Journal of Speech and Hearing Research, Volume 33, 761–775 December 1990

ACOUSTIC MEASUREMENTS OF MEN'S AND WOMEN'S VOICES: A STUDY OF CONTEXT EFFECTS AND COVARIATION

SUSAN NITTROUER
Boys Town National Institute, Omaha, NE

RICHARD S. McGOWAN
Haskins Laboratories, New Haven, CT

PAUL H. MILENKOVIC
University of Wisconsin, Madison, WI

DONNA BEEHLER
Boys Town National Institute, Omaha, NE

Several acoustic measures of laryngeal activity were made on adult speech to help answer two questions left unresolved by previous work: (1) how each measure varies, if at all, with phonetic structure, and (2) what aspect of laryngeal activity each measure specifies. Speech samples of 15 syllables (three vowels in five prevocalic consonantal contexts) were collected from men and women at two times of the day (early morning and late afternoon). Eight measurements were made, mainly on slices extracted from the middle of the vocalic portions, and inferential and correlational statistics were applied to these measures. Results of the inferential tests indicated differences between men and women in how laryngeal adjustments are made, affecting relative amounts of vocal jitter and spectral tilt of the voicing source. In addition, the voicing and manner characteristics of the prevocalic consonant were found to affect fundamental frequency, cycle-to-cycle perturbations, and amount of aspiration noise. To a lesser extent, vowel height and front/back tongue placement also affected these acoustic source characteristics. Results of the correlational tests showed that different laryngeal mechanisms contributed differentially to signal-to-noise ratios for men and women, and these mechanisms were more greatly affected by fundamental frequency for men's samples. Finally, various acoustic measures of laryngeal noise were found to be related to the same underlying mechanism.

KEY WORDS: laryngeal source, acoustic measurements, normal voice

Several acoustic measures have been used to investigate the vocal source for speech. Although the goal of most studies using these methods is to describe various aspects of laryngeal activity during speech production without the use of invasive techniques (Ladefoged, Maddieson, & Jackson, 1988), the specific objective of each study differs. As a result, the choice of acoustic measure varies among studies. For example, clinical speech scientists are interested in acoustic measures of voice to describe chronic source characteristics that distinguish between normal speakers and those with laryngeal disorders. Common acoustic measures used in clinical applications include fundamental frequency, jitter, and shimmer, with these last two measures quantifying perturbations in the duration and amplitude, respectively, of glottal-cycle periodicity. In addition, measures of the relative contributions to the source spectrum of harmonic and non-harmonic components have been developed (e.g., Milenkovic, 1987; Muta, Baer, Wagatsuma, Muraoko, & Fukuda, 1988; Yumoto, Gould, & Baer, 1982). Under most conditions, the normal larynx creates a voicing source with a stronger harmonic component, relative to the background noise, than does the disordered larynx.

Linguists are interested in quantifying differences in laryngeal activity among vowels exhibiting phonologically distinct source characteristics. These are not chronic states distinguishing among individual speakers, but rather, rapidly changing states that must display similarities across speakers of the same language. One major difference among vowels with different sources of characteristics is the spectral tilt, which is the rate of amplitude decrease as a function of frequency. For example, breathy vowels exhibit steeper spectral tilts than clear vowels (e.g., Bickley, 1982; Fischer-Jorgensen, 1967; Ladefoged, 1983; Ladefoged & Antoñanzas-Barroso, 1985). To quantify this distinction, the amplitude difference is computed between the fundamental (i.e., first harmonic) and higher frequency components of the speech signal, typically either the second harmonic or the first formant. Assuming that both the frequency of the fundamental and the formant structure remain constant, the amplitude of the fundamental is greater relative to the higher frequency components in breathy than in clear vowels, as would be expected given their steeper spectral tilts. In terms of laryngeal activity, these amplitude measures actually describe differences in volume velocity at the glottis: In clear vowels, the laryngeal musculature controls vocal-fold vibration in such a way that the volume-velocity waveform is highly skewed (i.e., there is a relatively slow increase to maximum velocity with a more rapid return to baseline), and there is a distinct baseline phase. In breathy vowels, the increases and decreases in volume velocity are more symmetrical, due to a slower return to baseline than in clear vowels, and the baseline phase is abbreviated.

Scientists concerned with synthesis and automatic recognition of speech have studied the acoustic consequences of natural variations in laryngeal activity to help develop more natural-sounding speech synthesizers and to find algorithms for extracting formant frequencies robust to variations in source characteristics. Measures used in these investigations include the amplitude of the fundamental relative to the higher frequency components (Klatt & Klatt, 1990; Monsen & Engebretson, 1977), as well as fundamental frequency (Shadle, 1985), and measures of the amount of additional noise in the signal, particularly in the higher spectral regions (Klatt & Klatt, 1990).

0022-4685/90/3304-0761$01.00/0

Previous studies investigating acoustic measures of laryngeal activity have succeeded in meeting some goals of each group of investigators: In general, clinical speech scientists have acoustic measures for distinguishing between speakers with normal and disordered larynges, linguists have measures that rather accurately describe vowels differing in source characteristics, and methods have been developed to make synthesized speech more natural sounding. Nevertheless, several other issues remain unresolved. First, we do not know how most acoustic measures of laryngeal activity vary with phonetic structure for normal speakers. Previous work has either not used a variety of phonetic structures or not measured more than one acoustic parameter, usually fundamental frequency. In other words, coarticulation of laryngeal gestures with supralaryngeal gestures has not been adequately studied using acoustic measures. It is important that information regarding laryngeal/supralaryngeal coarticulation be gathered using acoustic measures with normal adult speakers before this phenomenon is investigated with other populations, such as children. Otherwise, neither the acoustic measures appropriate for such investigations nor typical patterns of such coarticulation would be known.

The second issue left unresolved by existing studies concerns which aspect of laryngeal activity is actually quantified by each acoustic measure. Obviously, this issue is related to the first in that it is not enough just to measure the acoustic consequences of laryngeal/supralaryngeal coarticulation; we also wish to describe the coarticulation itself. For this reason, it is necessary to know the laryngeal activity specified by each acoustic measure. The most direct method for obtaining this information is to use physiologic techniques in conjunction with acoustic measurement. However, it is often difficult to get clear acoustic records in conjunction with other types of signals, and such procedures would involve the very invasive techniques many investigators seek to avoid. Less direct methods, such as modelling (e.g., Fant, 1986; Fant, Liljencrants, & Lin, 1985; Titze, 1974) and correlating acoustic measures with perceptual judgments (e.g., Bickley, 1982; Fischer-Jorgensen, 1967), allow inferences of what aspect of laryngeal activity is specified by each measure, but are not conclusive when used by themselves.

A method that would provide additional information concerning the laryngeal activity specified by each acoustic parameter would be to study correlations among these measures. Those measures that were found *not* to covary clearly could be hypothesized to be measuring *different* glottal phenomena. Supportive evidence would also be obtained for suggesting that some acoustic measures that covary are evaluating the same phenomenon, assuming, of course, that there were other bases for making such judgments. For example, measures of signal-to-noise, or harmonic-to-noise, ratios, and amplitude measures of the fundamental relative to the higher frequency components of the signal were each developed to measure breathiness. However, the former have been used mainly in clinical studies, whereas the latter have been used almost exclusively as linguistic descriptors. Consequently, it is not clear that the label "breathy" applies to speech signals with the same acoustic (and therefore, articulatory) characteristics in clinical and linguistic descriptions, or, as expressed by Ladefoged (1983), that "one person's voice disorder is another person's phoneme" (p. 351). On one hand, it is reasonable to expect that glottal sources with more symmetrical volume-velocity waveforms would contribute more noise to the signal because of their greater open quotients (i.e., periods of glottal flow). Therefore, the relative amplitude of the fundamental would be greater, and the signal-to-noise ratio would be poorer. On the other hand, the shape of the volume-velocity waveform and the relative amount of noise created at the larynx (i.e., aspiration noise) may be determined independently, a possibility suggested by looking across studies: Fischer-Jorgensen (1967), Bickley (1982), and Ladefoged and Antoñanzas-Barroso (1985) all reported that the increase in relative amplitude of the fundamental is a more important cue to the perception of phonologically breathy vowels than is the presence of aspiration noise. However, an increase in noise is a major acoustic consequence of many laryngeal disorders (Muta et al., 1988; Yanagihara, 1967; Yumoto et al., 1982). Correlational studies could help determine the relative dependence or independence of these two laryngeal phenomena (the symmetry of the glottal-source cycle and the amount of aspiration noise).

The main goal of the present work was to investigate the effects of variation in phonetic structure on acoustic measures of voice, and thus determine which acoustic measure(s) might be appropriate for examining coarticulation among laryngeal and supralaryngeal gestures. Specifically, inferential tests were designed to evaluate possible effects on these measures of vowel identity and consonantal context. Men and women served as subjects to investigate the effect of speaker sex, and measurements were made for samples collected at two different times of the day to examine possible differences due to this factor. Finally, it was hoped that we might contribute to hypotheses concerning which aspect of laryngeal activity each measure specifies by conducting correlational tests among the measures. These correlational tests also permitted us to investigate possible artifacts in several of these measures.

REVIEW OF EXISTING LITERATURE

Results of studies done for clinical, linguistic, and synthesis/recognition purposes have provided the following information regarding the effects on acoustic voice measures of the independent variables we planned to investigate.

Speaker sex. Differences in fundamental frequency between men's and women's speech are commonly recognized, with some authors reporting women's fundamental frequencies to be as much as 1.7 times those of men (e.g., Klatt & Klatt, 1990; Peterson & Barney, 1952).

However, others have reported female/male differences of as little as 1.45 (e.g., Monsen & Engebretson, 1977). Most studies indicate mean male values of approximately 120–130 Hz (e.g., Horii, 1980, 1982; Monsen & Engebretson, 1977; Peterson & Barney, 1952), with greater variability reported for female values. Across studies, these values range from roughly 190 Hz (e.g., Monsen & Engebretson, 1977; Sorensen & Horii, 1982) to 220 Hz (e.g., Cooper & Sorensen, 1981; Peterson & Barney, 1952).

In addition, the amplitude of the first harmonic relative to that of the second (i.e., the harmonic-amplitude difference) is generally greater in women's voices than it is in men's (Henton & Bladon, 1985; Klatt & Klatt, 1990; Monsen & Engebretson, 1977). This amplitude difference ensures that the tilt of men's and women's source spectra are similar when measured in terms of absolute frequency, given that women's higher fundamental frequencies result in more widely spaced harmonics. Furthermore, this difference between men's and women's voices indicates that women's volume-velocity waveforms are more symmetrical (and consequently have longer open quotients) than men's. In addition, stroboscopic evidence has shown that, for a much larger percentage of women than of men, a posterior portion of the glottis remains open throughout vowel production (Bless, Biever, Campos, Glaze, & Peppard, 1989). Although it is not clear what the exact relation is between longer open quotients and the presence of posterior openings, both phenomena suggest that more noise might be created at the larynx in female than in male speakers.[1] In fact, more noise in the high-frequency regions of women's speech spectra has sometimes been reported (Klatt & Klatt, 1990), suggesting that higher relative amplitudes of the fundamental and greater amounts of aspiration noise might necessarily be related.

Many investigators even take measures of these two acoustic characteristics (i.e., the harmonic-amplitude difference and the amount of aspiration noise) to be interchangeable, basing the decision of which to use on other factors, such as ease of measurement (e.g., Henton & Bladon, 1985). However, the available data regarding the effects of speaker sex on measures of aspiration noise remain equivocal, due to the lack of an appropriate measure: Klatt and Klatt (1990) reported greater aspiration noise for women's samples than for men's, but a subjective scale was used. Yumoto et al. (1982) and Milenkovic (1987) reported similar harmonic-to-noise and total signal-to-noise ratios, respectively, for samples from men and women. These results would demonstrate that women's source spectra are no more noisy than men's, if it can be assumed that the measures accurately quantify the amount of aspiration noise in the speech signal. Although these measures are more objective than the scale used by Klatt and Klatt, it is not clear that they are precise indicators of the amount of aspiration noise per se in the speech signal. These measures are heavily weighted towards the lower frequencies because the glottal vibration source decreases with increasing frequency, probably making them highly susceptible to the effects of jitter and shimmer. Aspiration noise, on the other hand, is more heavily weighted in the higher frequencies, and therefore may not be gauged accurately by these measures. A direct and objective measure of aspiration noise in the higher frequencies, roughly in the third-formant region, is needed to accurately assess the effect of speaker sex on this acoustic parameter, and simultaneously evaluate the relation between the amount of aspiration noise in the source signal and the harmonic-amplitude difference.

Most studies measuring vocal jitter have used only male speakers producing vowels in isolation. Values from these studies give average jitter for normal young men to be between roughly .040 ms and .075 ms (e.g., Horii, 1982; Wilcox & Horii, 1980). The scarce available evidence comparing jitter in male and female voices is equivocal, but generally seems to indicate lower jitter for female speakers. For example, Ludlow, Bassich, Connor, Coulter, and Lee (1987) reported mean jitter values of .047 ms for a group of normal male speakers and of .033 ms for a group of normal female speakers. Milenkovic (1987) reported mean values of .022 ms and .015 ms for men and women, respectively. In contrast, Orlikoff and Baken (1988) found similar absolute jitter values for men and women at their preferred pitch (.042 ms for men and .044 ms for women), but showed that jitter tended to decrease with increasing frequency. Specifically, mean absolute jitter decreased with increases in fundamental frequency obtained by instructing subjects to vary pitch relative to their most comfortable value. However, the slope of this regression (jitter as a function of fundamental frequency) was greater (more negative) for men than for women. This trend of decreasing jitter with increasing fundamental frequency was also reported by Horii (1979) for samples for male speakers.

The available evidence regarding the effect of speaker sex on shimmer is also inconclusive. Ludlow et al. (1987) reported similar shimmer values for men and women (5.1% and 5.3%, respectively), but Milenkovic (1987) reported slightly higher shimmer values for men (1.66% as opposed to 1.18% for women).

Consonantal context. Numerous studies have confirmed higher fundamental frequencies during the acoustic vowel segment following voiceless obstruents than following voiced, in the order of: voiceless stops > voiceless fricatives > voiced stops > voiced fricatives (e.g., House & Fairbanks, 1953; Lehiste & Peterson, 1961; Mohr, 1971; Ohde, 1984; Umeda, 1981). The aspect(s) of laryngeal activity responsible for this consonantal context effect is still not entirely understood. Two separate, though not mutually exclusive, hypotheses sug-

[1]Effects similar to these sex-related differences have been reported in linguistic studies of phonologically clear and breathy vowels: (a) As already discussed, the harmonic-amplitude differences are higher in breathy than in clear vowels; (b) A posterior glottal gap has been observed for speakers during the production of breathy vowels, but not during the production of clear vowels (Fischer-Jorgensen, 1967); and (c) Spectrograms of breathy vowels sometimes exhibit greater noise in the high-frequency regions (Ladefoged & Antoñanzas-Barroso, 1985, but cf. Fischer-Jorgensen, 1967).

gest that either increased vocal fold stiffness during the production of voiceless stops, or a lowered larynx during the production of voiced stops, is responsible for the effects, but each hypothesis is accompanied by its own set of problems [see Ohala (1978) for a complete discussion]. Physiologic studies have shown some differences between voiced and voiceless obstruents in patterns of activity of several laryngeal muscles. For example, Löfqvist, McGarr, & Honda (1984) found that activity of the lateral cricoarytenoid and vocalis muscles was suppressed during abduction for voiceless stops. However, activity of these muscles at vowel onset was the greater following voiceless than following voiced stops, with this difference continuing into the vowel segment. Löfqvist, Baer, McGarr, & Story (1989) observed an increase in the level of cricothyroid activity during production of voiceless obstruents, but not during voiced, with higher levels of activity also apparent during vowels following the voiceless consonants. Finally, Löfqvist and Yoshioka (1981) noted differences in the amplitude and duration of abduction of devoicing as a function of consonant manner. All these observed differences in muscle activity would result in differences in vocal fold tension that could explain the consonantal context effect. However, there are many other aspects of laryngeal activity for consonant voicing that are not yet fully understood. More study is needed to clarify these mechanisms and to understand the contribution of each to this effect.

Consonantal context effects on acoustic parameters, other than fundamental frequency, have not been studied, but inverse filtering has demonstrated several differences in the vicinity of voiced and voiceless obstruents. For example, Gobl's (1988) work showed consonantal context effects on excitation strength that lasted well into the vowel. Also, Löfqvist and McGowan (1989) found that measures of peak flow, minimum flow, and open quotient were all greater during the initial portions of vowel segments following voiceless obstruents than following voiced. Furthermore, one of the two speakers in that study demonstrated these context effects for the open quotient through the first 20 pitch periods of the acoustic vowel segment. Therefore, some consonantal context effects on acoustic measures of laryngeal activity other than fundamental frequency (such as signal-to-noise ratios and spectral tilt) might be expected.

Vowel identity. Several authors have reported generally higher fundamental frequency during the production of high vowels than during the production of low vowels, the phenomenon known as the "intrinsic pitch" of vowels (Honda, 1983; House & Fairbanks, 1953; Mohr, 1971; Ohala, 1973; Shadle, 1985; Zawadzki & Gilbert, 1989). The physiologic activity responsible for this effect is not clearly understood. Honda (1983) suggested that it is mainly due to the hyoid bone being displaced forward, increasing longitudinal tension along the vocal folds, when the tongue root moves forward to produce high vowels. In contrast, Zawadzki and Gilbert (1989) found fundamental frequency to be most strongly correlated with mandible height. Moreover, both Honda and Shadle found interactions between the vowel effect and other linguistic effects: Honda found that the horizontal position of the hyoid bone was similar for unstressed /i/ and stressed /ɑ/. Shadle found that the magnitude of the vowel effect on fundamental frequency was influenced by sentence position, such that the effect was progressively reduced as position moved toward the end of the sentence.

Two other studies have investigated the effects of vowel identity on acoustic correlates of laryngeal activity, other than fundamental frequency. The results of Milenkovic (1987) indicated differences in the amount of jitter and shimmer, and in signal-to-noise ratios among the vowels /i/, /ɑ/, and /u/, although statistics were not applied to these trends. Generally, /u/ demonstrated the least jitter and shimmer and the highest signal-to-noise ratios. The other two vowels demonstrated similar signal-to-noise ratios, but /ɑ/ samples had higher jitter and shimmer values. Horii's (1980) results for a group of normal male speakers producing the same three vowels are in agreement with Milenkovic's for shimmer, but jitter values were greatest for /i/ in his samples, with values for /ɑ/ and /u/ roughly equal. In contrast to both these studies, Horii (1982) found no differences for samples from normal males on measures of either jitter or shimmer across eight vowels. Furthermore, shimmer was substantially less than in previously reported studies. Horii (1982) suggested that these cross-study differences, especially for shimmer, may be associated with the use in his later study of a throat accelerometer, which seems to capture laryngeal waveforms more accurately than acoustic recordings.

METHOD

Speakers

Eight adults (4 male and 4 female) between the ages of 20 and 40 years served as speakers. All were native speakers of American English, nonsmokers, and had no history of speech, language, or hearing problems.

Stimuli

The stimuli were 15 consonant-vowel (CV) syllables, consisting of one of three vowels (/i/, /ɑ/, or /u/), with one of five consonants (/s/, /ʃ/, /t/, /d/, or /k/) in the prevocalic position. Each syllable was produced in the carrier phrase "I want a _____ please" in chest register at individually preferred pitch and intensity.

Procedure

All speakers were recorded twice, once in the morning between 8:00 and 10:00, and once in the afternoon between 4:00 and 6:00. Each syllable served as a label for a hand-drawn picture. The syllables, produced in the carrier phrase, were spoken in response to these pictures, which were grouped into five sets of three. The pictures

comprising each group were varied among speakers, with the one stipulation being that a syllable containing each vowel be in each group. The three pictures in each group were presented 10 times in randomized order before moving onto the next group of pictures. Peak intensity was monitored on a VU meter by the experimenter, and subjects were encouraged to maintain a constant intensity level. Speech samples were recorded on a Nachamichi MR-2 cassette deck, using an AKG C-535EB condenser microphone with a Shure Model M268 mixer. This system has a flat frequency response to 20,000 Hz, and a signal-to-noise ratio of better than 68 dB.

The first eight tokens of each syllable containing no extraneous noise were digitized on an IBM-PC AT at a sampling rate of 20 kHz, with low-pass filtering below 10 kHz, but without high-frequency pre-emphasis. Twelve-bit resolution was used in digitizing. The vowel nucleus of each token was identified as the periodic portion of the waveform between the initial consonant and the following /p/ in "please." Ten pitch periods from the middle of the nucleus (i.e., 5-pitch periods to either side of the temporal center) were extracted and saved in a separate file for subsequent analysis.[2]

Acoustic Analysis

The duration (dur) of the vowel nucleus was obtained from waveforms of whole tokens. Formant frequencies, bandwidths, and amplitudes were extracted from a 10-ms window at the precise temporal center of the vowel nucleus, using Interactive Library Systems (ILS) software. For the present analysis, the frequency of the first formant (F1) was the only one of these values of interest. All other measurements were made on the 10-pitch periods extracted from the center of the vowel nucleus.

Five measures of the acoustic speech waveform were derived from an autocorrelational technique: fundamental frequency (F_0), jitter, shimmer, signal-to-noise ratio (SNR), and band-limited signal-to-noise ratio (BLSNR). F_0, measured in Hz, is the reciprocal of the measured pitch period. Jitter is the mean cycle-to-cycle change in pitch period, measured in milliseconds. Shimmer is the mean percent change in waveform amplitude among pitch periods. SNR is the decibel ratio of total energy in the acoustic speech signal to the energy in the aperiodic, or noise, component. BLSNR is the SNR measure applied to the frequency range of 2–4 kHz.

Calibrations on synthetic speech have indicated that measure of jitter, shimmer, and SNR are influenced by the choice of vowel (Milenkovic, 1987). Therefore, inverse filtering was performed prior to making measurements to remove the formant structure from the acoustic waveform. A 20-ms interval in the center of each 10-pitch-period segment was extracted, pre-emphasized at 6 dB per octave, and a 22-coefficient LPC model was determined using the covariance method (Markel & Gray, 1976). The 22-coefficient LPC inverse filter was then applied to the acoustic waveform segment without any pre-emphasis. The inverse filter output under these conditions provides an estimate of the first derivative of the glottal airflow waveform (Milenkovic, 1986). The use of a fixed inverse filter over the 10-pitch-periods avoids introducing voice perturbation artifact resulting from changes in the LPC coefficients.

Three of the measures (F_0, jitter, and shimmer) were obtained using a method described by Milenkovic (1987) and implemented in CSpeech, a commercially available software package. The autocorrelation function for lag values about the pitch period is computed by averaging over a pitch period interval. The location of the autocorrelation peak at a lag near the expected pitch period gives the value of the pitch period for use in determining F_0 and jitter. The peak amplitude is used to determine shimmer. Parabolic interpolation is employed to find the precise location of the peak between waveform sample positions. In a calibration performed on synthetic speech sampled at 8.33 kHz (Milenkovic, 1987), the worst case absolute error in pitch period estimated by this method was .008 ms, and the worst case relative error in pitch period, reflected in measured jitter for a zero jitter signal, was .002 ms. (Without interpolation, we would have had a worst case pitch error of .060 ms, half the .120 ms interval between waveform samples.) Because we used a 20-kHz sampling rate in the present study, the worst case absolute error in pitch period (assuming linear scaling) would be under .003 ms, and the worst case relative error would be under .001 ms.

We sought to improve upon the SNR measurement reported by Milenkovic (1987). In that work, SNR was measured by comparing the autocorrelation peak value at the pitch-period lag with the energy in each pitch period of the acoustic waveform. A deficiency in the autocorrelation peak is indicative of the strength of an aperiodic component of the waveform. This technique has the drawback of relying on parabolic interpolation to quantify the small differences in the autocorrelation peak attributable to the noise. A 1-% interpolation error of the peak value corresponds to an SNR floor of 20 dB. If we were to interpolate the acoustic waveform directly, a 1-% RMS interpolation error would correspond to a 40-dB floor on SNR because decibel units are 10 times the base 10 logarithm of magnitude squared (autocorrelation values), whereas they are 20 times the base 10 logarithm of magnitude (direct use of waveform values). Therefore, we performed interpolation on the waveform directly in the present work.

A pitch predictor was used to obtain a measure of aperiodicity noise, again, by direct use of the acoustic

[2]Titze, Horii, & Scherer (1987) suggested that at least 20 pitch periods are needed to obtain a stable estimate of jitter and shimmer, but sustained vowels were used in that study. For this investigation, including 20 pitch periods sometimes would have involved the transitional regions of the syllable, which could actually have increased variability among samples. Furthermore, pilot work with some of the longer samples had shown that obtained measures, as made in this study, did not vary with the addition of pitch periods beyond 10. The use of multiple tokens also helped to ensure reliable estimates.

waveform. A description of the pitch predictor, an elaboration of Atal's (1982) pitch predictor, can be found in Milenkovic, Bless, & Rammage (1989). The interpolation procedure used was that of Crochiere and Rabiner (1983). The low-pass filter used in the interpolation was produced with a Hamming window applied to the impulse response of an ideal low-pass filter (Oppenheim & Shafer, 1975). The first null of the filter occurs at 5.5 kHz, giving a passband from 0 to 3.3 kHz, a transition band from 3.3 kHz to 5.5 kHz, and a stop band of frequencies beyond 5.5 kHz. We selected this passband as being representative of the bulk of the signal energy in vowels. By attenuating that portion of the signal outside this band, we suppressed the noise artifact of the LPC inverse filter, which has a high gain for frequencies where the vowel signal has little or no energy.

Klatt and Klatt (1990) described glottal insufficiency as having the effect of reducing the periodic component of the acoustic signal in the higher frequencies, while increasing in these frequencies the aspiration noise, which results from aerodynamic turbulence. Therefore, the BLSNR measure was formulated for this study to evaluate noise in these higher frequencies, and in so doing, possibly separate the aspiration noise from the other noise components.

A Discrete Fourier Transform was performed on sections from syllables with the vowel /ɑ/ to determine amplitude values of the first two harmonics (H1 and H2), and these values were used to compute the harmonic-amplitude difference (H1-H2). This H1-H2 score was not computed for /i/ and /u/ because, in these vowels, the frequency of the first formant is often near, or even below, the second harmonic, making the procedure unreliable (Ladefoged et al., 1988).

RESULTS

For each measure, a mean of the eight tokens of each syllable, obtained from each speaker at each time of day, was computed and used in subsequent analyses.[3] In this way, reliable estimates of the measures for each speaker were obtained, while not artificially inflating the degrees of freedom. Log transforms of shimmer were used in all statistical tests because of the highly skewed distribution.

Main Effects

Analyses of variance (ANOVAs) were done on each measure, except for F1 and duration of the vowel nucleus. These last two values were used only to investigate the possibility of artifactual relations between them and the other measures. For the ANOVAs of F_0, jitter, shimmer, SNR, and BLSNR, the between-subjects' main effect was speaker sex, and the within-subjects' main effects were time of day, consonantal context, and vowel identity. For the ANOVA of H1-H2, vowel identity was not a main effect because it was not measured in two of the three vowels.

Speaker Sex

Table 1 gives means and standard deviations for each measure across time of day, consonantal context, and vowel identity for male and female speakers. The ANOVAs showed two of these differences to be statistically significant: F_0 [$F(1,6) = 22.94$, $p = .003$] and jitter [$F(1,6) = 12.78$, $p = .01$].[4] The difference in jitter between male and female speakers was not accounted for by differences in pitch-period durations: Ratios of mean jitter to mean pitch-period duration were .015 for men and .007 for women. Unlike ANOVAs for the other effects, the ANOVA for jitter also demonstrated significant interactions of speaker sex with other variables: specifically, with time of day [$F(1,6) = 8.77$, $p = .03$] and with consonantal context [$F(4,24) = 5.80$, $p = .002$]. Because of these interactions, a Simple Effects Analysis was done that examined these other effects for men and women separately (i.e., holding sex constant). This analysis showed these two factors to be significant for men only.

Time of Day

Only jitter demonstrated a significant overall effect of time of day [$F(1,6) = 8.34$, $p = .03$], but the Simple Effects Analysis showed this result to be attributable to a difference in jitter for male speakers only as a function of the time of

[3]Standard deviations for each set of eight tokens were computed. Medians for these 240 values (15 syllables × 2 times of the day × 8 speakers) were as follows: F_0 = 6 Hz; SNR = 1.9 dB; BLSNR = 2.2 dB; shimmer = 1.18%; jitter = .020 ms; H1-H2 (/ɑ/ only) = 1.25 dB.

[4]It may appear from Table 1 that the effect of speaker sex on H1-H2 should be statistically significant also, and this result would be expected from previous research. In the present study, the effect fell just short of statistical significance [$F(1,6) = 5.24$, $p = .06$].

TABLE 1. Means (and standard deviations) of each dependent variable for male and female speakers, across time of day, consonantal context, and vowel identity.

Dependent variable	*Males*	*Females*
F_0 (Hz)	136 (27)	207 (14)
SNR (dB)	22.9 (3.5)	22.6 (2.1)
BLSNR (dB)	7.87 (3.49)	7.87 (2.62)
shimmer (%)	3.74 (1.57)	3.16 (1.91)
jitter (ms)	.110 (.064)	.035 (.020)
H1-H2, /ɑ/ (dB)	−4.84 (3.49)	−0.24 (2.10)

day [$F(1,6) = 17.11$, $p = .006$]. Mean values for jitter for male speakers were .084 ms for samples recorded in the morning and .136 ms for samples recorded in the afternoon.

Consonantal Context

Although consonantal context was used in ANOVAs as a main effect, planned comparisons were done also to determine the effects on laryngeal activity of the consonantal features of voicing and manner, as well as combinations of these features. Specifically, planned comparisons were done to try to locate differences in effects for the three major groups of consonants used here: (1) the voiced stop /d/, (2) the voiceless stops /t/ and /k/, and (3) the voiceless fricatives /s/ and /ʃ/. These planned comparisons were as follows:

consonant voicing: /d/ versus the other four consonants
stop voicing: /d/ versus /t/ and /k/
the voiced stop versus voiceless fricatives: /d/ versus /s/ and /ʃ/
voiceless fricatives versus voiceless stops: /s/ and /ʃ/ versus /t/ and /k/.

Table 2 displays consonant means and standard deviations for all dependent measures. For SNR, BLSNR, and shimmer, these values represent means across all independent variables other than consonantal context (i.e., speaker sex, time of day, and vowel identity). Because F_0 varied significantly as a function of speaker sex, means across time of day and vowels are given separately for male and female speakers. Means for H1-H2, across time of day, are given separately for men and women, but for /ɑ/ only. For jitter, means are given separately for men and women, and the means for men are further separated by time of day.

Table 3 displays significant ANOVA results for the overall effect of consonantal context on the measures, as well as results for the planned comparisons. For jitter, the results of these analyses are shown for samples from male speakers only because samples from female speakers demonstrated neither a significant overall effect of consonantal context nor any significant planned comparison. Taken together, Tables 2 and 3 indicate a consistent pattern of change, depending on consonantal context: The voiced stop, /d/, is associated with the highest values for SNR and BLSNR, but with the lowest values for all other measures. The voiceless stops, /t/ and /k/, are affiliated with the opposite extremes for all measures, that is, the lowest values for SNR and BLSNR, but the highest values for all other measures. Values for the voiceless fricatives, /s/ and /ʃ/, fall midway between the voiced and voiceless stops.

Vowel Identity

Vowel identity was entered into the overall ANOVAs as a main effect, and planned comparisons were done also to evaluate the effects both of tongue height (/ɑ/ vs. /i/ and /u/) and of the positioning of the tongue body along the horizontal plane (/i/ vs. /u/). Table 4 displays means and

TABLE 2. Means (and standard deviations) of dependent variables for consonantal context. Each mean is computed across a different set of independent variables (SNR, BLSNR, and shimmer: across all other independent variables; F_0: across time of day and vowel identity; H1-H2: across time of day; jitter: across vowel identity for men, across vowel identity and time of day for women).

Dependent variable	*Consonantal context* /d/	/s/	/ʃ/	/t/	/k/
SNR (dB)	23.2	23.2	23.0	22.1	22.5
	(2.8)	(3.0)	(2.9)	(3.0)	(2.9)
BLSNR (dB)	8.24	8.09	7.96	7.33	7.74
	(2.94)	(3.43)	(3.18)	(2.71)	(3.15)
shimmer (%)	2.90	3.36	3.25	3.88	3.86
	(1.13)	(1.55)	(1.48)	(1.95)	(2.36)
F_0 (Hz)					
men	133	136	137	137	138
	(26)	(27)	(29)	(29)	(25)
women	198	207	207	212	212
	(10)	(14)	(13)	(13)	(14)
H1-H2, /ɑ/ only (dB)					
men	−6.03	−5.41	−5.02	−3.74	−3.99
	(3.23)	(3.74)	(3.47)	(3.85)	(3.52)
women	−1.01	−0.47	−0.10	0.28	0.11
	(1.76)	(1.98)	(2.74)	(2.10)	(2.07)
jitter (ms)					
men					
morning	.073	.074	.077	.098	.096
	(.036)	(.042)	(.049)	(.062)	(.059)
afternoon	.089	.126	.148	.150	.168
	(.046)	(.067)	(.063)	(.065)	(.061)
women	.034	.034	.032	.038	.038
	(.014)	(.016)	(.014)	(.028)	(.026)

TABLE 3. Significant ANOVA results for overall consonantal context effects and planned comparisons.

Dependent variable	*Effect*	*Degrees of freedom*	*F*	*P*
SNR	Overall C context	4,24	7.44	<.001
	/d/ vs /t/ & /k/	1,6	14.73	=.009
	/s/ & /ʃ/ vs /t/ & /k/	1,6	40.79	<.001
BLSNR	/d/ vs /t/ & /k/	1,6	6.46	=.04
shimmer	Overall C context	4,24	4.35	=.01
	/d/ vs /t/ & /k/	1,6	7.80	=.03
	/s/ & /ʃ/ vs /t/ & /k/	1,6	9.47	=.02
F_0	Overall C context	4,24	8.71	<.001
	/d/ vs others	1,6	13.99	=.01
	/d/ vs /t/ & /k/	1,6	15.76	=.007
	/d/ vs /s/ & /ʃ/	1,6	11.07	=.02
	/s/ & /ʃ/ vs /t/ & /k/	1,6	9.09	=.02
H1-H2 (/a/)	Overall C context	4,24	7.10	<.001
	/d/ vs others	1,6	13.34	=.01
	/d/ vs /t/ & /k/	1,6	17.94	=.006
	/d/ vs /s/ & /ʃ/	1,6	6.92	=.04
	/s/ & /ʃ/ vs /t/ & /k/	1,6	23.06	=.003
jitter (males)	Overall C context	4,24	13.54	<.001
	/d/ vs others	1,6	33.50	=.001
	/d/ vs /t/ & /k/	1,6	40.26	=.001
	/d/ vs /s/ & /ʃ/	1,6	18.23	=.005
	/s/ & /ʃ/ vs /t/ & /k/	1,6	19.73	=.004

standard deviations for all dependent measures as a function of vowel identity. The main effect of vowel was statistically significant across all speakers for both BLSNR and shimmer [$F(2,12) = 5.82$, $p = .02$, and $F(2,12) = 6.71$, $p = .01$, respectively], as was the planned comparison for front/back tongue positioning in the case of BLSNR [$F(1,6) = 10.47$, $p = .02$] and the planned comparison for tongue height in the case of shimmer [$F(1,6) = 7.03$, $p = .04$]. F_0 demonstrated a significant vowel effect [$F(2,12) = 10.96$, $p = .002$], and the planned comparisons for both tongue height and front/back positioning were significant [$F(1,6) = 10.74$, $p = .02$, and $F(1,6) = 15.37$, $p = .008$, respectively]. For jitter, the overall effect of vowel was significant for samples from male speakers only [$F(2,12) = 6.16$, $p = .01$], as was the comparison of tongue height [$F(1,6) = 6.25$, $p = .047$]. These results for vowel identity show similar trends to those found for consonantal context, although perhaps not as clearly. For changes in consonantal context, a consistent pattern of decreasing SNR and BLSNR values and increasing values on all other measures was observed. For changes in vowel identity, F_0 and jitter (for men) both increase from the low vowel to the higher vowels, and F_0 increases further from the high, front vowel to the high, back vowel. BLSNR shows the same relation to F_0 in different vowels as it showed in different consonantal contexts; it decreases as F_0 increases from the high, front to the high, back vowel. Only shimmer does not follow the pattern of increase with F_0 that it did for consonantal context effects because it decreases in value from the low vowel to the high vowels.

TABLE 4. Means (and standard deviations) for dependent variables for vowel identity. Each mean is computed across a different set of independent variables (SNR, BLSNR, and shimmer: across all other independent variables; F_0: across time of day and consonantal context; jitter: across consonantal context for men, across consonantal context and time of day for women).

Dependent variable	*Vowel* /i/	/a/	/u/
SNR (dB)	22.2	22.6	23.5
	(3.0)	(2.8)	(2.9)
BLSNR (dB)	8.37	8.33	6.92
	(3.12)	(3.34)	(2.54)
shimmer (%)	3.27	4.07	3.00
	(1.54)	(2.14)	(1.37)
F_0 (Hz)			
men	138	132	139
	(28)	(23)	(29)
women	210	197	215
	(13)	(7)	(14)
jitter (ms)			
men			
morning	.088	0.76	.086
	(.054)	(.036)	(.060)
afternoon	.151	.104	.153
	(.067)	(.042)	(.071)
women	.029	.050	.027
	(.011)	(.028)	(.007)

CORRELATIONS

Pearson product-moment correlation coefficients (r) were computed between certain dependent measures as a way of answering eight specific questions concerning the laryngeal activity specified by each measure, and of investigating possible artifacts.

1. *To what extent are SNR and BLSNR correlated?* This question was addressed first because, if it were found that these measures were very highly correlated,

then other correlational tests would not need to be done separately for each. The correlation between SNR and BLSNR was .78. Therefore, 61% of the variance in each measure is explained by the other, but there is still 39% of unexplained variance. Consequently, all other correlational analyses were done for SNR and BLSNR separately.

2. *Are these measures of signal-to-noise ratio related to either the amount of jitter or shimmer in the glottal cycles?* It would be valuable to have a measure that quantifies the amount of aspiration noise in the voice source independently of jitter and shimmer, but it was never clear whether or not measures used in earlier studies provided such information. Our analysis for SNR, which is one of the previously used measures, suggests that it does not: Correlation coefficients computed between SNR and jitter showed a moderately strong relation ($r = -.43$), but coefficients computed for male and female speakers separately indicated that this result was due entirely to samples from male speakers ($r = -.71$ for males and $r = -.11$ for females). The correlation, across speaker sex, between SNR and shimmer showed a moderately strong relation ($r = -.61$). Separate correlation coefficients were not computed for SNR and shimmer for men and women because neither measure showed a speaker-sex effect.

It had been hoped that the band-limited signal-to-noise ratio (BLSNR) used here would achieve the goal of quantifying aspiration noise more independently from the influences of jitter and shimmer. However, correlational analyses for BLSNR showed similar, though somewhat weaker, results to those obtained for SNR: The correlation coefficient computed across speaker sex between BLSNR and jitter was −.40, and coefficients computed for men and women separately were −.67 and −.02, respectively. This result for women was not significant. The correlation coefficient computed between BLSNR and shimmer, across speaker sex, was −.47.

3. *Are the measures of signal-to-noise ratio (SNR and BLSNR) related to the acoustic measure of volume-velocity waveform shape (H1-H2)?* The purpose of this question was to examine whether or not the relative amount of noise in the voicing source seems to vary with the relative amplitude of the fundamental, as might be predicted given the greater open quotient in speech with higher-amplitude fundamentals. If this were the case, strong, negative correlations would be expected between the signal-to-noise ratios and H1-H2. When SNR and BLSNR were correlated with H1-H2 for samples with the vowel /ɑ/, neither correlation was found to be statistically significant: $r = .21$ for SNR versus H1-H2 and $r = .06$ for BLSNR versus H1-H2. However, when correlation coefficients were computed for men and women separately, each showed moderately strong relations: r for SNR versus H1-H2 was .59 for men and −.43 for women; r for BLSNR versus H1-H2 was .44 for men and −.54 for women. These results for women confirm what might have been predicted: As the relative amplitude of the fundamental increases, the amplitude of the noise increases, relative to the signal. The results for men were unexpected because they indicate that those samples from male speakers with higher H1-H2 values actually have improved signal-to-noise ratios.

To try to make sense of this seeming inconsistency, Stepwise Multiple Regressions were done for samples from men and women separately. H1-H2 was used as the dependent variable, and SNR, BLSNR, jitter, and shimmer were the independent variables. Initial correlations of these last two variables with H1-H2 were negative and not particularly strong (for men and women, respectively, r for jitter versus H1-H2 was −.32 and −.40, and r for shimmer versus H1-H2 was −.10 and −.05). For women, BLSNR was most highly correlated with H1-H2, and so was entered into the Multiple Regression first. Once this was done, the partial correlation for SNR and H1-H2 was effectively zero (.09). Jitter, now with a partial correlation to H1-H2 of −.53, was entered second. This step left neither SNR nor shimmer with sufficiently high partial correlation coefficients to be entered into the analysis. For men, SNR was entered into the analysis first, and this resulted in partial correlations between H1-H2 and jitter of .09, and between H1-H2 and BLSNR of −.34 (which is now in the predicted direction). So, for women's samples, removing the variance of H1-H2 explained by BLSNR left virtually no additional variance that could be explained by SNR, but left quite a bit that could be explained by jitter. For men's samples, on the other hand, removing the variance in H1-H2 explained by SNR left virtually no additional variance to be explained by jitter, but left some to be explained by BLSNR. Therefore, it seems likely the positive correlation between H1-H2 and the signal-to-noise ratios (SNR and BLSNR) found for men's samples may have reflected the high degree to which SNR and BLSNR are inversely related to jitter.

4. *Is the amount of jitter related to F_0, as suggested by the increased jitter found for male speakers over female speakers?* The coefficient for F_0 versus jitter showed a moderately strong relation ($r = -.69$) when computed across speaker sex. This value is slightly weaker than those obtained by Orlikoff and Baken (1988) for men and women separately ($r = -.73$ and $r = -.88$, respectively). Neither coefficient computed for male and female speakers separately in this study demonstrated a relation of this strength, although the coefficient computed for men was stronger than the one for women ($r = -.44$ for men and $r = -.19$ for women). These results indicate that the correlation obtained by us across speaker sex reflected, to a great extent, the main effect of speaker sex on both jitter and F_0, but that male speakers did demonstrate, to some extent, the inverse relation between F_0 and jitter reported by Orlikoff and Baken.

Because F_0 and jitter increased together across consonantal contexts and vowels, it was possible that these main effects were attenuating the strength of the negative relation between F_0 and jitter. To examine this possibility, correlation coefficients were computed for each syllable separately, across speaker sex. All but two of these coefficients were greater than the overall r of −.69 obtained across syllables. Values ranged from −.60 to −.91, with a mean of −.74. Correlation coefficients for each

syllable were not computed for men and women separately because too few values would contribute to each.

5. *Is the measure of H1-H2 a function of F_0?* The purpose of this question was to investigate whether speakers make laryngeal adjustments (affecting the shape of the volume-velocity waveform) with variation in fundamental frequency due to consonantal context. Correlation coefficients between H1-H2 and F_0 for samples with the vowel /ɑ/ showed a rather strong relation (r = .79). However, when separate coefficients were computed for men and women, a significant relation was found for men only (r = .76 for men, and r = .10 for women).

6. *Is the measure of H1-H2 affected by variation in the frequency of F1?* In other words, to what extent is this measure of volume-velocity waveform shape affected by artifactual consequences of F1 moving into closer proximity to the second harmonic? If variation in H1-H2 were due only to spectral envelope changes with shifting F1, a positive correlation would be expected between H1-H2 and F1 (Fant, 1960). Correlation coefficients between H1-H2 and F1 did not demonstrate a very strong relation across speaker sex (r = .48), demonstrated a nonsignificant relation for male speakers (r = .10), but demonstrated a negative relation for female speakers (r = −.48). This last finding is illustrated in Figure 1. As can be seen, Speaker 4F has F1 values that are much more variable and generally lower than those of the other female speakers. In addition, her H1-H2 values are somewhat higher than most values for the other female speakers. When analyses were redone discarding her data, a much smaller correlation coefficient was found between F1 and H1-H2 across speaker sex (r = .29), and a nonsignificant one was obtained for female speakers (r = −.16). Thus, the negative correlation originally found for female speakers was attributable to a main effect of this 1 speaker being different from the others on both measures of F1 and H1-H2, and the positive correlation found across speaker sex was due to the main effects of men having lower values for both F1 and H1-H2.

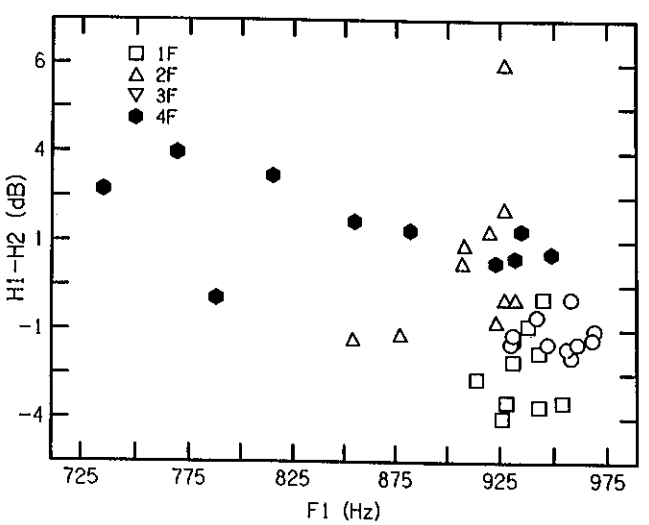

FIGURE 1. F1 and H1-H2 for female speakers for samples from the vowel /ɑ/. The ten points per speaker represent two times of day × five consonantal contexts.

7. *Are either the signal-to-noise ratios or H1-H2 related to the duration of the vowel nucleus?* In other words, was either more noise or more symmetric volume-velocity waveforms observed simply as an artifact when the 10 pitch periods analyzed occurred closer to glottal adduction and abduction gestures? The correlation coefficient computed across all vowels for dur versus SNR and dur versus BLSNR demonstrated only weak relations (r = .23 and r = .18, respectively), and the coefficient for dur versus H1-H2 for samples containing /ɑ/ demonstrated no relation (r = −.08).

8. *Are the measures of signal-to-noise ratio used here affected by the spacing of harmonics (i.e., fundamental frequency)?* The purpose of this question was to investigate the possibility that noisier spectra have been observed for women's speech than for men's due to the fact that voices with higher fundamental frequencies have more widely spaced harmonics. Therefore, women's speech spectra could appear noisier because there are relatively fewer harmonics contributing to the signal component, rather than because the noise component is enhanced. Correlation coefficients were computed between F_0 and both SNR and BLSNR to address this issue. Both correlation coefficients were quite small (r = .21 for F_0 vs. SNR; r = .19 for F_0 vs. BLSNR), suggesting that the spacing of harmonics did not greatly affect these signal-to-noise ratios. Correlation coefficients computed for men and women separately were higher (values for F_0 vs. SNR were .57 for men and .35 for women; those for F_0 vs. BLSNR were .40 for men and .31 for women), indicating that the degree of relation between these two variables was underestimated when groups were combined. However, these coefficients are positive, indicating that signal-to-noise ratios actually improved with increases in F_0. This is contrary to what would be predicted if more widely spaced harmonics were responsible for the observation of more noise in voices with higher F_0s.

DISCUSSION

MAIN EFFECTS

Speaker Sex

Although differences in procedures constrain the validity of cross-study comparisons, the effects of sex found here are in general agreement with those reported in other studies. Mean fundamental frequency for female speakers was 1.5 times higher than for male speakers, with male values similar to those of most other studies and female values lower than some. The amplitude of the fundamental, relative to that of the second harmonic, was 4.6 dB higher (i.e., less negative) for women than for men. This is slightly less than the 5.8 dB and 5.7 dB differences reported by Henton and Bladon (1985) and Klatt and Klatt

(1990), respectively. Values derived by us, in general, also appear a little lower than previously reported values.

Jitter values obtained for women were similar to those reported by Orlikoff and Baken (1988), but were a bit higher than those of Milenkovic (1987). By contrast, the obtained values for men were higher than those from any previous work. Female voices demonstrated less jitter than male voices, and the amount of jitter in samples from female speakers remained constant across other manipulations. In agreement with results of Horii (1979) and of Orlikoff and Baken (1988), the samples taken from male speakers in the present study showed greater variability in jitter as a function of F_0 than those from female speakers, but not necessarily in the direction found previously. According to these earlier studies, in which speakers were instructed to vary pitch while retaining the same isolated vowel, jitter should decrease with increasing fundamental frequency. All Pearson product-moment correlation coefficients from the present study met this prediction because they were negative, but correlations were not as great as might have been predicted from the earlier work. However, across linguistic variations in the utterances, which affected fundamental frequency, a pattern opposite to this prediction is observed. Any linguistic manipulation that served to increase fundamental frequency is associated with a discrete increase in jitter (see Tables 2 and 4).

The lack of a speaker-sex effect on the total signal-to-noise ratio is in agreement with the results of Yumoto et al. (1982) and Milenkovic (1987), but our correlational tests suggest that this result is due to different factors affecting the measure for each speaker group. Male speakers demonstrated substantially more jitter than female speakers, and also showed a rather strong correlation between jitter and SNR. Female speakers demonstrated relatively little jitter, and no correlation was found between jitter and SNR, although SNR values were comparable to men's. These results suggest that jitter accounted for the largest proportion of variance in SNR for men's samples because it was the greatest source of noise, and that some other noise component was primarily responsible for variation in SNR for women's samples. Our results alone, however, fail to specify whether the additional noise quantified by the signal-to-noise ratios for women's samples is aspiration noise per se or just an additional, undefined noise component. The observation of Klatt and Klatt (1990) of more aspiration noise in the high-frequency regions of women's spectra than in those of men's leads us to speculate that it is most likely aspiration noise.

Time of Day

In general, these measures of laryngeal activity were not affected by the time of day at which the sample was obtained. This should be welcome news to speech clinicians who need to evaluate patients with vocal pathologies, and have constraints on scheduling. The one exception was jitter in male voices. However, given the sensitivity of this measure to other factors, it may not be the best candidate for clinical use.

Consonantal Context

Ideally, what would be wanted for evaluating the effects of consonantal context on laryngeal activity (i.e., laryngeal/supralaryngeal coarticulation) would be a measure showing variation with all changes in consonantal context, but showing no variation as a function of vowel. All measures made here, except BLSNR and jitter, were sensitive to the overall effects of consonantal context across speaker sex. However, only two (F_0 and H1-H2) demonstrated significant effects on all planned comparisons. The first of these, F_0, is probably the simplest to measure, but also demonstrates strong intrinsic vowel effects. The second, H1-H2, can only be measured in low vowels, unless a correction for the influence of F1 is used. Thus, both these measures may be of limited value for measuring coarticulatory effects. Although SNR and shimmer did not demonstrate significant effects on all planned comparisons, they did demonstrate significant effects of voicing within a manner class (/d/ vs. /t/ and /k/), and significant effects of manner within a voicing class (/t/ and /k/ vs. /s/ and /ʃ/). Moreover, SNR can be measured for all vowels (unlike H1-H2), and is not sensitive to vowel effects (unlike F_0 and shimmer), making it a good choice for use in studies of coarticulation in which only one feature of consonant production is varied. Thus, four acoustic measures (F_0, H1-H2, SNR, and shimmer) seem appropriate for studies of laryngeal/supralaryngeal coarticulation, although no one measure is ideal. The choice of measure to be used should be determined by the phonological structure of the utterances.

The consonantal context effects found here have implications for hypotheses concerning the laryngeal mechanisms used in consonantal voicing. All results are consistent with physiologic studies showing differences in the level of activity of several laryngeal muscles as a function of consonant voicing and manner, with differences continuing into the vowel segment (e.g., Löfqvist & Yoshioka, 1981, 1984; Löfqvist et al., 1989). Many of these muscular adjustments used to control abduction/adduction have also been shown to affect vocal fold tension. The findings reported here fail to support suggestions that the fundamental-frequency difference at vowel onset observed between voiced and voiceless stops is due entirely to a lowering of the larynx during production of voiced stops to maintain voicing through the closure. If the voiced/voiceless stop distinction were due entirely to a lowering of the larynx for the voiced condition, then there should have been no difference in fundamental frequency between samples produced with prevocalic voiceless fricatives and stops. However, we observed a difference between these consonantal contexts.

The effects of consonantal context on jitter in samples from male speakers also provide interesting insights into laryngeal activity. Jitter increased in value across contexts along with all other measures (except SNR and BLSNR),

as described earlier, but was therefore related to fundamental frequency in a way that was not predicted by previous work. In earlier studies, jitter was found to be inversely related to fundamental frequency. However, in those studies, fundamental frequency was varied not by changing context, but by simply asking speakers to modify pitch while keeping linguistic factors constant. Apparently, changes in laryngeal activity made solely to vary fundamental frequency do not affect source characteristics in the same way as changes made to modify consonant voicing and manner. In the former case, decreasing jitter is associated with increasing fundamental frequency; in the latter case, jitter increases as the frequency of the fundamental does.

A possible explanation for the variation in jitter with consonantal context (for male speakers) may rest with the laryngeal adjustments necessary for switching from consonant to vowel. The adjustments involved in the adduction of the folds may last well into the central portion of the vowel following voiceless consonants (Gobl, 1988). Following voiced consonants, the adjustments appear to be completed much sooner (Löfqvist & McGowan, 1989), either because abduction did not occur at all or was not as great. Apparently, the adjustments required to make the transition from abducted to adducted folds lead to increased jitter, which overrides any decrease in jitter associated with increased fundamental frequency. Consequently, a positive correlation is obtained between jitter and fundamental frequency across variation in consonantal context. It is not clear why female speakers fail to show consonantal context effects on jitter, but it is probably because women's voices simply display low jitter, regardless of variation in independent variables. However, this fact makes jitter a poor choice for measuring coarticulatory effects across different populations.

The effects of consonantal context on the harmonic-amplitude difference also provide new insights into variations in laryngeal activity. It could simply be that these effects result from variation in fundamental frequency associated with consonant voicing and manner. However, two additional findings suggest that there may be something more happening. First, Table 2 shows that the maximum variation among consonantal contexts in mean H1-H2 is 2.29 dB for men, but only 1.12 dB for women. Second, there was a significant correlation between H1-H2 and F_0 for men, but not for women. These findings were obtained in spite of the fact that women actually showed greater variation than men in F_0 as a function of consonantal context (7% vs. 4%), and suggest that there may be a frequency limit above which it is no longer desirable to allow the relative amplitude of the fundamental to increase with increasing frequency. Female speakers may restrict the variation in the relative amplitude of the fundamental across variation in fundamental frequency, resulting in less-steep spectral tilts with increasing fundamental frequency, in order to preserve sufficient harmonic energy in the higher formants. This suggestion, that speakers may control the shape of the glottal waveform and consequently spectral tilt, is supported by Monsen and Engebretson (1977), who found evidence of such control for stress and sentence type.

The magnitude of the consonantal context effect on H1-H2 is also of interest here. Although a consistent effect was obtained across speakers, it was relatively small, especially compared to the interspeaker variability demonstrated within contexts. However, these measures were made on vocalic centers. Gobl and Ní Chasaide (1988) traced the change in spectral tilt through the vowel in /bɑ/ and /pɑ/ syllables. Their results showed the largest difference between voicing conditions at vowel onsets, with these differences reaching a minimum approximately halfway through the vowel. This combination of findings across studies suggests that normal adult speakers make substantial laryngeal adjustments early in vowel production, which affect the spectral tilt of the voicing source, but that some slight difference can continue throughout the vowel. Thus, H1-H2 should be able to provide information about differences in the strength and time course of these laryngeal adjustments among different speaker populations.

Vowel Identity

Three measures (F_0, BLSNR, and shimmer) were affected by vowel identity across speaker sex. For fundamental frequency, this effect was exactly what would have been predicted by earlier studies of intrinsic pitch: F_0 was higher for high than for low vowels. This finding by itself does not provide separate support for either hypothesis concerning the physiologic basis of the effect (i.e., that it is associated either with the forward displacement of the hyoid bone or with the height of the mandible). However, the finding of slightly higher F_0 for /u/ than for /i/ may be helpful in deciding between these hypotheses. Mandible height should be similar for /u/ and for /i/. Therefore, this finding suggests that intrinsic vowel effects, at least to some extent, are associated with the displacement of the hyoid bone, and that it is slightly more forward for high, back than for high, front vowels.

BLSNR varied with fundamental frequency across vowels in a way that was somewhat similar to effects found for consonantal context: Those conditions for which higher fundamental frequencies were observed demonstrated lower BLSNR. Jitter, which showed vowel effects for men only, also demonstrated a pattern similar to that obtained of consonantal context, increasing in the same conditions as fundamental frequency. These trends of increased noise and jitter with increased F_0 across vowels cannot be explained by laryngeal articulatory gestures as they were for consonantal context effects. We can only speculate that the laryngeal adjustments responsible for the intrinsic pitch of vowels also affect these parameters. Shimmer, on the other hand, displayed a pattern strikingly different from those of the other acoustic measures: It decreased in those conditions in which F_0 increased. Although we are unable to explain this result, it is in agreement with previously reported vowel effects for shimmer (Horii, 1980; Milenkovic, 1987). It

may be that the trends observed here for measures other than F_0 are related to acoustic source-filter interactions. These interactions have been shown to change the shape of the volume-velocity pulse independently of laryngeal tissue adjustments (Rothenberg, 1983). Although mechanical adjustments of the folds as vowels change may affect F_0, associated loading differences may affect voice quality.

CORRELATIONS

The correlational results found here address several issues regarding the laryngeal activity specified by each measure, as well as possible artifacts in these acoustic measures.

Laryngeal activity specified by each measure. Perhaps the most important correlational results concern those obtained for measures of the harmonic-amplitude difference and the signal-to-noise ratios. Upon first consideration, the correlation coefficients obtained across speaker sex for H1-H2 and each of the signal-to-noise ratios (SNR and BLSNR) do not seem to warrant the suggestion that these measures evaluate inextricably linked phenomena. However, the strength of the relations between H1-H2 and the signal-to-noise ratios increased when men's and women's samples were analyzed separately. At least for women's samples, these correlation coefficients clearly indicate that the amount of additional noise (probably aspiration noise) in the signal increases as the relative amplitude of the fundamental increases. Therefore, it seems reasonable to speculate that both the relative amplitude of the fundamental and the amount of noise in the signal are related to the shape of the volume-velocity waveform in women's speech. Furthermore, all results suggest that the signal-to-noise ratios used here provide a measure of aspiration noise per se for women's speech samples.

For men's samples, which demonstrated improved signal-to-noise ratios with increased H1-H2 values, it at first appears as if either the amount of aspiration noise in the signal does not increase (or, in fact, decreases) as the relative amplitude of the fundamental increases, or that the signal-to-noise ratios used here were not valid indicators of this additional noise. Related to this issue is the finding of decreased jitter with increases in H1-H2. This trend could actually have led to the unpredicted, positive correlations between H1-H2 and the signal-to-noise ratios for men's samples because jitter accounts for such a large proportion of variance in both SNR and BLSNR in men's samples. Consequently, as jitter decreased with increasing H1-H2, these signal-to-noise ratios showed an obligatory improvement in men's samples. Thus, it was difficult to obtain an accurate indication of the relation between aspiration noise and H1-H2, but not impossible. This goal was accomplished by removing the common source of variance (jitter) to H1-H2 and the signal-to-noise ratios, and then computing the partial correlation coefficients. When this was done, the partial correlation between H1-H2 and BLSNR of $-.34$ suggested that aspiration noise did increase with H1-H2. Overall, these analyses indicate that the amount of aspiration noise in men's voices was much less, and the amount of jitter was much more, than in women's voices. This combination of characteristics meant that the signal-to-noise ratios for men's samples were highly influenced by variation in jitter, and fairly immune to variation in the amount of aspiration noise.

We had expected that there would be interactions among the different measures of voice perturbation, based on the results of previous studies. (Milenkovic, 1987; Muta et al., 1988, Yumoto et al., 1982). The most serious interaction was expected to involve the influence of high levels of jitter on the signal-to-noise ratios, possibly decreasing the validity of these ratios as indicators of the amount of aspiration noise in the voice. We attempted to deal with this problem by inverse filtering the acoustic signal. Inverse filtering reduces the waveform overlap between successive pitch periods that occurs when formant oscillations excited in one pitch period carry over to the next. However, the only way that we could have completely isolated the effect of jitter from the signal-to-noise ratios would have been if the inverse filter were able to isolate each individual pitch period, and if the effect of jitter was to shift the pitch pulses in time without changing their shape. In the absence of this ideal condition, jitter adds to the apparent aperiodic component of the acoustic waveform, thereby reducing the measured signal-to-noise ratios. The effect of jitter on the signal-to-noise ratios is greatest for voices demonstrating high levels of jitter.

The results reported here also indicate that SNR is affected by shimmer in both men's and women's samples, and provide some insight into the source of this shimmer. To be precise, 37% of the variance in SNR is associated with shimmer for all speakers. Restricting this signal-to-noise ratio to the spectral region associated with formants higher than F1, primarily with F3, reduces but does not completely eliminate this association with shimmer: 22% of the variance in BLSNR is explained by shimmer. Because the association of shimmer with BLSNR is weaker than with SNR, it seems likely that the amplitude fluctuations being measured here are not due entirely to the addition of turbulence noise to the volume-velocity source. If this were the case, shimmer would account for at least as much of the variance in BLSNR as it accounts for in SNR. Moreover, the extremely weak partial correlations obtained between H1-H2 and shimmer, especially for women ($-.05$), suggests that these amplitude fluctuations are largely due to something other than the addition of turbulence noise, probably fluctuations in the amplitude of the glottal waveform. However, the exact source(s) of the shimmer measured here cannot be identified.

Possible artifacts. One conclusion of the correlational analyses done here was that the harmonic-amplitude difference, computed for low vowels, is not so strongly affected by the frequency of the first formant as to make it an unreliable measure of spectral tilt across speakers. In addition, both signal-to-noise ratios and the harmonic-amplitude difference were found to be independent of proximity of measurement to glottal adduction and abduction gestures when vowel centers are analyzed, as we did. Finally, the signal-to-noise ratios were not greatly affected by variation in the spacing of harmonics. The

correlation coefficients computed for F_0 versus SNR and BLSNR for men and women separately were between .31 and .57, with values for men higher than those for women. Initially, these findings seem difficult to interpret. The purpose of computing these correlation coefficients was to investigate whether aperiodic noise (mainly aspiration noise) seems to increase with rising F_0 merely as a consequence of the harmonics becoming more widely spaced. Negative correlations between F_0 and the signal-to-noise ratios would be expected if this were the case, but were not found. For women, the relative amplitude of the fundamental does not vary greatly as a function of F_0, thus creating flatter source spectra with increasing F_0. Therefore, even though the harmonics may be more widely spaced, the strength of that harmonic structure remains fairly constant relative to the background noise. For men, signal-to-noise ratios primarily index the amount of jitter, and jitter is inversely related to F_0. Therefore, signal-to-noise ratios improve with increasing F_0 because jitter is decreasing. This effect apparently cancels any possible effect of more widely spaced harmonics.

SUMMARY

The present study has provided some normative data on several acoustic measures of laryngeal activity for men and women producing vowels in various consonantal contexts. Such data for a variety of vowels and consonantal contexts previously have not been available.

A major goal of this work was to identify acoustic measures that could be used to investigate coarticulatory effects on laryngeal activity. All six measures used here varied to some extent as a function of the preceding consonant: F_0, jitter, shimmer, and H1-H2 tended to increase in value from the voiced-stop to the voiceless-fricative to the voiceless-stop contexts, whereas SNR and BLSNR tended to decrease in value across these contexts. However, only four measures demonstrated broad effects for consonant manner and/or voicing across speaker sex (F_0, H1-H2, SNR, and shimmer), suggesting that they would be most appropriate for future studies of laryngeal/supralaryngeal coarticulation. In addition to the consonantal context effects on laryngeal activity found here, variation was found in three acoustic measures with vowel identity across speaker sex (F_0, shimmer, and BLSNR).

Another goal of this work was to gain some insight into the laryngeal mechanism(s) responsible for these acoustic effects. Good evidence was found to support the suggestion that the observed consonantal context effects were due to laryngeal adjustments associated with the abduction and/or subsequent adduction gestures for voiceless consonants. Vowel effects on fundamental frequency seemed to be explained best by the horizontal placement of the hyoid bone, but the mechanism underlying vowel effects on other measures was less apparent. It was also found that the amount of aspiration noise (measured here by signal-to-noise ratios) seemed to be associated with the shape of the volume-velocity waveform so that more symmetrical waveforms were correlated with more noise.

Finally, an important conclusion that came out of these measurements was that covariation among acoustic parameters depends on linguistic context and the sex of the speaker. For instance, the variation of jitter with fundamental frequency depended on speaker sex and linguistic context. Similarly, modification in H1-H2 as a function of fundamental frequency was influenced by both speaker sex and consonantal context. Thus, there seem to be laryngeal adjustments that are made for linguistic purposes, but the exact nature of these adjustments varies across speaker populations. Acoustic measures can provide tools for investigating these differences.

ACKNOWLEDGMENTS

This work was supported by NIH Grant DC-00633 to the first author. We are grateful to Steven M. Barlow, Yoshiyuki Horii, Stephen T. Neely, and an anonymous reviewer for their comments on earlier drafts of this article.

REFERENCES

ATAL, B. S. (1982). Predictive coding of speech at low bit rates. *IEEE Transactions on Communications COM-30*, 600–614.

BICKLEY, C. (1982). Acoustic analysis and perception of breathy vowels. *MIT R.L.E. Speech Communications Group: Working Papers, 1*, 71–82.

BLESS, D. M., BIEVER, D. M., CAMPOS, G., GLAZE, L. E., & PEPPARD, R. (1989, August). *Videostroboscopic, acoustic, and aerodynamic analysis of voice production in normal adults.* Paper presented at the Vocal Fold Physiology Conference, Stockholm.

COOPER, W., & SORENSEN, J. (1981). *Fundamental frequency in sentence production*, New York: Springer-Verlag.

CROCHIERE, R. E., & RABINER, L. R. (1983). *Multirate Digital Signal Processing*, Englewood Cliffs, NJ: Prentice Hall.

FANT, G. (1960). *Acoustic theory of speech production*, The Hague: Mouton.

FANT, G. (1986). Glottal flow: Models and interaction. *Journal of Phonetics, 14*, 393–399.

FANT, G., LILJENCRANTS, J., & LIN, Q. G. (1985). A four-parameter model of glottal flow. *Speech Transmission Labs QPSR 4*, Royal Institute of Technology, Stockholm, 1–13.

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

GOBL, C. (1988). Voice source dynamics in connected speech. *KTH Speech Transmission Laboratory-QPSR, 1*, 123–189.

GOBL, C., & NÍ CHASAIDE, A. (1988). The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study. *KTH Speech Transmission Laboratory-QPSR, 2/3*, 23–59.

HENTON, C., & BLADON, R. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication, 5*, 221–227.

HONDA, K. (1983). Relationship between pitch control and vowel articulation. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 286–297). San Diego: College-Hill Press.

HORII, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research, 22*, 5–19.

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

HORII, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research, 25*, 12–14.

HOUSE, A., & FAIRBANKS, G. (1953). The influence of consonant

environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25,* 105–113.

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

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

LADEFOGED, P., & ANTOÑANZAS-BARROSO, N. (1985). Computer measures of breathy voice quality. *UCLA Working Papers in Phonetics, 61,* 79–86.

LADEFOGED, P., MADDIESON, I., & JACKSON, M. (1988). Investigating phonation types in different languages. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms, and functions* (pp. 297–317). New York: Raven Press.

LEHISTE, I., & PETERSON, G. (1961). Some basic considerations on the analysis of intonation. *Journal of the Acoustical Society of America, 33,* 419–425.

LÖFQVIST, A., BAER, T., MCGARR, N. S., & STORY, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85,* 1314–1321.

LÖFQVIST, A., MCGARR, N. S., & HONDA, K. (1984). Laryngeal muscles and articulatory control. *Journal of the Acoustical Society of America, 76,* 951–954.

LÖFQVIST, A., & MCGOWAN, R. S. (1989, August). *Voice source variations in running speech.* Paper presented at the Vocal Fold Physiology Conference, Stockholm.

LÖFQVIST, A., & YOSHIOKA, H. (1981). Interarticulator programming in obstruent production. *Phonetica, 38,* 21–34.

LÖFQVIST, A., & YOSHIOKA, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication, 3,* 279–289.

LUDLOW, C., BASSICH, C., CONNOR, N., COULTER, D., & LEE, Y. (1987). The validity of using phonatory jitter and shimmer to detect laryngeal pathology. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 492–508). Boston: Little & Brown.

MARKEL, J. D., & GRAY, A. H., JR. (1976). *Linear prediction of speech.* Berlin: Springer-Verlag.

MILENKOVIC, P. (1986). Glottal inverse filtering by joint estimation of an AR system with linear input model. *IEEE Transactions of Acoustics, Speech, and Signal Processing ASSP-34,* 28–42.

MILENKOVIC, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research, 30,* 529–538.

MILENKOVIC, P., BLESS, D. M., & RAMMAGE, L. A. (1989, August). *Acoustic and perceptual characterization of vocal nodules.* Paper presented at the Vocal Fold Physiology Conference, Stockholm.

MOHR, B. (1971). Intrinsic variations in the speech signal. *Phonetica, 23,* 65–93.

MONSEN, R., & ENGEBRETSON, A. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America, 62,* 981–993.

MUTA, H., BAER, T., WAGATSUMA, K., MURAOKA, T., & FUKUDA, H. (1988). A pitch-synchronous analysis of hoarseness in running speech. *Journal of the Acoustical Society of America, 84,* 1292–1301.

OHALA, J. J. (1973). The physiology of tone. In L. Hyman (Ed.), Consonant types and tone. *Southern California Papers in Linguistics, 1,* 1–14.

OHALA, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.

OHDE, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75,* 224–230.

OPPENHEIM, A. V., & SHAFER, R. W. (1975). *Digital signal processing.* Englewood Cliffs, NJ: Prentice Hall.

ORLIKOFF, R., & BAKEN, R. (1988, June). *Vocal jitter at different fundamental frequencies.* Paper presented at the 17th Symposium: Care of the Professional Voice, New York, NY.

PETERSON, G., & BARNEY, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

ROTHENBERG, M. (1983). An interactive model for the voice source. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 155–165). San Diego: College-Hill Press.

SHADLE, C. (1985). Intrinsic fundamental frequency of vowels in sentence context. *Journal of the Acoustical Society of America, 78,* 1562–1567.

SORENSEN, D., & HORII, Y. (1982). Cigarette smoking and voice fundamental frequency. *Journal of Communication Disorders, 15,* 135–144.

TITZE, I. R. (1974). The human vocal cords: A mathematical model. *Phonetica, 29,* 1–21.

TITZE, I. R., HORII, Y., & SCHERER, R. C. (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research, 30,* 252–260.

UMEDA, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *Journal of the Acoustical Society of America, 70,* 350–355.

WILCOX, K., & HORII, Y. (1980). Age and changes in vocal jitter. *Journal of Gerontology, 35,* 194–198.

YANAGIHARA, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research, 10,* 531–541.

YUMOTO, E., GOULD, W., & BAER, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustic Society of America, 71,* 1544–1549.

ZAWADZKI, P. A., & GILBERT, H. R. (1989). Vowel fundamental frequency and articulator position. *Journal of Phonetics, 17,* 159–166.

Received December 18, 1989
Accepted June 4, 1990

Requests for reprints should be sent to Susan Nittrouer, Ph.D., Haskins Laboratories, 270 Crown Street, New Haven, CT 06511.