottal chink") which persists during the "closed" phase of the breathy phonatory cycle (Hillenbrand, Metz, Colton, & Whitehead, 1990; Zemlin, 1968). Breathy quality, along with lowered habitual pitch, may be a social marker of feminine sexuality among English speakers (Klatt &

Klatt, 1990; McKay, 1987). Recently, researchers have found that breathy voice is often present during phrase endings as the larynx begins to prepare for the next inhalation (Klatt & Klatt, 1990). Klatt and Klatt surmised that this laryngeal posture is achieved in one of two ways:

> (1) a general 'relaxed' separation of the arytenoids or (2) a 'laryngealized' mode in which the abduction is accompanied by a rotational motion of the arytenoids such that some medial compression is applied to keep the folds vibrating in a nearly normal way in spite of the opening at the posterior. (p. 821)

They further speculate that this latter form of "breathy-laryngealized" phrase ending "may be a social marker of maleness" (p. 821). Phonation through a very narrow persistent opening along the length of the vocal folds has also been observed in the production of the English "murmured h" (Ladefoged, 1975).

Irrespective of the cause, when the glottis does not achieve complete approximation during the phonation cycle a certain amount of the breath stream is spent in the generation of turbulence. The narrowing of the airway through an existing glottal chink creates an increase in air particle velocity and a proportional drop in transglottal air pressure. Air particle collision at the glottal narrowing causes airflow patterns to eddy above the glottis. A number of authors have reported the importance of the Reynolds' number as a predictor of turbulent air flow (Zemlin, 1968; Minifie, 1973). The Reynolds' number is computed as $Re = vh/V$, where $v$ = particle velocity, $h$ = the width of the opening, and $V$ = the coefficient of viscosity for air. It has been estimated that a Reynolds' number of about 1,800 or greater is necessary for turbulent air flow (Zemlin, 1968).

The present study examines the psycoacoustic nature of breathy voice. Clear and breathy voice tokens were the basis of comparison between digital acoustic analyses and listeners' subjective ratings. The project began by collecting recorded samples of clear and simulated breathy voice from normal talkers. These samples were edited and then analyzed using a number of acoustic measurements that are thought to correlate with breathy vocal quality. The voice samples were also played for a group of listeners who rated them for degree of breathiness. These ratings were then correlated with the acoustic analyses to determine which objective measures appeared to best predict perceived breathiness.

# REVIEW OF THE LITERATURE

## Introduction

Research which directly investigates the acoustic parameters of perceived breathy vocal quality (apart from hoarseness or roughness) is somewhat sparse. While a number of speech scientists have been interested in quantifying breathiness (Fritzell, Hammarberg, Gaufinn, Karlsson, & Sundberg, 1986; Fukazawa, El-Assouty, & Hanjo, 1988; Hillenbrand et al., 1990; Klatt & Klatt, 1990; Rothenberg, 1983), the earliest investigations into the acoustic nature of breathy (murmured) voice, appear to have been written more from the linguist's perspective (Fischer-Jørgensen, 1967; Dave, 1967; Pandit, 1957). For example, Pandit examined the phonemic features of nasalization, aspiration, and murmur in Gujarati, a language spoken predominantly in Western India. Gujarati includes a set of murmured vowels which contrast phonemically with a set of fully voiced cognates. Using kymographic and spectrographic analysis, Pandit (1957) demonstrated that murmured vowels were not "suction stops" or "implosives" as had been argued by others. A suction stop or an implosive requires some period of glottal stop closure. He reported that no stopping of the voice occurs in the case of murmured vowels.

Ladefoged (1983), in a brief comparative analysis of phonation types used among a variety of languages, pointed out that one culture's voice disorder (i.e., breathy voice) may be a phonemic feature of another. Ladefoged encouraged speech therapists and scientists to study the phonation types found in other languages in order to obtain data from a uniform population. In examining breathy voice, a number of investigators have utilized vowel tokens from languages in which breathiness is a distinctive feature (Bickley, 1982; Fischer-Jørgensen, 1967; Huffman, 1987). Some have used pathologically breathy voicing samples (Fritzell et al., 1986; Fukazawa et al., 1988). Still others have used normal non-breathy (Klatt & Klatt, 1990; Ptacek & Sander, 1963) and normal simulated-breathy voice signals (Hillenbrand, 1990; Klich, 1982). At times the use of the human voice has been abandoned altogether in favor of synthetic speech (Bickley, 1982; Klatt & Klatt, 1990). The parameters of synthetic speech can be more easily manipulated for the purposes of listening experiments.

The following review covers many of the investigations into acoustic and perceptual correlates of breathy voice that have been conducted since Pandit's (1957) early study.

## Fundamental Frequency

Phoneticians have investigated fundamental frequency ($f_0$) differences between clear and murmured (breathy) cognates in languages which have such phonemic distinctions. Pandit (1957) described murmur as "voiced breath, low pitched and simultaneous with the vowel" (p. 169) and "as *sotto voce*, with voicing and slight lowering of the pitch" (p. 170). Interestingly, Pandit's contention that Gujarati murmured vowels have a lower $f_0$ than do clear vowels was subjectively supported by Dave (1967), but Dave's acoustic analysis revealed no such difference. Fisher-

Jorgensen (1967), utilizing both the recordings previously analyzed by Dave and other recordings of native Gujarati speakers, examined $f_0$ differences more carefully. She measured fundamental frequency at four selected points within the vowel. Her results supported Pandit's earlier contention that breathy vowels are generally lower pitched than their clear counterparts. Lower frequency at the beginning of the murmured vowels seemed to account for most of this difference. According to Fischer-Jørgensen, Dave's conclusion that no $f_0$ difference exists was the result of measurement error. His single measurement at the midpoint of the vowel did not account for the suprasegmental feature of intonation. In samples which were collected from ten speakers of !Xóõ (a language of Southern Africa in which breathiness is a distinctive feature), Ladefoged (1983) also noted that murmured vowels were lower in pitch than their fully voiced counterpart.

Ptacek and Sander (1963) described a listening experiment that used recordings of the vowel /a/ spoken by normal talkers. One part of their study examined the effects of vowel $f_0$ on listeners' perception of breathiness. Each talker was recorded using two phonation frequencies (400 & 210 Hz for females and 250 & 130 Hz for males). While murmured vowels in both Gujarati and !Xóõ appear to be lower pitched than their clear cognates, Ptacek and Sander demonstrated that on average listeners rated the higher pitched vowels to be breathier for both male and female talkers. On the contrary, Klatt and Klatt (1990) found that within gender perception of breathiness was not at all correlated with the average $f_0$ of the talker.

## Formant Changes

Dave (1967) stated that murmured vowels show a weakening of the first formant ($F_1$) with weak and split higher formants ($F_2$, $F_3$, etc.). However, these same features were found in a few spectrograms of clear vowels as well. Formant frequencies for murmured and clear vowels were not found to be significantly different. Due to the added tracheal coupling in breathy vowels a higher frequency $F_1$ was anticipated by Fischer-Jørgensen (1967). This was found to be the case for some vowels and some speakers but not for others. Fischer-Jørgensen concluded that the formant frequency differences for breathy and clear vowels are small and inconsistent.

When the glottis remains partially open during the voicing cycle, effects of the acoustic coupling of the trachea with the vocal tract have been seen in spectral analyses. Klatt and Klatt (1990) investigated the [h] portion of [ha] syllables. They found that extra poles (resonant peaks or formants) and zeros (anti-resonances or local energy minimums) occur in the spectrum envelopes of [h] and could be seen, albeit reduced, in the following vowels. These extra poles tended to occur at about 600, 1400, and 2200 Hz for female voice samples and at three slightly lower frequencies for males. Tracheal coupling was also thought to be responsible for a flattening of the $F_1$ peak and for corresponding increases in $F_1$ bandwidth.

Glottal Waveform and First Harmonic Amplitude

Since breathy phonation is believed to be chiefly the result of changes in the glottal source wave (Bickley, 1982) arising from changes in glottal posture (Hillenbrand et al., 1990; Zemlin, 1968), a number of researchers have utilized the technique of inverse filtering. Inverse filtering attempts to extract the glottal source volume velocity waveform from the oral-output signal by canceling the resonant filtering effects of the vocal tract and lip radiation (Miller, 1959; Rothenberg, 1973). Traditionally, the technique of inverse filtering has employed either a long reflectionless metal tube (Monsen, 1981; Sondhi, 1975) or a series of analogue or digital filters (Davis, 1981). The latter has been used most often by researchers who were interested in changes in glottal volume velocity waveforms that occur with breathy phonation (Bickley, 1982; Huffman, 1987; Fischer-Jørgensen, 1967; Fritzell et al., 1986). Rothenberg (1973) suggested that:

> the empirical definition that non-breathy voicing in the modal (chest) register be identified by an air flow waveform that has a marked flat or almost flat region, due to the coming together of all or part of the vocal folds, and that the minimum air flow be less than 10% or 15% of the peak air flow. (p. 1642)

Bickley (1982) found that inverse filtered waveforms for Gujarati clear phonation exhibited the above noted flat closed phase. This closed phase comprised about one third of the total vibratory cycle. She also found a typical asymmetry in which the opening phase of each glottal cycle was slower, sloping more gradually than the closing phase. Her Gujarati breathy samples, however, demonstrated more symmetrical opening and closing phases with little or no complete closed phase. The graphic representations of the breathy signals more closely resembled sinusoids rather than the normal glottal pulse train. It is believed that the more sinusoidal contour of the breathy glottal wave created by a longer open phase is responsible for measured increases in first harmonic ($H_1$) amplitude (Bickley, 1982; Chapin-Ringo, 1988; Fischer-Jørgensen, 1967; Klatt and Klatt, 1990; Ladefoged, 1983). Enhanced $H_1$ amplitude in the spectra of breathy speech signals has been observed by a number of investigators and is generally believed to be an acoustic correlate of breathy voice (Bickley, 1982; Huffman, 1987; Fischer-Jørgensen, 1967; Klatt and Klatt, 1990; Hillenbrand, 1990).

Some limited inverse filtering experiments conducted by Fischer-Jørgensen (1967) indicated that the murmured glottal wave has a shortened and less dramatic closed phase, and a more sinusoidal shape when compared with clear vowels. She considered the most salient spectral feature of Gujarati murmured vowels to be the relatively high intensity of $H_1$ compared with higher frequency components (i.e., $F_1$ through $F_4$, $H_2$, and $H_3$). Highpass filtered murmured vowels showed greater attenuation than did highpass filtered clear vowels. That is, when the low frequency components were filtered out the breathy vowels appeared to lose more energy than did the clear vowels. This was taken as a measure of greater first harmonic strength in the breathy samples. Breathy and clear spectrograms were analyzed for the absolute amplitude of $H_1$, as well as the amplitude of $H_1$ relative to higher frequency components. For the most part, these amplitudes and relative amplitudes of $H_1$ were greater for the breathy vowels.

Ladefoged (1983) found the difference in amplitude between $H_1$ and the peak of $F_1$ to be the most promising indicator of degree of breathiness. This $H_1/F_1$ dB difference was found to be significantly greater for clear phonation than for murmured phonation. That is to say, the amplitude of $H_1$ was found to increase along with the amount of breathiness.

In a study regarding correlates of voicelessness in stop-plosives, Chapin-Ringo (1983) reported on $H_1$ and $H_2$ relative amplitudes in vowels following voiced and unvoiced stop consonants. Interestingly, the first few pitch periods of vowels following voiceless stops generally showed $H_1$ amplitude to be greater than $H_2$ amplitude. Vowels adjacent to voiced stops showed the opposite. A pilot listening study using synthetic stimuli revealed that the difference in voice onset time between perceived voiced and voiceless tokens could be reduced when $H_1$ was greater than $H_2$.

Huffman (1986) used inverse filtering to derive glottal waveforms from samples of four phonation types used in Hmong. This Southeast Asian language includes one breathy and three non-breathy voicing features. Breathy samples showed a mean $H_1/H_2$ ratio of 9.48 dB, while non-breathy samples showed no significant differences between $H_1$ and $H_2$.

Klatt and Klatt (1990) generally found amplitude ratios of $H_1$ to $H_2$ were greater for female subjects than males by an average of 5.7 dB. This implies a longer open quotient in the female glottal cycle. (Open quotient is defined as the duration of the open phase divided by the duration of the entire glottal cycle [Hirano, 1981].) Considerable within-gender subject-to-subject variability in $H_1$ measures was noted.

While [ha] utterances were perceived as significantly breathier than [?a] utterances, no significant difference in $H_1/H_2$ amplitude measures was found at the mid-vowel.

### Spectral Slope and Aspiration Noise

In the normal voicing cycle complete closure occurs at some point in the cycle (Stevens, 1977). The point of complete glottal closure is thought to be responsible for the high frequency excitation of the vocal tract which helps determine the strength of $F_2$ and higher formants (Rothenberg, 1983). The harmonic spectra of normal glottal volume velocity waveforms have a slope of about -12 dB per octave with regularly occurring harmonics beyond 4 kHz. In the case of breathy voice no sharp closure occurs during the glottal cycle. Vocal fold adduction begins anteriorly and migrates posteriorly along the vocal folds, stopping short of complete closure before the opening phase begins. The resulting smoothed waveform creates a reduction of glottal source high frequency spectral components (Klatt & Klatt, 1990; Rothenberg, 1983; Stevens, 1977). Along with the increase in $H_1$ amplitude, the harmonic spectra of breathy glottal signals have a steeper slope (about -18 dB/octave) with regular harmonics only up to about 1.5 to 2 kHz. Keep in mind, however, that it is not the glottal volume velocity wave that the listener hears. While the slope of the glottal source spectrum is relatively steep, the slope of the oral-output signal appears to flatten out as a function of breathiness. The missing harmonics are replaced by a dense aspiration noise spectrum which excites the higher frequency vocal tract resonances.

In his early study, Pandit (1957) noted that the spectrograms of murmured vowels contained randomly distributed energy, most noticeable in the higher frequencies, while the spectrograms of fully voiced vowels did not. Based on her spectrographic observations, Fisher-Jørgensen (1967) believed differences in amounts of spectral noise between murmured and clear vowels were small and inconsistent.

Klich (1982) inspected narrowband spectrograms of simulated breathy vowels. He reported decreases in harmonic definition and increases in noise as breathiness was increased. Four speaking conditions were recorded for analysis: normal, mildly breathy, severely breathy, and whisper. Based on spectrograms, Klich measured relative amounts of energy in three frequency bands as a function of breathiness. As breathiness increased, relative energy appeared to decrease in the 100 - 500 Hz range, remain the same in the 1500 - 2500 Hz range, and increase in the 3500 - 4500 Hz range. This reduction of low frequency energy and concomitent increase in high frequency energy resulted in a flattening of the breathy spectrum.

Bickley (1982) observed "irregularities" at high frequencies in wideband spectrograms of !Xóõ breathy vowels. Although she speculated that additive noise may be a characteristic of breathy voice, she noted that the variability of noise between breathy vowel tokens made it difficult to quantify. Ladefoged (1983) reported that the amount of irregularity (noise) seen in the upper frequency spectrum was greater for breathy recordings. He recorded time domain waveform plots of normal oral output signals which clearly revealed the fundamental period along with the periodic $F_1$ damping effects of the vocal tract. On the other hand, waveforms of breathy (murmured) signals were more sinusoidal in appearance with superimposed aperiodic high frequency components. Evidence of $F_1$ damping on the time domain wave were difficult to detect in breathy waveforms, ostensibly due to the amount of added aperiodicity. Like Bickley (1982), Ladefoged did not feel that additive noise was the best means for quantifying breathiness, at least when relatively brief vowel tokens were being analyzed. Ladefoged found the harmonic spectrum morphology of fully voiced and murmured utterances are distinctly different from one another. He considered spectral tilt as a measure of breathiness since murmured samples exhibited a flatter spectrum. While the more sinusoidal glottal source waveform results in decreased upper harmonic amplitude, glottal turbulence produces greater high frequency excitation. Due to these two seemingly opposing factors, Ladefoged felt that spectral tilt was a more difficult measure than $H_1/F_1$ amplitude.

Huffman (1986) tried to measure spectral tilt for inverse filtered Hmong breathy and non-breathy signals. Spectra between 0 and 2 kHz derived from discrete Fourier transforms were fitted with regression lines across harmonic peaks. However, these were found to be unreliable characterizations of tilt for many of the spectra.

Fukazawa et al., (1988) reported the effects of pathological breathiness on oral output voice signal periodicity. Acoustic analysis of a normal talker's /a/ revealed a clearly periodic complex signal in the time domain and a well defined harmonic spectrum which extends to about 4.5 kHz. The same data for a talker with a vocal cord polyp showed decreased periodicity at the end of each time domain cycle. Power spectrum analysis for this talker showed clearly defined harmonics only up to about 2.0 kHz. A third sample taken from a patient with laryngeal cancer revealed a time domain plot similar to white noise with a nearly non-distinguishable fundamental period and a relatively flat harmonic spectrum with $H_1$ being the only harmonic present.

Fukazawa et al. devised what they felt was an objective means for indexing pathologic breathiness independent of hoarseness. This measure is essentially reflective of changes in high frequency spectral energy. Subjects chosen for this study included 24 normals and 31 patients presenting a variety of laryngeal pathologies characterized by hoarse voice. Recordings of a sustained /a/ were made from each of the subjects. These were digitally processed, high-pass filtered (+12 dB/octave slope) to measure the energy of high frequency components, and given a breathiness (Br) index rating. The Br index was defined as the energy of the high frequency components divided by the energy of the total signal. For normal voices the Br index was almost always computed to be below 50, while it increased to several hundred for most of the pathological voices, depending on the degree of turbulence created (Fukazawa et al., 1988).

More recently, Klatt and Klatt (1990), looking for gender related breathiness differences, used a wide (600 Hz) bandpass filter centered at $F_3$ in order to isolate the third formant of [ha] samples. Unfiltered wave plots, which are dominated by low frequency periodic components, were deemed unsuitable for noise estimation. (This was supported earlier by Kasuya, Ogawa, Mashima, and Ebihara [1986] whose research revealed that even pathological voices sometimes exhibit well developed harmonic structures in the lower frequencies.) A four step scale was then employed for visually estimating the random noise content of these filtered $F_3$ wave plots. Females, on average, were found to generate more random noise, although within-gender variation was large compared to between-gender average differences. Greater noise was also estimated for unstressed and utterance-final syllables. It was conjectured that there is less complete glottal closure during these two conditions. The authors believed that pharyngeal friction may also contribute to additive noise.

## Psychoacoustic Correlational Studies

Fisher-Jørgensen (1967) devised a listening experiment using filtered recordings of breathy and clear vowels. The recordings were presented within the context of Gujarati words to a single listener. Lowpass filtering with a cutoff frequency at about 3.2 kHz, which only affected spectral energy above $F_4$, did not change perception of murmured vowels. Band-stop filtering between about 200 - 500 Hz, preserving the fundamental but reducing $H_2$, also appeared to have little effect on murmured vowel recognition. Contrary to the expected outcome, highpass filtering at 230 Hz, which reduced $H_1$ amplitude by about 25 dB, did not significantly decrease correct identification of murmured vowels. The author concluded that, while she believed the relative amplitude of $H_1$ to be "the most obvious and constant feature" (pp. 133-134), no single acoustic feature was sufficient to cause the sensation of breathiness. Fisher-Jørgensen's attempts at synthesizing breathy Gujarati vowels were largely unsuccessful. It has been hypothesized that the voicing source synthesizer available at the time may have been insufficient and/or aspiration noise should have been incorporated into the signals (Klatt and Klatt, 1990).

A perceptual study was carried out by Bickley (1982) using synthesized Gujarati breathy and clear word pairs which were judged by four native Gujarats. The variables of $H_1$ amplitude and aspiration noise were controlled to generate a continuum of clear to breathy vowels (/a/, /i/, /o/). Formant locations and amplitudes remained constant. Aspiration was added in 5 dB steps over a range of 20 dB. No correlation was found between breathiness ratings and additive noise. First harmonic amplitude was boosted 0, 9, 12, or 15 dB above clear vowel values. Breathiness ratings clearly increased as a function of $H_1$ amplitude. Bickley proposed that what listeners were sensitive to, with regard to a breathiness percept, was the amplitude of $H_1$ relative to that of $F_1$.

Klich (1982) used recordings of females speaking the sentence "He pets a rabid horse every day" under four degrees of simulated breathiness, ranging from normal to whisper. The stimuli were presented to a group of 27 listeners in one 30 minute listening session. The results of this study showed that five vowel parameters were significant in the prediction of listener's ratings. Pearson correlations showed that overall sound pressure level (SPL) was most strongly related to percieved breathiness. A step-wise multiple regression analysis determined that of six measures used only speaking rate did not contribute significantly to the prediction of perceived breathiness. Vowel duration, which increased as a function of perceived breathiness, appeared to be the most significant predictor, although all five parameters were very close in importance. The next most significant predictor, mean overall sound pressure level (SPL), decreased as breathiness ratings increased. The relative amounts of energy within three bands were estimated at vowel midpoints via spectrograms. Energy between 100 and 500 Hz and between 1500 and 2500 Hz relative to total vowel energy were negatively correlated with breathiness. The relative energy between 3500 and 4500 Hz increased along with perceived breathiness.

Fukazawa et al. (1988), in a small listening study, used pathological voice recordings which were rated in terms of breathiness and roughness by two experienced ENT physicians. A significant correlation was found between the two physicians' ratings. There was a strong correlation between the perception of breathiness and the Br index, while there was no significant correlation found between the Br index and roughness ratings. Because the Br index is a digitally automated process, the authors felt it could be a quick, inexpensive, easily repeatable, and useful tool for disease detection and documentation of degrees of breathiness.

Hillenbrand (1988) used synthetically generated voice samples in a breathiness perception experiment. Listening samples were generated using two different techniques for varying the slope of the output spectrum. The first technique altered the synthetic glottal source signal. Low pass filtering was used to produce three variations in the shape of the glottal pulse train which in turn produced three different output spectra. The second technique controlled formant amplitudes to generate three different slopes in the periodic spectrum. The source signals were mixed with 13 different amplitudes of noise. Vowel formant frequencies were set appropriately for

[a]. Ten speech pathologists rated the various [a] samples that had been generated with various combinations of spectral slope and degrees of added noise. As was anticipated, decreases in signal-to-noise ratio were closely associated with a greater perception of breathiness. However, spectral slope had no effect on the listeners' ratings.[1]

Klatt and Klatt (1990) investigated male/female differences in voice quality through acoustic analysis and subjective ratings of recordings made from ten female and six male talkers. The speech materials included two five-syllable sentences, "Steve eats candy cane" and "The debate hurt Bob" and two reiterant imitations of those sentences. The imitations used [?V] and [hV] to replace the five syllables while maintaining the same stress patterns. Five different vowels were used but only data taken from [a] was reported. Reiterant imitations of "Steve eats candy cane" from all 16 subjects were randomized and played to eight listeners. The listeners were asked to rate the sentences on a seven-point breathiness scale. In general, female talkers were judged to be breathier than males, and [ha] was perceived as breathier than [?a]. The authors correlated acoustic measures with subjective breathiness ratings. Of the acoustic measures used, only two were found to have statistical significance: $H_1/H_2$ ratio and the degree of random noise visually present in $F_3$ extractions (Klatt & Klatt, 1990).

---

[1]This contradicted an earlier study by Yanagihara (1967) which investigated the "hoarseness" percept, even when four of Hillenbrand's subjects were asked to give ratings based on perceived hoarseness.

Klatt and Klatt (1990) described a new speech synthesis model (KLSYN88) which was designed to account for a number of naturally occurring acoustic parameters, such as jitter, diplophonia, aspiration noise, source-tract interactions, $F_1$-$f_0$ interactions, changes in the source waveform due to $F_1$ standing waves, and extra poles and zeros due to tracheal coupling. It was felt that this more interactive source-tract design could more accurately simulate the human voice. Control over aspiration noise would enhance the synthesis of female voices which appear to be breathier than male voices. This new synthesis model was tested in a second listening study. Twelve synthetic vowels (one reference-standard and eleven modifications of the reference) were used to create imitations of "Steve eats candy cane" with [ha] and [?a]. The sentences were modeled after one particular female subject's voicing patterns which included greater aspiration for unstressed and sentence final syllables. The acoustic manipulations of the reference-standard vowels included: (a) two increases in $H_1$ amplitude; (b) a 10 Hz decrease in $f_0$ over the first 100 ms of the vowel; (c) increases in $F_1$ and $F_2$ bandwidth; (d) two increases in the downward spectral tilt; (e) two increases in aspiration noise; and (f) three different combinations of increases in spectral tilt, aspiration noise, $F_1$ bandwidth, and open quotient. Listeners were asked to judge each of the sentences in terms of perceived naturalness, breathiness, and nasality.

Increases in spectral noise were found to be the single most important cue to perceived breathiness. However, a combination of added noise, increased spectral tilt, and an increase in $H_1$ was judged most positively, both in terms of breathiness and naturalness. With increases in spectral tilt, less additive noise was required to elicit judgements of breathiness.[2] Increases in $H_1$ alone were heard as breathy by some listeners and nasal by many others. It was speculated that the frequency of the $H_1$ in this case (about 200 Hz) corresponds to the lowest pole that occurs during nasopharyngeal coupling. Signals were never judged as nasal when noise was added to increases in $H_1$. When only $F_1$ bandwidth was increased, the result was perceived as unnatural but not breathy. Downward spectrum tilting was only slightly suggestive of breathiness and was thought to be unnatural when it occurred in isolation. Fundamental frequency decrease elicited no perceptions of breathiness.

Hillenbrand, Metz, Colton, and Whitehead (1990) examined the nature of breathy voice by using high-speed film and acoustic analysis of normal and simulated breathy phonation. High-speed film and simultaneous digital recordings were made of a normal female subject producing normal, breathy, and very breathy [a]. Glottal area analyses for the three voicing conditions illustrated the presence of a glottal chink during the closed phase of the breathy phonation cycle. They also revealed less abrupt opening and closing phases for the breathy conditions and variations in maximum glottal area across conditions. To analyze the recordings acoustically, Hillenbrand et al. (1990) used six measures: (1) "H1A" - $H_1$ amplitude relative to $H_2$ amplitude; (2) "BRI" - a modification of the Fukazawa et al. (1988) Br index; (3) "F3A" - energy of a 1 kHz band centered around $F_3$ relative to overall energy at lower frequencies;

---

[2]This demonstrates that spectral slope does affect the perception of dysphonia. This tends to support Yanagihara (1967) and contradicts Hillenbrand (1988).

(4) "ZCR" - the number of zero crossings counted within a 1 kHz band centered around $F_3$; (5); "CPP" - the prominence of a cepstral peak corresponding to $f_0$ measured from a 1 kHz band centered around $F_3$; and (6) "APK" - the height of the auto-correlation peak corresponding to $f_0$ relative to the overall energy measured from a 1 kHz band centered around $F_3$. Both BRI and F3A measured the relative amounts of high frequency energy. ZCR, CPP, and APK were intended to estimate signal periodicity in the mid frequency range where it was anticipated that breathy phonation would be less periodic than normal phonation. ZCR and APK measured periodicity in the time domain, while CPP made estimates in the frequency domain.

The autocorrelation "peakedness" measure (APK) was found to correlate most closely with subjective estimations of breathiness, followed by the relative energy of a 1 kHz band centered around $F_3$ (F3A) and the Fukazawa et al. (1988) Br index (BRI). The authors concluded that BRI and F3A may prove to be better predictors of breathiness as APK autocorrelation measures are more prone to measurment error. Contrary to the findings of Fisher-Jørgenson (1967) but consistent with those of Klatt and Klatt (1990), Hillenbrand et al. (1990) found that $H_1/H_2$ measures (H1A) were the least likely to correlate with perceived breathness when compared with the other five acoustic measures used in their research.

# METHODS

## Voice Sample Collection

### Talkers

Fifteen normal talkers (7 women and 8 men) were recruited from among graduate students in Speech-Language Pathology and Audiology at Western Michigan University, their friends and relatives. They were selected from a relatively homogeneous population; e.g., native Midwestern American English (Ladefoged, 1975) speakers ranging in age from 22 to 37 with a mean age of 27.3 (*SD*=5.1). Subjects reported no history of voice or hearing problems. Each subject demonstrated ability to perform the required task during subject training.[3]

### Talker Training

A sample recording of a normal, moderately breathy, and very breathy [u] was used to familiarize subjects with the voicing task. Each talker was trained to produce three voicing variations (normal, moderately breathy, and very breathy) at the following three pitch levels: (1) estimated average speaking fundamental (avg. $f_0$), (2)

[3]Nine additional talkers attempted, but were unable to perform, the required task and were not used as subjects in this study.

avg. $f_0$ + 4 semitones,and (3) avg. $f_0$ + 8 semitones.[4] Each talker's long term average speaking fundamental was estimated by asking each subject to read a portion of the "Rainbow Passage" (Fairbanks, 1940). This reading was recorded and frequency averaged via a Kay Elemetrics Visi-Pitch/IBM PC Interface (Horii, 1983). The $f_0$ frequency averaging and talker training took place just prior to the recording session.

## Recordings

Recordings were made in a sound-treated-chamber (IAC 401A), using an Audio-technica 250XL unidirectional microphone at a distance of approximately 7 to 10 cm in front of the lips and 3 cm above the breath stream. The original signals were digitized at 44,100 16-bit samples/sec using a Sony digital audio processor (PCM-F1) and stored on video cassette tape using a Canon portable video recorder (VR-40).

A Yamaha PSR-6 portable electronic keyboard channeled to a loudspeaker inside the recording booth was used to provide talkers with target pitches. A Kay Elemetrics Visi-Pitch (6087) provided subjects with visual feedback regarding fundamental frequency and duration during the recording of vowel tokens. Talkers wore a bone vibrator (Radio Ear B-70-A) adjacent to the larynx. The bone vibrator acted as a transducer coupling the voice signal to the Visi-Pitch via a custom designed voice-activated trigger.

---

[4]Semitones were based on the tempered scale (Hirano, 1981; Rigden, 1977).

The speech materials recorded included the four vowels [a], [æ], [i], and [o]. The two running speech samples, "Joe took father's shoe bench out. He left it lying on the lawn." and "Where were you when we were away?" were also recorded. A randomized list of these vowels and sentences was provided for each talker. Talkers were asked to sustain each vowel token for approximately 3 seconds. For all talkers the speaking tasks were repeated at least once during a recording session that lasted between 45 and 90 min., including a 10 - 15 min. break. Five talkers were asked to return to re-record particular vowel sets which did not demonstrate sufficient variation in breathiness. For example, when a talker's recorded breathy /i/ (at a given pitch level) sounded the same as either its normal or very breathy counterparts, those three vowel samples were later re-recorded under the same conditions as the originals. A total of 36 vowel samples (4 vowels x 3 pitch levels x 3 voicing variations) and six running speech samples (2 utterances x 3 voicing variations) were selected from each of the 15 talkers' recordings. This yielded a total of 540 vowel samples and 90 running speech samples.

The recorded signals, stored on video tape, were sent through a 30 dB attenuator pad, an ATI model M1000-2 precision amplifier, and an Alison Laboratories AL-2ABR Variable Filter (high cutoff setting 7.2 kHz), digitized at 20,000 12-bit samples/sec and stored in the disk memory of a Digital Equipment Corporation (DEC) PDP-11/73 computer. Each vowel token was digitally edited to the most stable (in terms of signal amplitude) 1 sec sample.[5] This editing process occasionally yielded signals that, when auditioned by the experimenter, were deemed unsatisfactory in terms of pitch stability or other voicing qualities. In these instances a more satisfactory 1 sec was selected from the original 3 sec recordings. The 540 1-sec edited vowel samples and 90 running speech samples were stored on computer disk. Only the 180 vowel tokens recorded at the talkers' estimated average fundamental frequency were used for the following acoustic analysis and listening experiments. To prevent onset and offset transients, signals were ramped on and off using a 20 msec cosine function.

## Breathiness Perception Experiment

### Listeners

Twenty listeners (19 women, 1 man) were recruited from among graduate students and faculty in Speech-Language Pathology at Western Michigan University. They ranged in age from 21 to 47 with a mean age of 27.5 (SD = 6.8). Listeners were audiometrically screened at 20 dB HL for five frequencies (0.5, 1, 2, 4, 6 kHz). All listeners passed the screening.

---

[5]This was accomplished by computing non-pitch-syncronous amplitude perurbation. "Shimmer" was averaged across 100 consecutive 10 msec intervals.

Instrumentation

Listening experiments took place in a sound-treated room (ambient noise level < 30 dBA). The stimuli consisted of the 180 vowel tokens used in the acoustic analyses experiments described above. Stimuli were stored in the disk memory of a DEC PDP 11/73 computer. Signals were sent through a Shure Model M67 Preamplifier, an Alison Laboratories AL-2ABR Variable Filter (low-pass frequency setting 7.2 kHz), a MacIntosh-250 amplifier, and then presented over a single Boston Acoustics A60 speaker.

Procedures

Listeners were tested individually. Signals were presented in a sound field at approximately 80 dBA at one meter. This approximated common speaking conditions.

Listeners were asked to rate each of the 180 signals used in the listening experiment according to the amount of perceived breathiness. A direct magnitude estimation procedure required the listeners to determine their own numerical rating scales. (No anchor stimulus or modulus was used.) Listeners could hear each signal as often as they wished and take as much time as necessary before entering their judgements on a computer terminal. Each subject was given the following instructions:

> In this experiment you will hear recorded voice signals presented one at a time over a loudspeaker. Your task will be to judge how breathy each signal sounds. Enter a large number for large amounts of breathiness and a small number for small amounts of breathiness. You can use any whole number

> equal to or greater than zero. It makes no difference what range of numbers you use; that is you can scale your judgements from 0 to 10, from 10 to 90, from 100 to 10,000, or whatever range of numbers you choose. If you wish to hear a signal again before making your judgement enter '-9' for the trial. You can ask for a signal to be repeated as often as you wish. It is important to emphasize that it is your job to judge how breathy each signal is. You should try, as much as possible, to ignore any other variations in voice signal quality.

Listeners participated in two listening sessions at least 24 hrs apart, each lasting between 30 and 45 minutes. Signals were presented in randomized order within three blocks of all 180 signals each. The order of presentation within each 180-stimulus block was rerandomized for every listening session. The 180 signals were presented to each subject six times (3 times per session) for a total of 120 ratings (20 subjects x 6 ratings) per signal. A practice session of 60 signals began all listening sessions to familiarize subjects with the task and allow them to determine their rating scales. The practice trials were identical to the listening session except that ratings from these trials were disregarded.

## Acoustic Analysis

A number of acoustic analyses were performed with each of the 180 signals used in this study. This work was accomplished using an Advance Logic Research (ALR) 486 computer with custom software (Hillenbrand et al., 1990). The measures chosen were based on the following four assumptions regarding breathy voice: (1) an increased glottal open quotient, resulting in a more sinusoidal glottal signal, will enhance $H_1$ amplitude (Bickley, 1982; Ladefoged, 1983; Klatt & Klatt, 1990); (2)

increased glottal turbulence will result in decreased periodicity for both time and frequency domain signals, especially at higher frequencies (Klatt & Klatt, 1990; Ladefoged, 1983); (3) lack of complete closure during the glottal cycle will result in a reduction of higher frequency periodic components in both time and frequency domain signals (Rothenberg 1983); and (4) increased glottal turbulence can be measured as increased amplitude in the high frequency output spectrum (decreased spectral tilt) (Fukazawa, et al., 1988).

The acoustic measures are described as follows:

Measures of Spectral Slope

**BRI**: This is a slightly modified version of the "breathiness index" described by Fukazawa et al. (1988).[6] It is an overall measure of spectral slope. Increases in glottal turbulence (additive noise) are expected to be reflected by decreases in the slope of the spectrum envelope (i.e., increased high frequency energy).

**F-Sum**: This is a measure of average spectral energy above 4 kHz divided by the average spectral energy below 4 kHz.[7] Increases in glottal turbulence are expected to be measurable as increases in high frequency energy.

---

[6]While Fukazawa et al. measured 25.6 ms from any given sample, **BRI** averages a series of 25.6 ms samples, each shifted by 12.8 ms, for the entire 1 sec signal.

[7]Other variations of **F-Sum** were tested, including the following: band-limited (4-7 kHz) energy divided by total energy, band-limited energy divided by energy below 4 kHz, and energy above 4 kHz divided by total energy. These were less successful than **F-sum** at predicting perceived breathiness.

Measures of Signal Periodicity

**APR-BP**: This is a measure of the height of the autocorrelation peak corresponding to $t_0$ relative to a regression line fit to the autocorrelation function, divided by the overall RMS energy in the band-limited signal.[8] Signals were bandpass filtered between 2.5 and 3.5 kHz.

The autocorrelation function is "equal to the average product of the signal...and a time-shifted version of itself, and is a function of the imposed time-shift" (Lynn, 1973). The function reflects the degree of periodicity in the time domain signal. A regression line was used in order to normalize differences in signal amplitude. Relative decreases in the $t_0$ peak are expected to be associated with increases in glottal turbulence. Figure 1 shows a sample autocorrelation function. The tallest peak is associated with $t_0$. The second tallest peak is associated with $2 \times t_0$. The correlation is in arbitrary units.

**APR-HP**: This is a measure of the height of the autocorrelation peak relative to a regression line fit to the autocorrelation function, divided by the overall RMS energy in the band-limited signal. Signals were highpass filtered at 2.5 kHz.

**APKR-BP**: This is a measure of peak-to-trough height of the autocorrelation peak divided by the overall RMS energy in the band-limited signal. Signals were bandpass filtered between 2.5 and 3.5 kHz.

---

[8]Versions of the autocorrelation peak measure were tested which were not divided by overall band-limited signal energy. These were much less successful at predicting perceived breathiness.

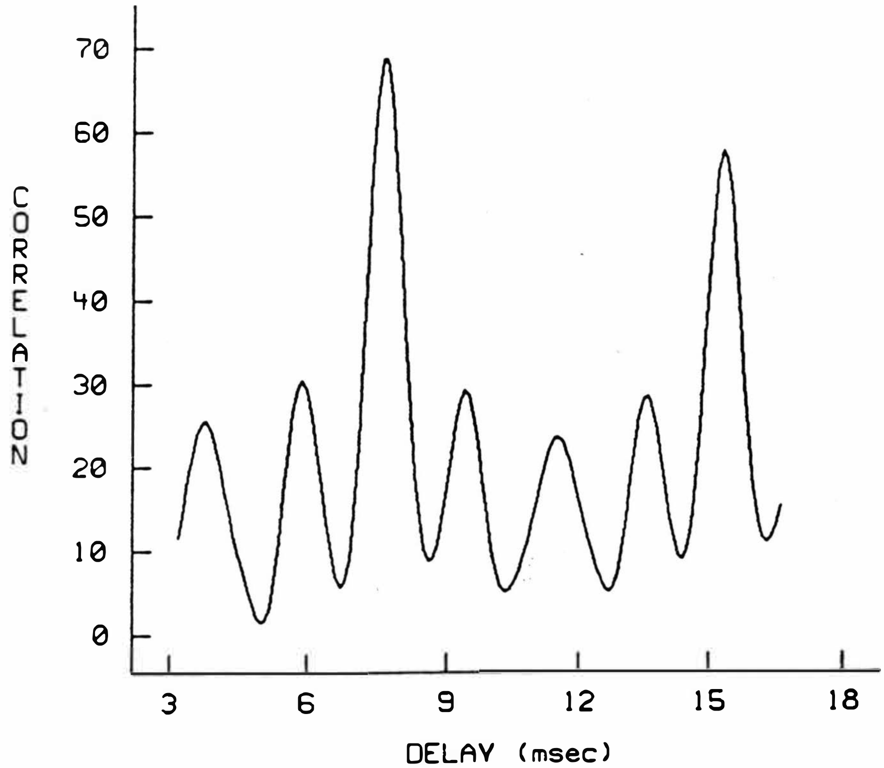


Figure 1. Sample Autocorrelation Function.

**APKR-HP**: This is a measure of peak-to-trough height of the autocorrelation peak divided by the overall RMS energy in the band-limited signal. Signals were highpass filtered at 2.5 kHz.

**CPP-BP**: This is a measure of cepstral peak amplitude normalized for overall cepstral amplitude. Overall amplitude was estimated based on a regression line fit to the cepstrum. Signals were bandpass filtered at 2.5 to 3.5 kHz.

The cepstrum or "power cepstrum" (Randall, 1973) was originally defined as "the power spectrum of the logarithm of the power spectrum" (Bogert, Healy, &

Tukey, 1963). It is a measure of "harmonic periodicity" in the frequency domain. The cepstral peak corresponds to $f_0$ in the power spectrum. It is anticipated that increases in glottal turbulence will be measurable as decreases in cepstral peak prominence (see Figure 2).

**CPP-HP**: This is a measure of cepstral peak amplitude normalized for overall cepstral amplitude. Signals were highpass filtered at 2.5 kHz.[9]

**ZCR-BP**: This is the average zero-crossing count for ten msec segments of the time domain signal. Signals were bandpass filtered between 2.5 and 3.5 kHz. A high zero-crossing count should be associated with low periodicity (i.e., increased noise).

**ZCR-HP**: This is the average zero-crossing count for ten msec segments of the time domain signal. Signals were highpass filtered at 2.5 kHz.[10]

First and Second Harmonic Ratio

**H1**: This is a measure of the dB amplitude of $H_1$ relative to $H_2$. Harmonic amplitudes, based on fast Fourier transforms (FFT), were measured from the approximate center of the vowel sample. Increases in H1 amplitude are anticipated for increases in open quotient (see Figure 3).

---

[9]Versions of **CPP-BP** and **CPP-HP** were tested in which pitch tracking was limited to + and - 25% to reduce error. This did not improve prediction of perceived breathiness.

[10]Zero-crossing counts in the unfiltered signal were found to be less predictive of perceived breathiness than was the band-limited versions.

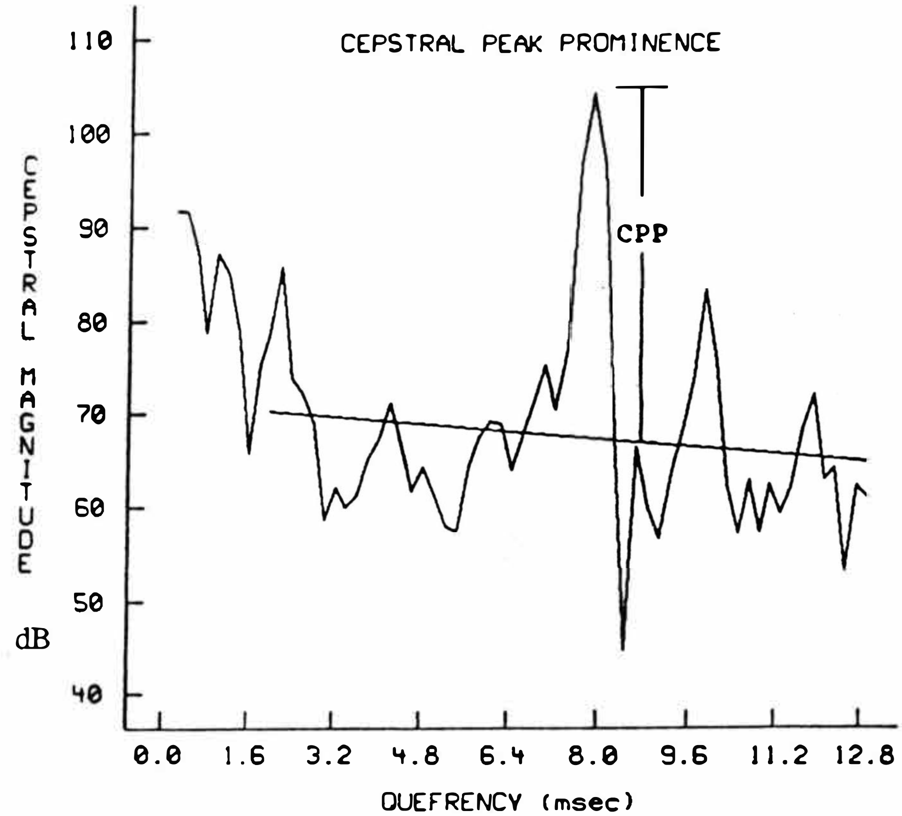


Figure 2. Sample Cepstrum Showing Regression Line Used to Measure Cepstral Peak Prominence (CPP).

**H1** was the only one of the eleven acoustic measures that was not fully automated. It required that the experimenter read the digital Fourier analysis output files, select the first and second harmonic peak amplitudes, and then calculate their ratio.

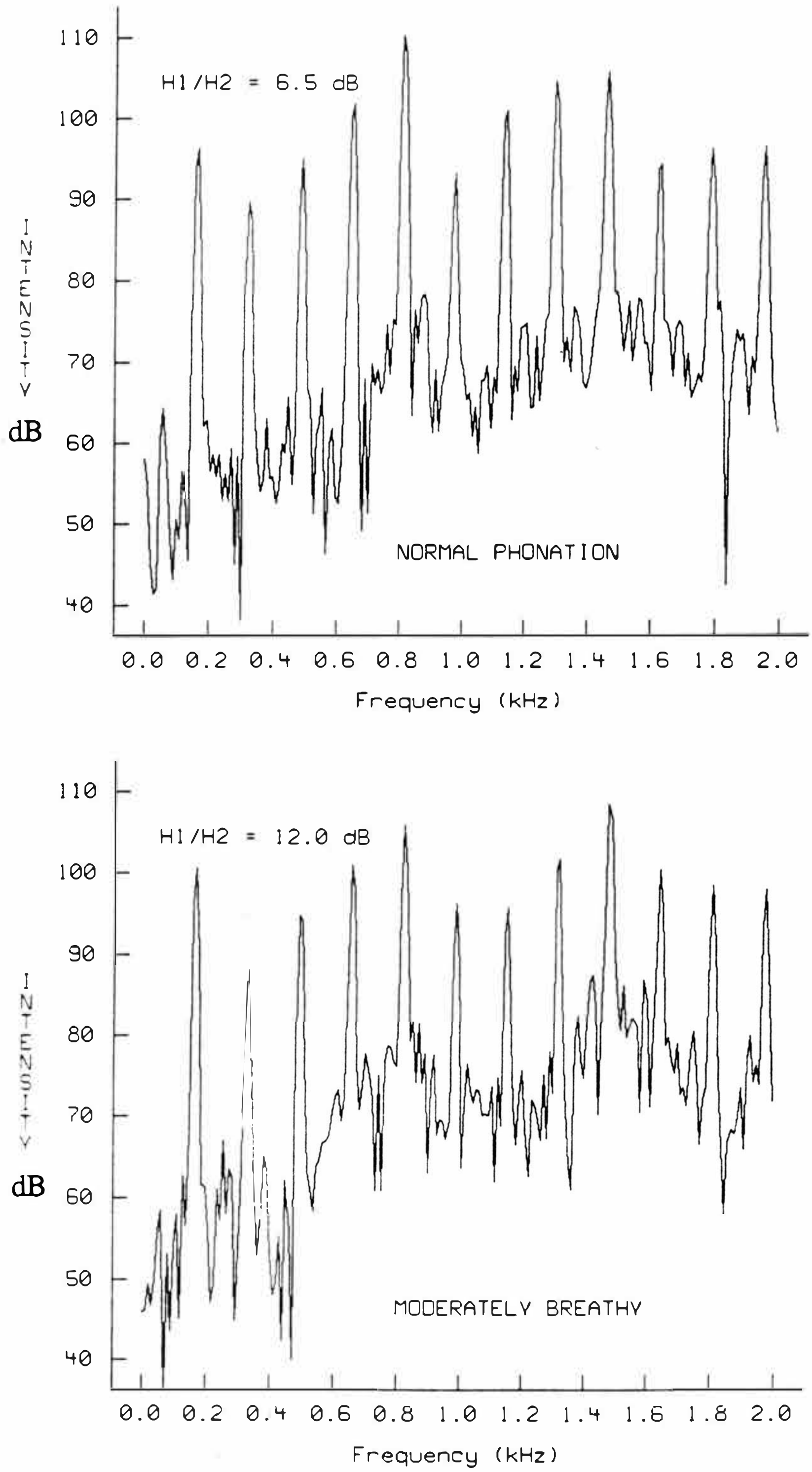


Figure 3. Sample Spectra Showing $H_1/H_2$ Ratios for Normal and Moderately Breathy Phonation.

# RESULTS

## Listener Reliability

Each listening subject participated in two listening sessions. Each listener rated all 180 vowel tokens three times during the first session and three times during the second. The means of the three ratings per token for the first session were compared with those of the second session for each subject. Pearson product moment correlations and a paired *t*-tests were used in the comparison. The results are summarized in Table 1. Pearson within-subject correlation coefficients were high overall and significant at the 0.0001 level. Within-subject *r*-values ranged from 0.813 to 0.966. The mean within-subject *r*-value (based on Fisher's *z*-transformation) was approximately 0.91 ($SD \approx 0.23$).

Paired *t*-test results were found to be at the $p < 0.05$ significance level for 14 of the 20 listeners. Of these 14, ten rated the vowel tokens as being breathier overall during the first listening session than the second. The remaining four rated the second session as being breathier overall. All listeners had been instructed to use the same rating scale for both sessions. However, these *t*-test results reflect differences in mean ratings between listening sessions. For six of the listeners there was no significant difference in rating severity between the first and second listening sessions.

Table 1

Summary of Pearson Correlation Coefficients and
Paired $t$-Tests for Within-subject Ratings

| Listeners | Pearson $r$ $p$=0.0001 | Paired $t$-test $t$ | Listeners | Pearson $r$ $p$=0.0001 | Paired $t$-test $t$ |
|---|---|---|---|---|---|
| 1 | 0.9069 | -2.4616* | 11 | 0.9148 | 0.6827 |
| 2 | 0.9431 | 3.0693* | 12 | 0.9514 | -6.0356* |
| 3 | 0.8521 | 8.2266* | 13 | 0.9191 | -1.9156 |
| 4 | 0.8673 | 11.1368* | 14 | 0.9279 | 12.7054* |
| 5 | 0.8836 | 7.3145* | 15 | 0.9082 | -2.7184* |
| 6 | 0.9493 | 9.2903* | 16 | 0.8627 | 4.0644* |
| 7 | 0.9409 | -1.5713 | 17 | 0.8727 | 5.5282* |
| 8 | 0.8854 | 11.0795* | 18 | 0.9662 | -0.6108 |
| 9 | 0.9377 | 0.9883 | 19 | 0.8135 | -2.4966* |
| 10 | 0.9124 | 6.8109* | 20 | 0.8744 | -1.6753 |

* Significant at the 0.05 level.

A Cronbach Coefficient Alpha was computed between the mean ratings of the 20 listeners as a measure of between-subject reliability. Each listeners mean rating was correlated with the group mean of all the other listeners. Table 2 summarizes the results. The coefficients were all significant at the 0.01 level. It was assumed that an arithmetic mean of all listeners' ratings[11] would be a reasonable representation of perceived breathiness and could used as a dependent variable in a regression analysis.

Table 2

Cronbach Coefficient Alpha
for Between-subject Ratings

| Listeners | Correlation with Set of All Others* | Alpha ($\alpha$) | Listeners | Correlation with Set of All Others* | Alpha ($\alpha$) |
|---|---|---|---|---|---|
| 1 | 0.9497 | 0.9944 | 11 | 0.9231 | 0.9945 |
| 2 | 0.9593 | 0.9943 | 12 | 0.9588 | 0.9943 |
| 3 | 0.9429 | 0.9944 | 13 | 0.9602 | 0.9943 |
| 4 | 0.9598 | 0.9943 | 14 | 0.9613 | 0.9943 |
| 5 | 0.9444 | 0.9944 | 15 | 0.9531 | 0.9943 |
| 6 | 0.9349 | 0.9944 | 16 | 0.9206 | 0.9945 |
| 7 | 0.9594 | 0.9943 | 17 | 0.9543 | 0.9943 |
| 8 | 0.9522 | 0.9943 | 18 | 0.9786 | 0.9942 |
| 9 | 0.9662 | 0.9942 | 19 | 0.8816 | 0.9948 |
| 10 | 0.9566 | 0.9943 | 20 | 0.9500 | 0.9943 |

* $p$=1-$\alpha$ Significant at the 0.01 level.

[11]A geometric mean of the ratings was also computed as suggested by Stevens (1977) but was found to be strongly correlated with the arithmetic mean ($r$=0.9996).

## Effects of Talker Variables on Listeners' Ratings

An analysis of variance (ANOVA) was performed in order to investigate the effects of talker gender, voicing condition, and vowel type ([a], [æ], [i], or [o]) on the listeners' perception of breathiness (see Table 3). The vowel tokens selected from

Table 3

ANOVA Results for Regressions of Talker Gender, Phonation Type, and Vowel Type on Breathiness Ratings

| Source | Mean Square | *F* | *p* |
|---|---|---|---|
| Gender | 6.28 | 4.83 | 0.0295* |
| Vowel | 1.31 | 1.01 | 0.3911 |
| Phonation | 430.76 | 330.94 | 0.0001* |
| Gender/Vowel | 2.39 | 1.84 | 0.1416 |
| Gender/Phonation | 4.62 | 3.55 | 0.3100 |
| Vowel/Phonation | 0.90 | 0.69 | 0.6545 |
| Gender/Vowel/Phonation | 0.52 | 0.41 | 0.8743 |

* Significant at the 0.05 level.

each talker's recordings were chosen to represent three variations along a breathiness continuum (i.e., normal, moderately breathy, and very breathy phonation). As was anticipated, there was a significant effect for phonation type on the ratings of breathiness. There was also a slightly significant effect for talker gender. Vowel type showed no significant effect on the perception of breathiness.

An ANOVA performed within talker gender revealed that females were not perceived as being breathier than males when voicing was normal. This does not support some evidence that English speaking females are perceived as being breathier than their male counterparts (Klatt & Klatt, 1991). Under the very breathy condition, males were perceived as breathier than females. (This finding has little importance within the context of the present study since breathiness was the result of vocal control and not a "normal" or usual speaking condition.) It seems possible that, for whatever reason, males tended to generate more glottal turbulence during their extreme breathy simulations. It might also be the case that listeners allowed for greater breathiness in female talkers, while the breathy male voice was perceived as being more abnormal.

## Effects of Talker Variables on Acoustic Measurements

An ANOVA was performed in order to investigate the effects of talker gender, voicing condition, and vowel on each of the 11 acoustic measurements being reported. There was a significant effect for voicing condition on each of the acoustic measures except for **APR-BP**, which also showed no correlation to listeners' ratings of breathiness (see below). **H1** (a measure of $H_1/H_2$) was found to be higher for females than males under the normal voicing condition. The mean difference between genders was 1.7 dB. This is in agreement with previous findings of gender differences in first harmonic amplitude (Klatt & Klatt, 1990; Monsen & Engebretson, 1977). However, this does not agree with listening experiment results in which no between gender difference was noted.

## Psychoacoustic Correlations

Pearson product moment correlation coefficients were determined between the mean breathiness ratings and each of the 11 acoustic measures. Table 4 lists the

Table 4

Pearson Correlation Coefficients for Breathiness Ratings With Acoustic Measurements

| Acoustic Measure | *r* | *p* | Acoustic Measure | *r* | *p* |
|---|---|---|---|---|---|
| **CPP-BP** | -0.896 | 0.0001 | **F-Sum** | 0.468 | 0.0001 |
| **CPP-HP** | -0.891 | 0.0001 | **ZCR-HP** | 0.463 | 0.0001 |
| **APR-HP** | -0.699 | 0.0001 | **BRI** | 0.410 | 0.0001 |
| **APKR-HP** | -0.675 | 0.0001 | **ZCR-BP** | 0.261 | 0.0004 |
| **H1** | 0.662 | 0.0001 | **APR-BP** | -0.100 | 0.1810 |
| **APKR-BP** | -0.510 | 0.0001 | | | |

results in descending rank order. Figure 4 shows the squared correlations between breathiness ratings and each acoustic measure. Ten of the 11 acoustic measures showed significant correlation with the listeners' ratings of breathiness ($p<0.05$). The

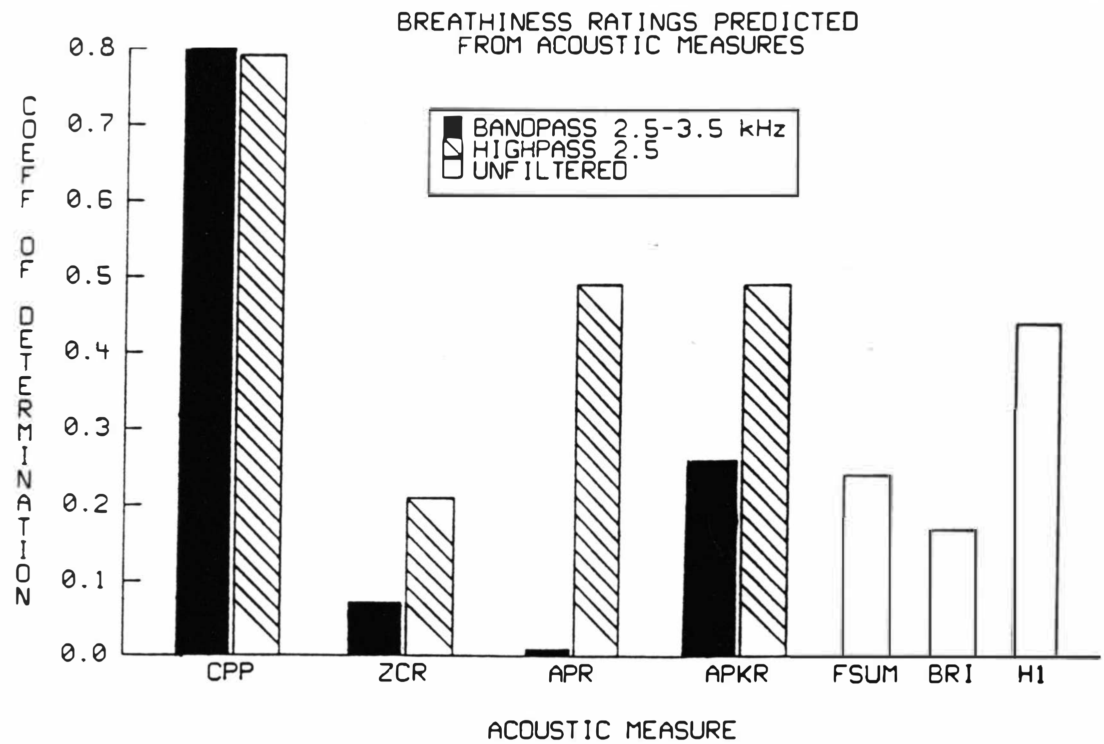


Figure 4. Squared Pearson Correlation Coefficients for Eleven Acoustic Measures With Listeners' Ratings of Breathiness.

two cepstral peak prominence measures (**CPP-BP** & **CPP-HP**) were the strongest predictors of perceived breathiness. The one measure that did not correlate with the breathiness ratings ($p$>0.05) was a variation of the autocorrelation measure (**APR-BP**).

Pearson correlations were also measured between the 11 acoustic analyses and the listeners' ratings based on the two separate conditions of talker gender and vowel type. When correlations were made between the acoustic measures and listeners' ratings for the female voice samples separate from the males and vice versa, prediction of breathiness was not improved. This was also the case when the four vowel types ([a], [æ], [i], or [o]) were separated and the psychoacoustic correlations were made for each vowel independent of the other three.

A stepwise multiple regression was performed in order to determine if some linear combination of the acoustic measures would improve the prediction of breathiness ratings. Table 5 summarizes the results. A variable was entered and retained in the stepwise regression procedure only if it contributed significantly to the variance. Only the six variables listed in order of relative contribution in Table 5 met the 0.15 significance level when all 11 variables were tested within the model. CPP-HP, which was previously found to correlate strongly with the perceptual ratings did not contribute significantly to the prediction of breathiness due to its strong intercorrelation with **CPP-BP**. The first variable retained (**CPP-BP**) was a strong predictor of breathiness on its own and accounted for just over 80% of the variance. The next two variables (**F-Sum** & **H1**) improved prediction most abruptly (9% of variance explained). After the first three variables in the model, improvement in prediction increases relatively slowly. Very little improvement occurred with the addition of the lower three steps in the regression model.

Table 5

Stepwise Multiple Regression Analysis Showing Prediction of Breathiness Ratings With Acoustic Measures*

| Step | Acoustic Measure | % Variance Explained (Total $r^2$) | Partial $r^2$ |
|---|---|---|---|
| 1 | **CPP-BP** | 0.8029 | 0.8029 |
| 2 | **F-Sum** | 0.8566 | 0.0537 |
| 3 | **H1** | 0.8913 | 0.0347 |
| 4 | **APKR-HP** | 0.9118 | 0.0205 |
| 5 | **APR-BP** | 0.9171 | 0.0053 |
| 6 | **BRI** | 0.9201 | 0.0030 |

* Variables were entered and removed based on a $p<0.15$ significance level.

# DISCUSSION

## Listening Experiment

Between-subject agreement in the listening experiment was considered to be very high. This conclusion was based on high Pearson correlation coefficients and Cronbach coefficient alphas. Within-subject agreement was also considered to be high. However, paired *t*-tests indicated some difference between the two sets of listening session results for 14 of the 20 listeners. Four listeners rated the signals as being breathier in the first session than in the second. The ratings of ten other listeners showed a difference in the opposite direction. At the same time, within-subject Pearson correlation coefficients averaged 0.91 with no correlation below 0.81 ($p$=0.0001). It seems likely that the 14 listeners in question did not use the very same rating scales for both listening sessions, yet the relative ratings for each signal remained consistent. Perhaps a longer practice period would have helped listeners develop a consistent rating scale. A third listening session could have been used, allowing elimination of results from the least consistent session. Another possible solution to intersubject inconsistency may be the use of an anchor stimulus during magnitude estimation procedures. In such a procedure subjective ratings are based on a comparison between a given anchor stimulus and the stimuli being rated.

A variety of rating systems have been used in the past for the subjective evaluation of disphonia. These include paired-comparisons (Coleman, 1969), magnitude estimation with or without an anchor stimulus (Coleman, 1969; Hillenbrand, 1988), a four point scale without a reference stimulus (Fukazawa et al., 1988), a five point scale without a reference (Fritzell, et al., 1986), and a seven-point scale with or without a reference (Klatt & Klatt, 1990; Klich 1982; Ptacek & Sander, 1963).

In his discussion of rating scales, Stevens (1975) distinguished between "metathetic" and "prothetic" continua. Metathetic continua, such as pitch, vary positionally from high to low. For example, as frequency decreases the subjective sensation of pitch changes from high to low. Prothetic continua, such as loudness, vary in degree or magnitude rather than position. Stevens demonstrated that when prothetic stimuli are rated using equal interval scales, the lower end of the continuum is subdivided more finely than the upper (Stevens, 1974). He indicated that direct magnitude estimates are more appropriate than equal interval scaling for the judgement of prothetic stimuli. In the current study, listeners were asked to rate vowel tokens according to the perceived amount of breathiness. Listeners were not asked to distinguish from among subjective qualities (e.g., breathy, rough, hoarse, harsh, clear, etc.) Although it has not been born out experimentally, it seems likely that degree of breathiness is better defined along prothetic continua. Therefore, a direct magnitude estimation technique was chosen for this study.[12]

---

[12]A pilot study, using only three listeners, suggested that the type of scale used made little difference.

ANOVA results for the listening experiment indicated that the normal vowel tokens spoken by females were not perceived to be breathier than those spoken by males. Klatt and Klatt found that females were judged to be breathier than males. However, those judgements were based on sentence length utterances. The listeners in the current study made judgements based on 1 sec normal vowel samples interspersed with moderately breathy and very breathy 1 sec samples. One second may not have been time enough to judge breathiness differences across the normal samples. In general, any difference in perceived breathiness between male and female normal talkers may be a relatively small one. Given the wide variations in simulated breathiness used here, listeners may not have been attuned to such relatively fine distinctions.

Contrary to listeners' ratings, first harmonic amplitude measures were found to be higher for females than for males, implying a greater open quotient for females. This is in agreement with earlier findings (Holmberg, Hillman, & Perkell, 1988; Klatt & Klatt, 1990; Monsen & Engebretson, 1977). This may also support the notion that $H_1$ amplitude is not the most important acoustic feature in the perception of breathiness. No other acoustic measurement revealed a difference between male and female normal vowels.

## Psychoacoustic Correlations

Based on correlational data, ten of the eleven acoustic measures reported were successful in the sense that they correlated with the perception of breathiness. Some were more successful than others. The data indicates that signal aperiodicity is a stronger acoustic correlate of perceived breathiness than is amplitude of the first harmonic. This generally supports findings of Hillenbrand et al. (1990), Klatt and Klatt (1990), but contradicts conclusions drawn by Bickley (1982) and Fisher-Jørgensen (1967).[13] Cepstral peak prominence measures (**CPP-BP** and **CPP-HP**) clearly proved to be the best predictors of perceived breathiness within the context of the present study. Their $r^2$ (coefficient of determination) values for prediction were approximately 0.80 and 0.79 respectively. In a study using only one talker, three listeners, and six acoustic measurements (Hillenbrand et al., 1990), **CPP-BP** was less successful at predicting breathiness ($r^2 \approx 0.65$).

**APR-HP** and **APKR-HP** were the third and fourth best predictors, respectively. Hillenbrand et al. (1990), found that the measurement represented here as **APKR-BP** was the strongest predictor of perceived breathiness ($r^2 \approx 0.78$).[14] In the current study, when correlated with listeners' ratings of breathiness, **APKR-BP** had an $r^2$ value of only approximately 0.26. Hillenbrand et al. have suggested that due to their dependence on locating the $f_0$ peak, autocorrelation measures are subject to greater measurement error, especially for breathy signals. **APR-BP** showed no

[13] It should be noted that Klatt and Klatt (1990) were able to manipulate the acoustic parameters of synthetic vowel samples to a greater degree than were Fisher Jørgensen or Bickley who lacked control over source spectrum tilt.

[14] It should be noted that Hillenbrand et al. (1990) were principally measuring glottal area of three breathiness conditions based on high speed film. The psychoacoustic correlations were intended to be preliminary.

correlation with the listeners ratings. In ANOVA results **APR-BP** was also the only acoustic analysis that did not measure a difference between normal, breathy, and very breathy voicing types. The variation in results between the four autocorrelation measures, both within the current study and in comparison to Hillenbrand et al., may reinforce concern over measurement error. As a result they seem less promising than some of the other measures for predicting breathiness.

**APR**, **APKR**, and **ZCR** were all measures of signal periodicity in the time domain. Each was tested with both bandpass and highpass filtered signals. Highpass filtering showed marked improvement in breathiness prediction for each case. It seems likely that turbulent noise is effecting the output spectrum first and most strongly in the frequency range above 3.5 kHz. This would account for the improved measurement of aperiodicity in the highpassed signals. **ZCR-BP** appeared to be a more promising predictor ($r^2 \approx 0.51$) in the smaller study by Hillenbrand et al. than in this study ($r^2 \approx 0.07$).

Hillenbrand et al. reported $r^2 \approx 0.35$ for **H1** correlations compared to $r^2 \approx 0.44$, a slight improvement in the current research. A number of researchers have observed increased $H_1$ amplitude in breathy samples (Bickley, 1982; Fischer-Jørgensen, 1967; Huffman, 1987; Klatt & Klatt, 1990; Ladefoged, 1983). The current correlation of first harmonic amplitude with degree of perceived breathiness is in agreement with Klatt and Klatt (1990), and Bickley (1982). On the otherhand, while Fischer-Jørgensen (1967) described $H_1$ amplitude as an important feature, she found that reducing $H_1$ amplitude did not affect correct identification of breathy vowels.

**F-Sum** and **BRI**, both measured spectral slope. **BRI** demonstrated greater predictive power in an earlier study ($r^2 \approx 0.66$) (Hillenbrand et al., 1990) than it did currently ($r^2 \approx 0.17$). **F-Sum** was only slightly better at predicting breathiness ($r^2 \approx 0.24$) than **BRI**. (**F-Sum** was not used by Hillenbrand et al.) It seems noteworthy that **F-Sum** and **BRI** were less likely to predict breathiness than was the relative height of $H_1$. The relatively low predictive power of **BRI** and **F-Sum** contrasts with a number earlier studies. Fukazawa et al. (1988), Klatt and Klatt (1990), and Klich (1982) all found the amount of high frequency energy to correlate more closely with perceived breathiness.

Multiple regression analysis revealed some improvement in the prediction of perceived breathiness when the acoustic measurements were combined. **CPP-BP**, **F-Sum**, and **H1**, accounted for just over 89% of the variance explained. This concurs with both Klatt and Klatt (1990) and Fischer-Jørgensen (1967) who suggested that the perception of breathiness is most likely dependent upon a combination of acoustic features.

## Summary and Questions for Future Research

Results of the present study show measurements of signal aperiodicity in both the time and frequency domains are more highly correlated with the perception of breathiness than is the relative amplitude of $H_1$. Aspiration noise, source spectrum tilt and $H_1$ amplitude all appear to play a role in the perception of breathiness. What is the relative role of spectral tilt (i.e., reduction in higher frequency harmonics) in predictions of perceived breathiness? No clear answer can be given. Hillenbrand (1988) found that the downward tilt of the source spectrum had no correlation to listeners' ratings of breathiness or hoarseness. Yanagihara (1967) found that spectral tilt did correlate to perceptions of hoarseness. In listening experiments using synthetic vowels, Klatt & Klatt (1990) found that downward tilting of the frequency spectrum alone increased breathiness only slightly. However, by decreasing spectral tilt, less aspiration noise was required to elicit the perception of breathiness. It was the combination of increase in aspiration noise, downward spectral tilt, and $H_1$ amplitude that was judged to be the breathiest and most natural.

If a dependable, objective measure of breathiness is to be developed, there is growing evidence that at a minimum it should measure aperiodicity in the higher frequencies (e.g. above 2 kHz). Better still, a measurement of breathiness should also incorporate first harmonic amplitude. Much needs to be understood before any such diagnostic tool can be developed. The measure of first harmonic amplitude used in the current study was not fully automated and required greater human intervention for its calculation than the other ten measures. Complete automation would need to be achieved before a measure of $H_1$ could be incorporated into a practical diagnostic tool.

More and larger studies are also needed to determine if a breathiness difference exists between male and female speakers of English. If a difference does exist how large a difference is it? Would it need to be accounted for in any assessment of the female English speaking voice?

The present study does not directly investigate the relationship between fundamental frequency and breathy voice but, clearly, there are a number of questions to be answered. What role does $f_0$ play in perception of breathiness? Some have found that higher pitched voice samples are perceived as having increased breathiness (Ptacek & Sander, 1963). Others show no correlation between sample $f_0$ and the perception of breathiness (Klatt & Klatt, 1990). Can the origins of observed pitch differences between murmured and clear vowel cognates (Fisher-Jorgensen, 1967; Ladefoged, 1983; Pandit, 1957) be linked to their historical development? What mechanical role, if any, does glottal vibratory frequency have on glottal flow? Cleveland and Sundberg (cited in Klatt & Klatt, 1990) found evidence that for male singers, increases in $f_0$ are accompanied by slight increases in open quotient.

A persistent glottal chink can facilitate an acoustic coupling of the tracheal cavity with the pharyngial/oral cavity. Very little work has been directed toward the understanding of changes in vocal tract transfer function due to such tracheal coupling. Fisher-Jorgensen (1967) and Dave (1957) reported little if any effect of murmur on vowel formants. On the other hand, Klatt and Klatt (1990) found evidence of increased $F_1$ bandwidth along with extra formants due to tracheal coupling in breathy signals.

A review of the literature revealed that the major body of research dedicated to breathy voice has been based on normal voices simulating breathiness. One logical extension of the current research would be its replication using a pathologically breathy population.

The larger goal of this and related research is greater understanding of both the normal and disphonic voice. Such knowledge can only benefit the fields of laryngology, speech pathology, and voice synthesis. A more specific goal is the development of a diagnostic tool that can objectively and automatically quantify breathiness. Clearly, much more remains to be done.

Appendix A

Human Subjects Institutional Review Board Approval Letter

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-3899

WESTERN MICHIGAN UNIVERSITY

Date: September 17, 1991

To: Ronald Cleveland

From: Mary Anne Bunda, Chair Mary Anne Bunda

Re: HSIRB Project Number 91-08-15

This letter will serve as confirmation that your research protocol, "Acoustic and perceptual correlates of breathy voice" has been approved after expedited review by a subcommittee of the HSIRB. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the approval application.

You must seek reapproval for any change in this design. You must also seek reapproval if the project extends beyond the termination date.

The Board wishes you success in the pursuit of your research goals.

xc: Robert Erickson

Approval Termination: September 17, 1992

## BIBLIOGRAPHY


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, Research Laboratory of Electronics, MIT, Cambridge, MA, 71-82.

Bogert, B. P., Healy, M. J. R., & Tukey, J. W. (1963) The quefrency alanysis of time series for echoes. In Rosenblatt, M. (Ed.), Proceedings of the symposium on time series analysis (pp. 209-243). New York: Wiley.

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

Broad, D. J. (1973). Phonation. In Minifie, F. D., Hixon, T. J., & Williams, F. (Eds.), Normal aspects of speech, hearing, and language (pp. 127-167). Englewood Cliffs, NJ: Prentice Hall.

Chapin-Ringo, C. (1988). Enhanced amplitude of the first harmonic as a correlate of voicelessness in aspirated consonants. Journal of the Acoustical Society of America, Supplement 1, 83, S70.

Cleveland, T., & Sundberg, J. (1983). Acoustic analysis of three male voices of different quality. Speech Trans. Lab. Q. Prog. Stat. Rep., 4, Royal Institute of Technology, Stockholm, 27-38.

Coleman, R. F. (1969). Effect of median frequency levels upon the roughness of jittered stimuli. Journal of Speech and Hearing research, 12, 330-336.

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

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research, 12 246-269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Cluster of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research, 12 642-696.

Dave, R. (1967). A formant analysis of the clear, nasalized, and murmured vowels in Gujarati. Indian Linguistics, 28, 1-30.

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

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

Fisher-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. Indian Linguistics, 28, 71-139.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I., & Sundberg, J. (1986). Breathiness and insufficient vocal fold closure. Journal of Phonetics, 14, 549-553.

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

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

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

Hirano, M. (1981). Clinical examination of voice. New York: Springer-Verlag.

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

Holmberg, E.B., Hillman, R.E. & Perkell, J.S. (1988). Glottal air flow and pressure measurements for soft, normal and loud voice by male and female speakers. Journal of the Acoustical Society of America, 84, 511-529.

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

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

Jensen, J. P. (1965). Adequacy of Terminology for Clinical Judgment of Voice Quality Deviation. Eye, Ear, Nose and Throat Monthly, 44, 77-82.

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.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. (1991). Perceptual evaluation of voice quality: Review, tutorial, and a new model. (Submitted for publication).

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: Contemporaary research and clinical issues (pp. 351-360). San Diego: College Hill.

Ludlow, C. L. (1981). Research needs for the assessment of phonatory function. ASHA Reports, 11, 3-8.

Lynn, P. A. (1973). An introduction to the analysis and processing of signals. New York: John Wiley & Sons.

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

Michel, J. F., & Wendahl, R. (1971). Correlates of voice production. In Travis, L. E. (Ed.), Handbook of speech pathology and audiology. (pp. 465-479). New York: Appleton-Century-Crofts.

Miller, R. L. (1959). Nature of the vocal cord wave. Journal of the Acoustical Society of America, 31, 667-677.

Minifie, F. D. (1973). Speech acoustics. In Minifie, F. D., Hixon, T. J., & Williams, F. (Eds.), Normal aspects of speech, hearing, and language. (pp. 235-341). Englewood Cliffs, NJ: Prentice-Hall.

Monsen, R. B. & Engebretson, A. M. (1977). Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America, 62, 981-993.

Pandit, P. B. (1957). Nasalization, aspiration and murmur in Gujarati. Indian Linguistics, 17, 165-172.

Perkins, W. H. (1971). Vocal function: A behavioral analysis. In Travis, L. E. (Ed.), Handbook of speech pathology and audiology. (pp. 481-503) New York: Appleton-Century-Crofts.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Porkess, R. (1991). The Harper Collins dictionary of statistics. New York: Harper Collins.

Ptacek, P. H., Sander, E. K. (1963). Breathiness and phonation length. Journal of Speech and Hearing Disorders, 28, 267-272.

Rammage, L. A., Peppard, R. C., & Bless, D. M. (1992). Aerodynamic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study. Journal of Voice, 6, 64-78.

Randall, R. B. (1977). Application of B & K equipment to frequency analysis. Nærum, Denmark: Brüel & Kjær.

Rigden, J. S. (1977). Physics and the sound of music. New York: John Wiley & Sons.

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.

Rothenberg, M. (1983). Source-tract acoustic interactions in breathy voice. In Titze, I. R., & Scherer, R. C. (Eds.), Vocal fold physiology: Biomechanics, acoustics and phonatory control (pp. 465-481). Denver: Denver Center for the Performing Arts.

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

Schiavetti, N., Metz, D. E., & Sitler, R. W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired. Journal of Speech and Hearing Research, 24, 441-445.

Shearer, W. M. (1982). Research procedures in speech, language, and hearing.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. Phonetica, 34, 264-279.

Stevens, K. N., & House, A. S. (1954). Development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America, 27, 484-493.

Stevens, K. N., & House, A. S. (1961). An acoustical theory of vowel production and some of its implications. Journal of Speech and Hearing Research, 4, 303-320.

Stevens, S. S. (1974). Perceptual magnitude and its measurement. In E. C. Carterett & M. P. Friedman (Eds.), Handbook of perception (Vol. 2). New York: Academic Press, 1974.

Stevens, S. S. (1975). Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: John Wiley & Sons.

Weiss, C. E., & Lillywhite, H. S. (1976). Communicative disorders: A handbook for prevention and early intervention. Saint Louis: C. V. Mosby.

Yanagihara, N. (1967). Significance of harmonic change and noise components in hoarseness. Journal of Speech and Hearing Research, 10, 531-541.

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