Glottal characteristics of female speakers: Acoustic correlates

Helen M. Hanson

Citation: The Journal of the Acoustical Society of America **101**, 466 (1997); doi: 10.1121/1.417991

View online: https://doi.org/10.1121/1.417991

View Table of Contents: https://asa.scitation.org/toc/jas/101/1

Published by the Acoustical Society of America

Glottal characteristics of female speakers: Acoustic correlates

Helen M. Hanson[a)]

Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-549, 50 Vassar Street, Cambridge, Massachusetts 02139

(Received 17 January 1996; accepted for publication 10 September 1996)

The aim of the research reported in this paper is to formulate a set of acoustic parameters of the voicing source that reflect individual differences in the voice qualities of female speakers. Theoretical analysis and observations of experimental data suggest that a more open glottal configuration results in a glottal volume-velocity waveform with relatively greater low-frequency and weaker high-frequency components, compared to a waveform produced with a more adducted glottal configuration. The more open glottal configuration also leads to a greater source of aspiration noise and larger bandwidths of the natural frequencies of the vocal tract, particularly the first formant. These different attributes of the glottal waveform can be measured directly from the speech spectrum or waveform. A set of acoustic parameters that are likely to indicate glottal characteristics is described. These parameters are measured in the speech of a group of female speakers, and the glottal configurations of the speakers are hypothesized. This research contributes to the description of normal variations of voicing characteristics across speakers and to a continuing effort to improve the analysis and synthesis of female speech. It may also have applications in clinical settings. © *1997 Acoustical Society of America.* [S0001-4966(97)03001-4]

PACS numbers: 43.70.Gr, 43.70.Aj, 43.72.Ar [AL]

INTRODUCTION

Certain acoustic characteristics of speech give a voice its quality and individuality, and are the means by which listeners identify or distinguish speakers. These characteristics are complex, having contributions that range from those of the speech production mechanism, that is, differences between individual sound sources and the natural frequencies of the vocal tract, to the effects of prosody and dialect. The work reported in this paper focuses on individual variations in the glottal source waveform, particularly those associated with degree of glottal adduction. These variations are of interest for speech applications such as speaker verification and identification, speech recognition, and speech synthesis by computer, and also have applications for speech disorders and therapy, and in the field of forensics.

The values of the parameters that describe the glottal waveform can vary depending on the glottal configuration, and it is expected that these variations may lead to different voice qualities and intensities. Some voice qualities are usually associated with disordered voice, such as harshness, but our main concern for this paper are those that occur for voices that are not considered to be disordered. Voice qualities that occur frequently in normal speech are pressed, modal, and breathy, described in Laver (1980) and Titze (1995). There are other voice qualities that may occur in normal conversational speech (Laver, 1980), but the focus of this paper is restricted to a few different glottal vibratory patterns (to be described in Sec. I) that may be associated with pressed, modal, and breathy voice qualities. These patterns may be considered regions along one or more continua; we consider degree of glottal adduction, which is greatest for pressed voice, least for breathy voice, and somewhere in between for modal voice.

Direct examination of the vocal folds during normal phonation has revealed that female speakers are more likely than male speakers to have incomplete closure of the vocal folds (Hertegård *et al.*, 1992; Linville, 1992; Peppard *et al.*, 1988; Södersten and Lindestad, 1990; Södersten *et al.*, 1991). The degree of incomplete closure can vary, and posterior openings of the glottis may extend beyond the vocal processes to the membranous part of the folds. The implications of incomplete closure for the glottal waveform are that there is an airflow bypass even during the so-called closed phase of the glottal vibratory cycle, and that an abrupt discontinuity of the airflow derivative is not possible due to the mass of air in this pathway. As we will see later in this paper, the effects of this bypass on the glottal waveform parameters increase with its size, providing one source of variability in voicing characteristics.

The aim of our work has been to formulate a set of acoustic parameters that reflect individual and gender differences in voice quality and glottal configuration. These parameters are derived from measurements of the acoustic spectra of vowels produced by speakers. In this paper we provide the theoretical background for these parameters, and present acoustic data for female speakers. Physiological and perceptual data that support the acoustic data, data for male speakers, and comparison of data across gender will be described in future papers. The decision to initially study the male and female data separately was influenced by the gender differences in glottal configuration described above.

A. Related work

Previous studies of individual variations in glottal characteristics have been based on inverse filtering, visual in-

[a)]Electronic mail: hanson@speech.mit.edu

spections of the vocal folds, acoustic measures made directly on the speech spectrum or waveform, and perceptions of voice quality, particularly breathiness.

Holmberg and her colleagues have made aerodynamic measures of male and female voice (Holmberg *et al.*, 1988, 1989, 1994a, 1994b, 1995; Perkell *et al.*, 1994), and extracted glottal parameters directly from the glottal waveform. These studies included relatively large groups of subjects (from 15 to 45) phonating in different speech conditions. The main findings are the existence of parameter differences between females and males, and across speech conditions. Their studies have also included acoustic measures made on the speech spectrum, and they related these spectral measures to the aerodynamic results (Holmberg *et al.*, 1995), finding that adduction quotient, as measured from the glottal waveform, has a strong relationship with the relative amplitudes of the first two harmonics. In addition, the speed of glottal closure is in some cases reflected by the relative amplitudes of the first and third formants.

Karlsson (1986, 1988, 1989, 1990, 1991a, 1991b, 1992a, 1992b) also studied the glottal waveform obtained by inverse filtering, focusing on a small group of female speakers. Glottal parameters were derived by fitting a theoretical model to the resulting glottal waveform, and varied widely across speakers (Karlsson, 1988). She also attempted to correlate the glottal parameters with the voice qualities of her subjects, as judged by speech therapists. Different voice qualities were separated by the degree of spectral tilt of the voice source, and the presence of noise excitation at mid to high frequencies (Karlsson, 1988, 1992a). Speakers perceived to be breathy had higher minimum flows, steeper tilts, and more aspiration noise (Karlsson, 1988, 1992b). Voices with a tight, strained quality had weaker lower harmonics (Karlsson, 1992b).

Södersten and her colleagues observed glottal closure via fiberscopy and related the degree of closure to perceptions of breathiness. They also used an acoustic measure, the amplitude of the first harmonic relative to the amplitude of the first formant peak. Their results show that female speakers have a higher degree of incomplete closure and perceived breathiness than male speakers (Södersten and Lindestad, 1990), and that significant correlations exist between perceived breathiness and the relative amplitude of the first harmonic (Södersten *et al.*, 1991).

Gobl (1989) examined voice quality correlates by extracting glottal parameters from the inverse-filtered waveform and making measurements on the glottal spectrum. The acoustic measures were the average of the harmonic amplitudes in four frequency bands. Breathy voice was found to have a steeper spectral tilt than modal voice. Gobl and Ní Chasaide (1988) and Ní Chasaide and Gobl (1993) extracted glottal parameters both from the glottal waveform and from vowel spectra. They found that as glottal abduction increased, so did the downward spectral slope of the vowel spectrum, and that formant amplitudes, especially that of $F1$, decreased as well.

Klatt and Klatt (1990) studied variations in voice quality, from pressed to breathy, for both male and female speakers, using reiterant and synthesized speech. Through perception tests, they too found that female speakers were perceived to be breathier than male speakers. Unlike the work discussed above, they relied entirely on acoustic measures made directly on the speech spectrum and waveform. The relative amplitude of the first harmonic was higher for female speakers, who also showed more evidence of aspiration noise at high frequencies. Relevant cues to perception of breathy voice were increases in the amplitude of the first harmonic, aspiration noise, and lower formant bandwidths, with aspiration noise as the most important cue.

There have been other attempts to derive glottal parameters based on measurements of the spectrum of the glottal waveform, or of the speech waveform and spectrum. In particular, Fant (Fant, 1979, 1993; Fant *et al.*, 1985, 1994; Fant and Lin, 1988). Ananthapadmanabha (1984), and Gauffin and Sundberg (1989) have pioneered in developing these techniques.

B. Summary

Several researchers have studied variations in glottal vibratory patterns using inverse filtering or fiberscopic and endoscopic examination of the vocal folds. However, inverse filtering has several problems associated with it (Holmberg *et al.*, 1995), and examination of the vocal folds is necessarily invasive. Thus, there is a need to develop methods of measuring glottal characteristics directly from the acoustic sound pressure which do not require special equipment to record data. In the work to be reported here, we also estimate the spectrum of the glottal source, but rather than filtering out the effects of the vocal tract filter, we take them into account in our analysis. Besides avoiding the problems of inverse filtering, we are able to examine the effect of the glottal source on the filter, especially its bandwidths, thus gaining further information about the glottal configuration.

The contributions of this work are several. First, it adds to research efforts aimed at finding quantitative measures that describe dimensions along which normal voices vary across speakers. It also contributes to the analysis, synthesis, and modeling of female voice and speech. In addition, the work may have clinical applications.

We begin the next section by reviewing the theoretical basis for measurements made on the speech spectrum and waveform. As a result of this theoretical development, several measures of glottal characteristics are suggested. Acoustic data for 22 female speakers are then given, and we attempt to interpret these data in terms of the theoretical models and to classify individual differences based entirely on the inferences derived from the measurements of sound pressure.

I. THEORETICAL BACKGROUND

In this section we discuss several ways in which the configuration of the vocal folds and glottis may vary during vowel production. Specifically, we consider four types of configurations: (1) the arytenoids are approximated and the membranous part of the folds close abruptly; (2) the arytenoids are approximated, but the membranous folds close nonsimultaneously along the length of the folds; (3)

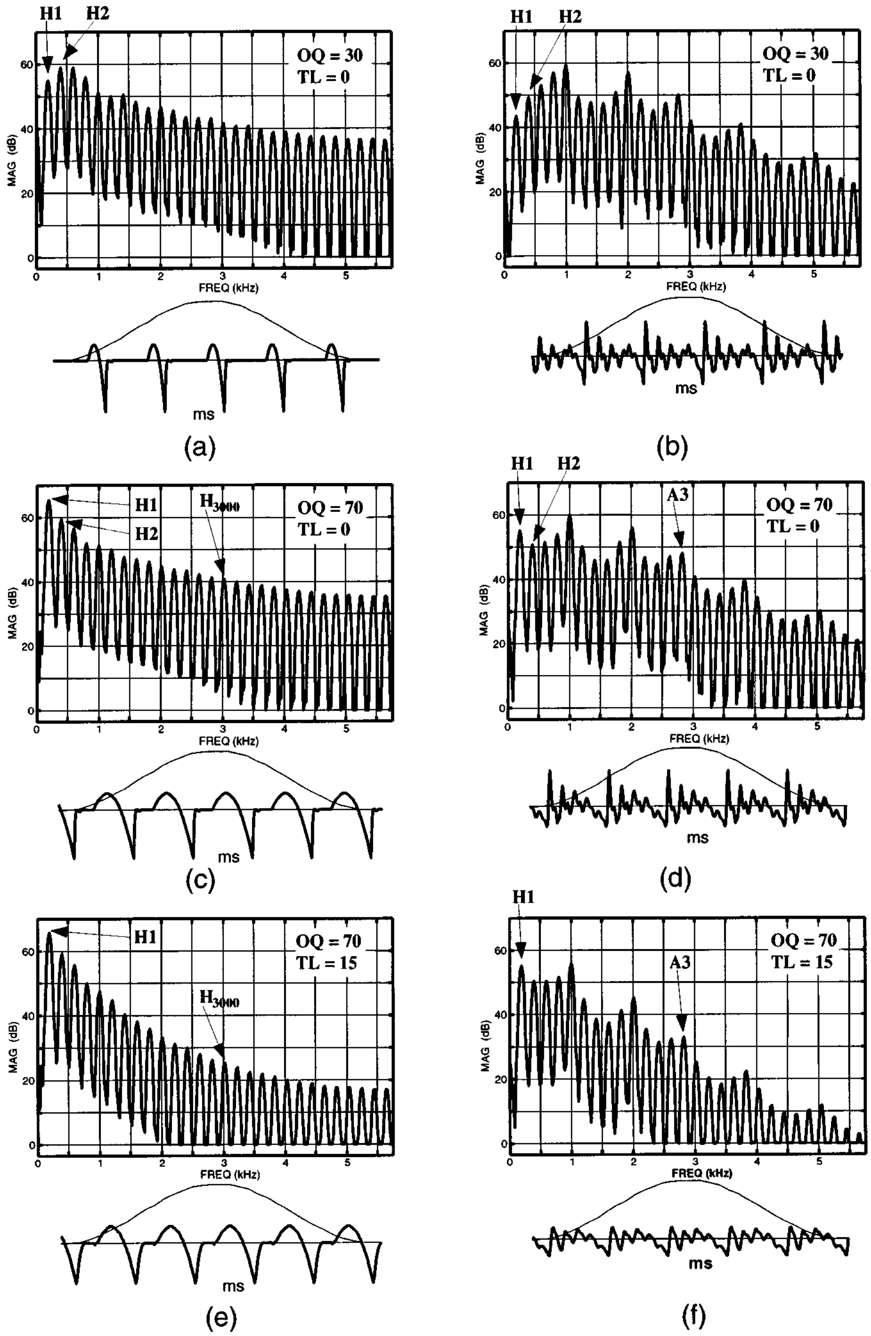

FIG. 1. Waveforms and spectra of the periodic glottal volume-velocity source corresponding to various manipulations of the glottis. The fundamental frequency is in the range for an adult female speaker. Panels (a), (c), and (e) show spectra and derivatives of the volume-velocity sources, while panels (b), (d), and (f) show the spectra of the vowel /æ/ synthesized using those volume-velocity sources. (a)–(b): Open quotient (OQ) is 30%, additional spectral tilt (TL) is zero; (c)–(d): OQ is 70%, TL is zero; (e)–(f): OQ is 70%, TL is 15 dB (i.e., spectrum is 15 dB lower at 3 kHz). The waveforms were generated with the KLSYN88 synthesizer (Klatt and Klatt, 1990).

there is a posterior glottal opening, or ''chink,'' at the arytenoids that persists throughout the glottal cycle, but the folds close abruptly; (4) a posterior opening extends into the membranous portion of the folds throughout the glottal cycle, forcing the folds to close nonsimultaneously. Through a combination of observation and modeling, we suggest several ways in which these various configurations affect the glottal airflow and are manifested in the speech spectrum or waveform. Note that there may be other glottal configurations in addition to the four that we consider.

A. Complete glottal closure during a vibratory cycle

For the case in which the glottis closes completely during a part of the glottal cycle, the glottal waveform can show several kinds of differences from one individual to another. For example, if a speaker modifies her production such that it results in a glottal waveform with a larger open quotient but the same rate of decrease of volume velocity at closure, the spectrum of the source undergoes a change only at low frequencies, with essentially no change in the spectrum amplitude at high frequencies (Klatt and Klatt, 1990). Figure 1(a) and (c) shows the derivatives of the volume-velocity waveform[1] and the spectra of these derivatives for two synthesized waveforms[2] having different values of open quotient (OQ) (30% and 70%, respectively). When open quotient varies from 30% to 70%, the amplitude of the first harmonic relative to that of the second $(H1-H2)$ changes by about 10 dB. Figure 1(b) and (d) shows spectra for the vowel /æ/ synthesized using these glottal waveforms. The difference between the values of $H1-H2$ that were observed in the glottal spectra is also evident in the spectra of the synthesized vowels. Other researchers have used $H1-H2$ as an indication of open or adduction quotient (see, for example, Holmberg *et al.*, 1995).

The spectrum of the derivative of the glottal waveform at middle and high frequencies, when the derivative has a discontinuity at the time of closing, has a downward slope of 6 dB/octave. This spectrum is influenced by the abruptness with which the flow is cut off when the membranous part of the vocal folds closes during the vibration cycle. This abruptness can be affected in two ways, for a given open quotient, when there is complete closure of the glottis during some part of the vibratory cycle (Fant *et al.*, 1985; Klatt and Klatt, 1990). One mechanism that leads to a change in abruptness is a glottal closing that does not occur simultaneously at all points along the anterior–posterior length of the vocal folds. Closing is a type of ''zipper'' action, with initial closure at the anterior end of the glottis and the closure sliding back along the length of the glottis (cf. Ananthapadmanabha, 1993). This type of closure leads to a more gradual cutoff of flow, resulting in a derivative of the glottal waveform that does not have a discontinuity. As illustrated with synthesized glottal waveforms in Fig. 1(e), the effect on the spectrum is to introduce an additional downward tilt, or slope, at high frequencies. If we define T_D as the time from initiation of the anterior closure to the time of closure at the posterior end, and if we approximate the gradual cutoff as an exponential, then a reasonable approximation to the time constant T of this exponential is roughly one-half of the time of the sliding closure,[3] that is,

$$T \approx \frac{T_D}{2}. \tag{1}$$

The breakpoint for the change in spectral slope is then given by

$$f_T = \frac{1}{2\pi T} = \frac{1}{\pi T_D}. \tag{2}$$

Above this frequency, the slope of the spectrum of the derivative of the waveform increases to 12 dB/octave if an exponential approximation is assumed (see, for example, Siebert, 1986). For f_T less than about 2000 Hz, the resulting increase in the tilt at 2750 Hz, an average location of $F3$ for female speakers, is

$$20 \log_{10} \frac{2750}{f_T}. \tag{3}$$

For example, if T_D is 0.5 ms, f_T is 637 Hz, and the increase in tilt at 2750 Hz is 13 dB. Likewise, if T_D is 1.0 ms, the increase in tilt at 2750 Hz is 19 dB. The glottal waveform of Fig. 1(e) is synthesized such that the amplitude of the source spectrum at 3000 Hz (H_{3000}) is 15 dB lower than the spectra in Fig. 1(a) and (c). The spectrum of a vowel synthesized using the glottal waveform of Fig. 1(e) is shown in Fig. 1(f). Note that the amplitude of the third formant $(A3)$ drops by about 15 dB compared with the spectrum in Fig. 1(d). Thus, the amplitude of the third formant relative to that of the first harmonic $(H1-A3)$ appears to be a reasonably accurate indication of source spectral tilt, except if $H1$ is weak, as in Fig. 1(b). Holmberg *et al.* (1995) have also used this measurement as an indication of how abruptly airflow is cut off.

Another way in which the abruptness at closure can be influenced is by manipulating the rate of decrease of flow at the instant of closure. For a given open quotient, this rate of decrease is related to the amount of skewness of the glottal pulse, as shown with synthesized glottal waveforms in Fig. 2(a) and (b). As the slope of the closing phase becomes faster relative to the slope of the opening phase, the spectrum amplitude at middle and high frequencies increases relative to the amplitude at low frequencies. In this example, the difference between the amplitudes of the first harmonic $(H1)$ and the harmonic at 3000 Hz (H_{3000}) increases by about 10 dB with a change in speed quotient from 140% to 320%. Speech spectra corresponding to these glottal waveforms are shown in Fig. 2(c) and (d), and again, the difference $H1-A3$ for the two spectra provides an indication of tilt.

The amplitude of the third formant is also influenced by other factors, one being the locations of $F1$ and $F2$. Another is that the bandwidth of $F3$ is affected by the radiation characteristic to a greater extent than are the lower formants, and the degree of this influence varies with the configuration of the vocal tract for the vowel. Corrections must be made for these effects if values of $A3$ are to be compared across vowels or speakers. Finally, the value of $A3$ will vary depending on whether or not $F3$ is centered on a harmonic. A minimum value of $H1-A3$ that can be expected for a neutral vowel with equally spaced formants when there is no additional spectral tilt due to nonabrupt airflow cutoff has been estimated to be about 9 dB for female speakers (Hanson, 1995), based on calculations made with the KLGLOTT88 source model (Klatt and Klatt, 1990).

B. Incomplete glottal closure during a vibration cycle

Glottal configurations exhibiting an airflow bypass can modify the spectrum of the glottal waveform and the transfer function of the vocal tract relative to those that would exist for the configuration in which the entire glottis is closed over part of the cycle of vibration. In this section we show that among the modifications introduced are: (1) an increase in the bandwidth of the first (and possibly the second) formant; (2) an increase in the tilt of the glottal spectrum at high frequencies; and (3) emergence of a turbulence noise source in the vicinity of the glottis that may be comparable in amplitude (at high frequencies) to the spectrum amplitude of the periodic source.

Cranen and Schroeter (1995) have also studied the acoustic consequences of glottal openings. Although their model exhibits an increase in source spectral tilt only when

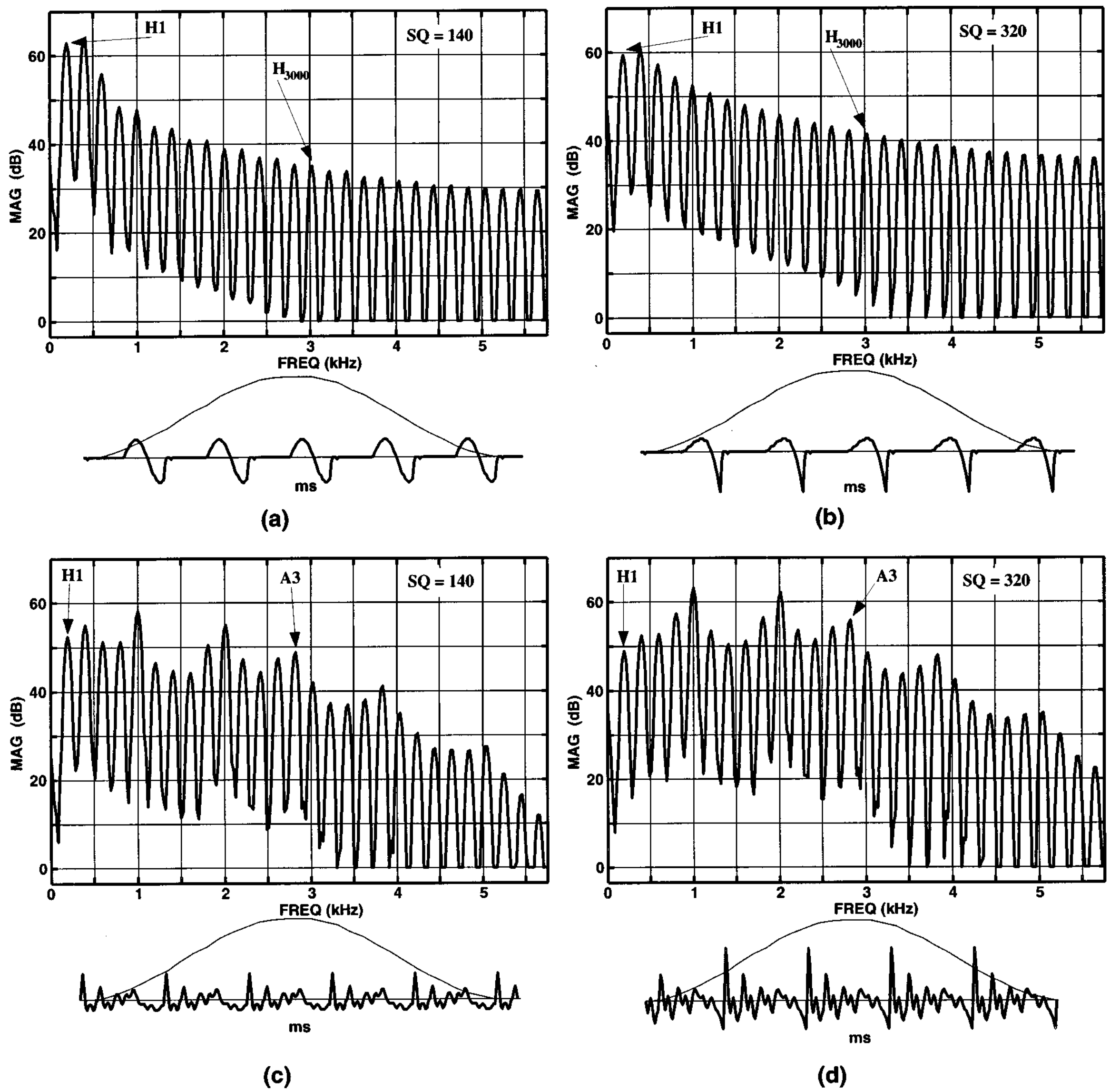

FIG. 2. Waveforms and corresponding spectra of the derivative of the periodic glottal volume-velocity source with different speed quotients (SQ), the speed of the closing phase relative to that of the opening phase. (a) Skewing of waveform decreased to give a SQ of 140%; (b) SQ is 320%; (c) The vowel /æ/ synthesized with the glottal waveform of (a); (d) The vowel /æ/ synthesized with the glottal waveform of (b).

the vocal processes are abducted, we show that all of the above-mentioned acoustic manifestations occur whether the bypass is a glottal chink or an abduction extending beyond the vocal processes. However, the second of these properties, an increase in tilt, is expected to be more marked when there is also abduction of the arytenoids at the vocal processes.

1. Effect on first-formant bandwidth

Formant bandwidths are related to the rate of acoustic energy loss in the vocal tract. The energy losses in the frequency range of the first formant come from several sources, including the resistance of the yielding walls of the vocal tract, and heat conduction and frictional losses at the walls. In earlier work (House and Stevens, 1958; Fant, 1962; Fujimura and Lindqvist, 1971), bandwidths due to vocal-tract losses were measured by exciting a subject's vocal tract while the subject held his or her glottis closed. A first-formant bandwidth of 40–95 Hz was found for female speakers in the closed glottis condition (Fujimura and Lindqvist, 1971; Fant, 1972). When the glottis is open and there is airflow through it, the glottal resistance can contribute further energy loss, particularly at low frequencies, thus adding significantly to the first-formant bandwidth. House and Stevens (1958) also measured bandwidths for the open glottis condition for their male subjects, and found that bandwidth did indeed increase under this condition. In fact, measurement of the $F1$ bandwidth can provide an indirect indication of the degree to which the glottis fails to close completely during a cycle of glottal vibration.

Bandwidths, particularly the first-formant bandwidth ($B1$), can also be estimated from the speech waveform. If the $F1$ oscillation is assumed to be of the form $e^{-\alpha t}\cos 2\pi f t$, that is, a damped sinusoid, where f is the frequency of the first formant, then the constant α (in s^{-1}) is

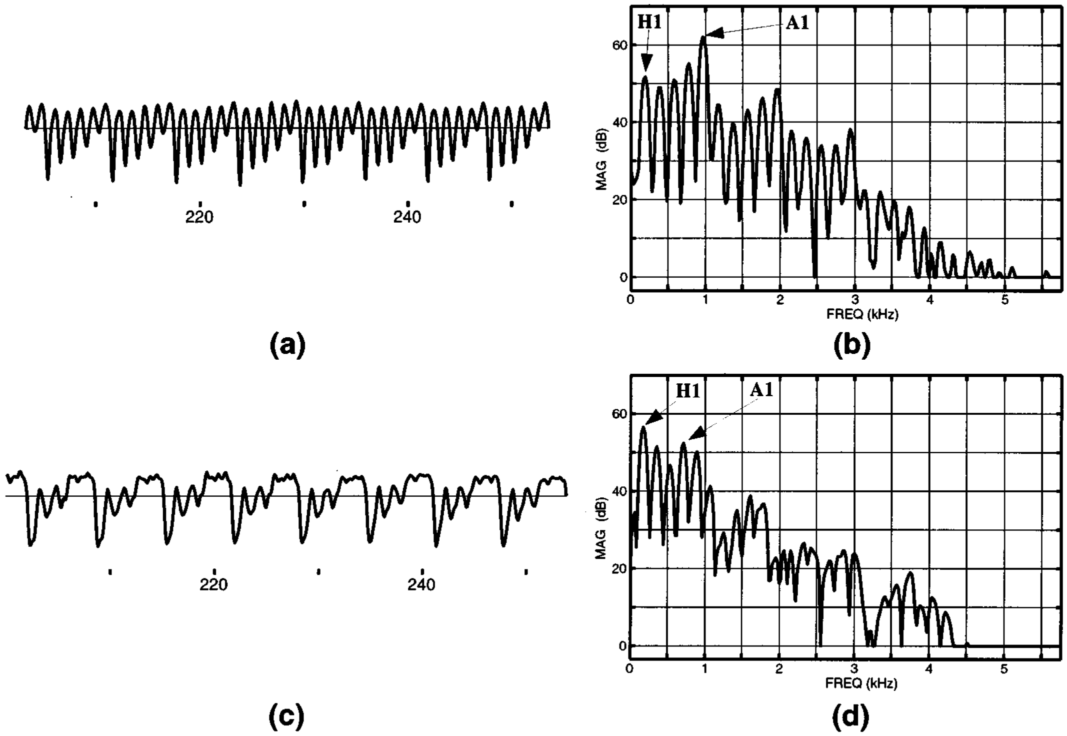

FIG. 3. Examples of waveforms and spectra of the vowel /æ/ produced by two different adult female speakers. The waveforms illustrate decay rates that are (a) slow and (c) rapid, corresponding to narrow and wide first-formant bandwidths, respectively. As estimated from the decay of the speech waveform, the bandwidths during the first part of the cycle are about 60 Hz for (a) and 275 Hz for (c). From the corresponding spectra in (b) and (d), we see that a narrow first-formant bandwidth results in a stronger, more prominent first-formant peak.

related to the bandwidth $B1$ by the equation $B1=\alpha/\pi$ Hz (see, for example, Siebert, 1986). Then by measuring the decay rate of the first-formant waveform during the early part of the glottal period, where the glottal area is expected to be smallest, one can estimate the first-formant bandwidth. Care must be taken to avoid the initial excitation spike when using this method, although bandpass filtering to isolate the first-formant waveform may alleviate this problem. To obtain an accurate estimate, there must be a high enough first-formant frequency and a long enough pitch period to get at least two oscillations during the closed part of the cycle.

As an example of how $F1$ bandwidth is manifested in the acoustic sound pressure, waveforms of the radiated sound pressure for the vowel /æ/ produced by two different female speakers are shown in Fig. 3(a) and (c). The sound-pressure waveform during the initial part of each glottal period is a damped oscillation, the largest component of which is at the frequency of the first formant. The increased rate of decay of the $F1$ oscillation during the last part of each cycle [particularly in Fig. 3(c)] reflects the increased losses at the glottis, and hence the increased bandwidth, during the open phase (Fant, 1979). If the glottis remains open throughout the cycle of vibration, the decay rate during the first part of the glottal cycle will also be increased relative to that for the closed-glottis condition. For the waveforms in Fig. 3, the first-formant bandwidths during the first part of the cycle, as estimated from the decay rate, are about 60 Hz for (a) and 275 Hz for (c).

Another indication of $F1$ bandwidth is the amplitude of the $F1$ peak in the speech spectrum. As predicted by theory, this amplitude should be proportional to the inverse of the bandwidth. Given the bandwidths estimated for the waveforms in Fig. 3, we might expect a difference in relative amplitudes of the $F1$ peaks in the corresponding spectra to be about 12 dB. From the spectra in Fig. 3(b) and (d), we see that the larger bandwidth in (c) results in a reduced $F1$ peak amplitude, making the peak less prominent relative to the amplitude of the first harmonic. The values of $H1-A1$ for these two spectra are about -10 dB for (b) and 4 dB for (d), resulting in a difference of about 14 dB between the two spectra, close to that predicted. This difference is mainly due to the difference in first-formant amplitude ($A1$), but is also partially due to the variation in the relative value of $H1$ across the two speakers.

This example suggests that the amplitude of $F1$ relative to that of the first harmonic ($H1-A1$) may also be a suitable indication of bandwidth. However, this method gives an average bandwidth over the entire glottal cycle, including those times when the glottis is open. During the open phase both $F1$ and its bandwidth increase. Thus, this method may give a larger value of bandwidth than that estimated from the waveform near the beginning of the glottal cycle. Variation across speakers in the relative amplitude of the first harmonic will also add some uncertainty to this acoustic parameter. In addition, $A1$ will vary depending on whether or not $F1$ is centered on a harmonic. The values expected for $H1-A1$ have been estimated to range from a minimum of about -11 dB to a maximum of 5 dB for female speakers (Hanson, 1995), based on theoretical analysis using the KLGLOTT88 glottal source model (Klatt and Klatt, 1990) and the range of

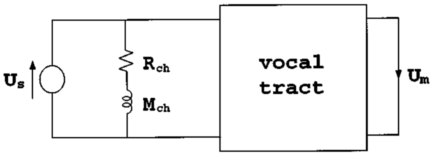

FIG. 4. Model of speech production when the membranous part of the folds have come together, but an opening remains at the arytenoid cartilages, the vocal processes, or both. R_{ch} and M_{ch} represent the resistance and mass of the glottal opening, U_s the volume velocity at the source, and U_m the volume velocity at the mouth.

minimum flows reported by Holmberg *et al.* (1994a).

The estimate of bandwidth based on the decay of the $F1$ oscillations can be used to estimate the area of the glottis during the maximally constricted part of the cycle. Theoretical estimates of the contribution of the glottal opening to the bandwidth $B1$ of the first formant can be made by calculating the value of the resistive termination at the glottis and determining the acoustic energy loss in this resistance (Fant, 1960). An equivalent circuit for calculating the losses is given in Fig. 4. The glottal impedance is represented by an acoustic resistance and an acoustic mass. If we assume that the glottis terminates a uniform vocal tract of length l_v and cross-sectional area A_v, the contribution B_g to the bandwidth of a formant is

$$B_g = \frac{\rho c^2}{\pi A_v l_v R_{ch}(1+4\pi^2 f^2 M_{ch}^2/R_{ch}^2)}, \tag{4}$$

where ρ is the density of air in the vocal tract (g/cm^3), c is the speed of sound (cm/s), f is the formant frequency (Hz), R_{ch} is the glottal resistance due to the chink (dyn s/cm^5), and M_{ch} is the acoustic mass of the glottal chink (g/cm^4) (Stevens, in preparation). From this equation we see that, for a given glottal opening, as f increases, B_g decreases, so the glottal opening has its greatest effect on the bandwidth of $F1$.

Assuming that the pressure drop across the glottis is equal to the subglottal pressure P_s, it has been shown that

$$R_{ch} = \frac{d\Delta P}{dU_{ch}} \approx \frac{\rho U_{ch}}{A_{ch}^2} \tag{5}$$

and

$$U_{ch} = A_{ch}\sqrt{\frac{2P_s}{\rho}}, \tag{6}$$

where U_{ch} is the airflow through the glottal chink and A_{ch} is the area of the chink (Fant, 1960). M_{ch} can be expressed as

$$M_{ch} = \frac{\rho l_g}{A_{ch}}, \tag{7}$$

where l_g is the vertical thickness of the glottis (Fant, 1960). Thus, for a given subglottal pressure and thickness of the glottis, we can calculate B_g and U_{ch} as functions of A_{ch}, using Eqs. (4)–(7). Table I lists a range of A_{ch} values and the corresponding values of B_g, $B1$, and U_{ch}, where $B1 = B_g + B_v$, B_v being the $F1$ bandwidth due to vocal-tract losses (with a closed glottis). For /æ/, B_v is approximately 50 Hz for female speakers of Swedish (Fujimura and Lindqvist, 1971; Fant, 1972). From this table we see that bandwidth increments B_g up to 200 Hz might be expected for glottal openings in the range up to 8 mm^2, when the subglottal pressure P_s is assumed to be 5500 dyn/cm^2. This minimum opening corresponds to a minimum flow of about 249 cm^3/s, which is about the upper limit observed by Holmberg *et al.* (1994a) for 15 female speakers of American English.

TABLE I. Range of glottal chink areas (A_{ch}) and corresponding estimations of: glottal contribution to first formant (B_g); bandwidth of first formant ($B1$); flow through chink (U_{ch}); time constant (T) of the rate of change of flow near closure; and resulting increase in spectral tilt at 2750 Hz. We assume a subglottal pressure of 5500 dynes/cm^2, and vocal tract losses of 50 Hz. See text for equations for B_g, U_{ch}, and T.

A_{ch} (cm^2)	B_g (Hz)	$B1$ (Hz)	20 $\log_{10} B1$ (dB)	U_{ch} (cm^3/s)	T (ms)	Tilt (dB)
0.00	0	50	34	0	0	0
0.01	25	75	38	31	0.13	7
0.02	50	100	40	62	0.16	9
0.03	76	126	42	93	0.20	11
0.04	101	151	44	124	0.23	12
0.05	126	176	45	155	0.27	13
0.06	151	201	46	186	0.30	14
0.07	176	226	47	217	0.33	15
0.08	202	252	48	249	0.37	16
0.09	227	277	49	280	0.40	17
0.10	252	302	50	311	0.43	18

2. Effect on spectral tilt

When there is a glottal chink with the arytenoid cartilages approximated at the vocal processes, the pattern of mechanical vibration of the vocal folds should be approximately the same as it is when there is no glottal chink. The shape of the airflow waveform will, however, be influenced by the bypass through the interarytenoid space, particularly at the time when the vocal folds come together. Although there may be a discontinuity in the rate of closure of the glottis, the acoustic mass of the airway and the presence of the bypass path prevent this discontinuous change from being present in the modulated portion of the glottal airflow.

The glottal flow in the vicinity of the time of closure when there is a glottal chink can be approximated by the response of the circuit model in Fig. 5. When the switch is closed, the circuit represents the case where the glottis is

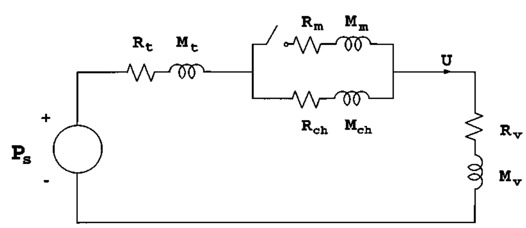

FIG. 5. Model of glottal flow when a fixed opening remains at the arytenoid cartilages during the ''closed'' phase of the glottal cycle. When the switch is closed the situation just prior to closure is modeled. At closure, the switch opens, and the rate of change of flow is limited by the time constant of the circuit. In this figure, P_s represents the subglottal pressure and U the glottal flow. R_t and M_t are the acoustic resistance and mass of the trachea, while R_v and M_v are those of the vocal tract. R_m and M_m are the acoustic resistance and mass of the opening at the membranous part of the vocal folds, and R_{ch} and M_{ch} are those of the glottal chink.

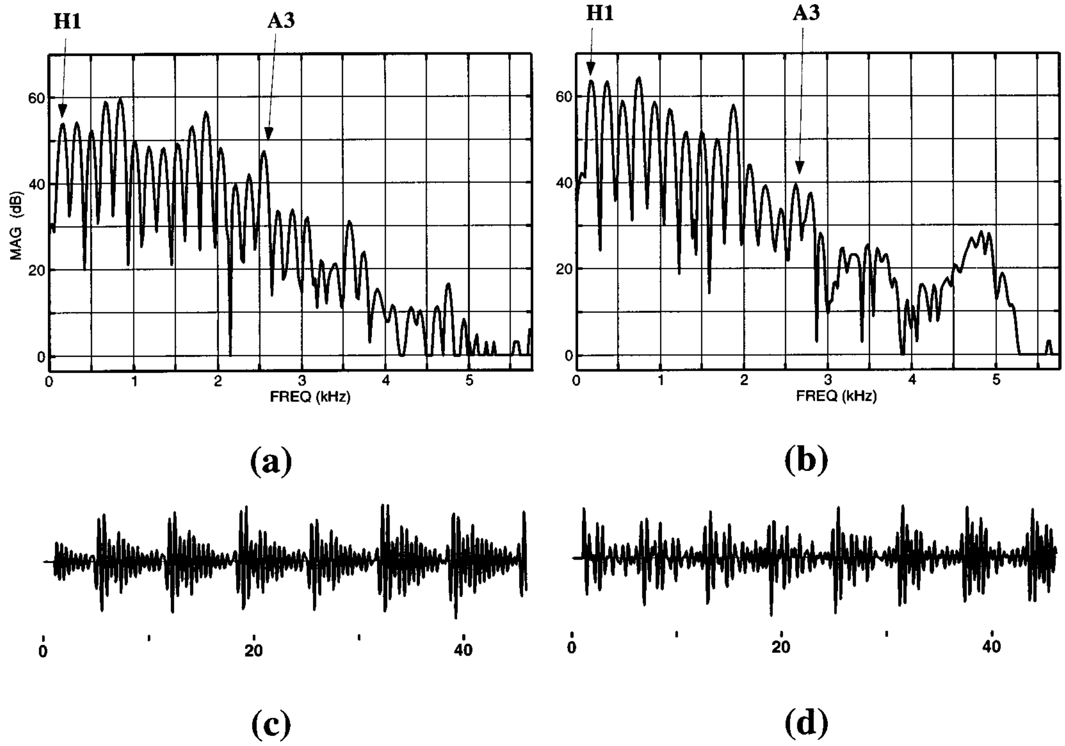

FIG. 6. (a)–(b) Spectra of the vowel /æ/ produced by two adult female speakers with different amounts of spectral tilt. Time window for calculating spectrum is 22.3 ms; (c)–(d) Bandpass-filtered third-formant waveforms corresponding to the spectra in (a)–(b). The center frequency of the filter was 2525 and 2650 Hz for (c) and (d), respectively, and the bandwidth was 600 Hz for both (c) and (d).

open at the folds, and there is flow in both branches of the circuit, giving a total flow U. The closure of the folds is modeled by the sudden opening of the switch. Thus, the rate of change of flow U at the instant of closure is limited by the time constant $T=M/R_{\mathrm{ch}}$, M representing the acoustic mass of the air in the trachea (M_t), glottal chink (M_c), and vocal tract (M_v),

$$M=M_t+M_c+M_v=\rho\left(\frac{l_t}{A_t}+\frac{l_g}{A_{\mathrm{ch}}}+\frac{l_v}{A_v}\right) \tag{8}$$

(Stevens, in preparation), where l_t is the length of the trachea, A_t is the cross-sectional area of the trachea, l_g is the effective vertical length of the glottis, A_{ch} is the cross-sectional area of the glottal chink, l_v is the length of the vocal tract, and A_v is the cross-sectional area of the vocal tract. R_{ch} is the acoustic resistance of the glottal chink, and we have assumed that the acoustic resistance of the trachea and the vocal tract are negligible in comparison to R_{ch}. The length and cross-sectional area of the trachea are about 11 cm and 2 cm^2, respectively, for females (Zemlin, 1988), and l_g is about 0.3 cm (based on data from Titze, 1989a, 1989b, and taking into account end effects). If we assume a vocal tract in a neutral setting, with a length l_v of 15 cm and cross-sectional area A_v of 3 cm^2, then

$$M\approx\rho\left(10.5+\frac{0.3}{A_{\mathrm{ch}}}\right). \tag{9}$$

Using Eqs. (5) and (6), the time constant is then

$$T=\frac{M}{R_{\mathrm{ch}}}=\sqrt{\frac{\rho}{2P_s}}\,(10.5A_{\mathrm{ch}}+0.3). \tag{10}$$

This time constant leads to an additional 6 dB/octave tilt in the spectrum at high frequencies (cf. Siebert, 1986), with the extra tilt beginning at a frequency $f_T=1/2\pi T$. This breakpoint can be translated into a measure of the number of decibels reduction in spectrum amplitude at 2750 Hz, which is approximately the frequency of the third formant for a female speaker. Table I summarizes some time constants T and the corresponding increases in spectral tilt that might be expected for a range of glottal chink areas. Based on minimum airflows measured from inverse-filtered waveforms (Holmberg *et al.*, 1994a) which have a range up to 256 cm^3/s, the maximum increase in tilt that one should expect due to a glottal chink is about 16 dB.

When the arytenoid cartilages remain abducted at the vocal processes throughout the glottal vibration cycle, the membranous part of the folds does not close abruptly, but rather closes nonsimultaneously along the length of the glottis. As discussed in Sec. I A, this nonabrupt closing can contribute significantly to the spectral tilt at mid to high frequencies, depending on the time it takes for the folds to close. With a glottal configuration that has both a fixed space between the arytenoids and some separation at the vocal processes, the effect on the spectral tilt in the third-formant frequency region could be considerable. Thus, depending on the positioning of the arytenoids, including the vocal processes, during phonation, one might expect variations in the $F3$

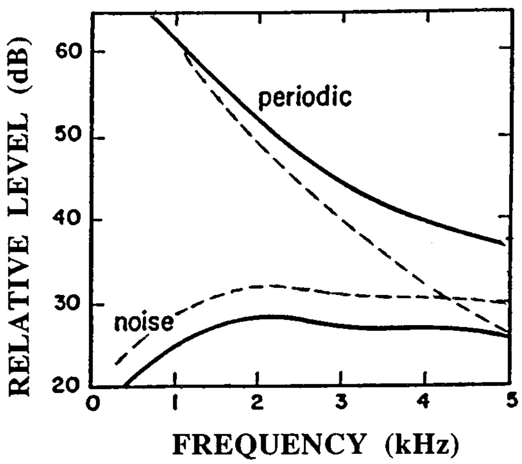

FIG. 7. Calculated spectra and relative amplitudes of periodic volume-velocity source and turbulence-noise source for two different glottal configurations: a configuration in which the glottis has a relatively small minimum opening over one-half of the cycle (solid lines), and a configuration for which the minimum glottal opening has increased (dashed lines). The spectrum for the periodic component gives the amplitudes of the individual harmonics. The noise spectrum is the spectrum amplitude in 50 Hz bands. The calculations are based on theoretical models of glottal vibration and of turbulence noise generation (Stevens, 1993; Shadle, 1985). (From Stevens and Hanson, 1995 and Stevens, in preparation.)

range of the high-frequency spectrum that are substantially greater than 16 dB.

Examples of spectra for the vowel /æ/ produced by two different female speakers are displayed in Fig. 6(a) and (b). These spectra illustrate two extremes of spectral tilt. As discussed in Sec. I A, the amplitude of the first harmonic relative to that of the third-formant peak ($H1-A3$) may be a suitable acoustic correlate of spectral tilt, if certain corrections are made. The values of $H1-A3$ in these two examples are 6 and 23 dB.

3. Turbulence noise at the glottis

Another acoustic consequence of a glottal opening is the generation of turbulence noise in the vicinity of the glottis. It is possible to make estimates of the amplitude and the spectrum of the turbulence noise source at the glottis when the glottal area and the transglottal pressure are known (Shadle, 1985; Stevens, 1993). We can then compare the spectrum of the periodic glottal source to the effective spectrum of the noise source.

When there is vocal-fold vibration with a relatively small glottal opening during half of the cycle, the comparison of the periodic and noise source spectra is shown by the solid lines in Fig. 7. These source spectra are the result of calculations based on theoretical and experimental data from turbulence noise sources and from periodic glottal sources (Shadle, 1985; Stevens, 1993). During phonation both of these sources are filtered by essentially the same vocal-tract transfer function to yield formant prominences. The ratio of the amplitude of the harmonics at 3 kHz to the noise amplitude in a 50-Hz band at the same frequency is 17 dB. Over the entire frequency range up to 5 kHz the noise spectrum is well below the spectrum of the periodic source, so that the combined spectrum is expected to show well-defined harmonics.

When the minimum glottal opening becomes larger, the spectra given by solid lines in Fig. 7 change in two ways. The spectrum amplitude of the periodic component becomes weaker at high frequencies, as noted in Sec. I B 2, and the amplitude of the turbulence noise increases because of the increased flow.[4] For a given subglottal pressure, the amplitude of the turbulence noise source at the glottis is expected to increase approximately in proportion to $A_g^{0.5}$, where A_g is the average glottal area during a cycle of vibration (Stevens, 1971). The two spectra now have the form given as dashed lines in Fig. 7, with the noise spectrum being comparable to the periodic spectrum at high frequencies.

Figure 6(a) and (b) shows the effect of turbulence noise at the glottis in the spectra of natural vowels. The harmonic structure of the spectrum in Fig. 6(b), which has a more extreme tilt, is less apparent at high frequencies (2.5 kHz and above) than that of Fig. 6(a), presumably because of the effect of the aspiration noise. The influence of aspiration noise can also be seen by examining a vowel waveform when it is bandpass filtered at $F3$, with a bandwidth of 600 Hz. The two $F3$ waveforms corresponding to Fig. 6(a) and (b) are displayed in Fig. 6(c) and (d). The waveform in Fig. 6(c), while showing signs of noise excitation, still has a strongly periodic nature. However, the waveform in Fig. 6(d) shows mainly noise, and less evidence of periodic excitation.

Numerous researchers have developed objective measures of the noise present in the speech waveform during glottal vibration (see, for example, Yumoto *et al.*, 1982; Ladefoged and Antoñanzas-Barroso, 1985; Kasuya *et al.*, 1986; Klingholz, 1987; de Krom, 1993; Hillenbrand *et al.*, 1994; Mori *et al.*, 1994). Usually these methods involve isolating the periodic component of the speech waveform from the noisy component. This separation can be done through spectral- or cepstral-based analysis, or through comparison of pitch periods in the time domain, measuring the differences between pitch periods that result from the statistical variability of noise. However, as pointed out by Ladefoged and Antoñanzas-Barroso (1985), these methods do not measure just the noise that is due to an aspiration source, but rather the noise that results from a combination of factors. These other factors include jitter and shimmer. Their solution was to use only *part* of a vibratory cycle and compare it with the corresponding part of the next cycle. However, unless the fundamental frequency is an exact multiple of the sampling period, even a perfectly periodic waveform will appear aperiodic, due to frequency components near the Nyquist frequency that are represented by only a few samples (Klatt and Klatt, 1990). This source of variation can only be remedied by significant oversampling.

Klatt and Klatt (1990) used another technique to estimate aspiration noise, avoiding the inclusion of noise due to other factors. In this method, the amount of aspiration noise in relation to the periodic component is estimated subjectively by examining the bandpass-filtered waveform in the $F3$ region, such as those in Fig. 6(c) and (d). It is also possible for an observer to make estimates of the amount of

TABLE II. Dialect history of the 22 female subjects, age 0–12 years.

	Age	Dialect history
F1	30	Massachusetts (north shore)
F2	22	Rhode Island
F3	27	0–5 North Carolina; 5–6 Brussels; 6–12 Zaire
F4	38	New York; Illinois; Maryland
F5	44	St. Louis, MO
F6	36	Atlanta, GA
F7	25	Virginia; California
F8	23	Massachusetts
F9	32	Schenectady, NY
F10	41	Massachusetts
F11	30	Attleboro, MA
F12	22	0–8 California; 8–12 New Hampshire
F13	31	Southeastern Connecticut
F14	14	0–3 western Missouri; 3–5 eastern Pennsylvania; 5–12 southeastern Iowa
F15	27	Massachusetts
F16	37	0–3 Queens, NY; 3–12 Long Island, NY
F17	30	Massachusetts
F18	38	Buffalo, NY
F19	30	Haverhill, MA
F20	26	0–3 New York, NY; 3–8 Haiti; 8–12 Brockton, MA
F21	37	0–6 Arkansas; 6–7 Berkeley, CA; 7–8 Arkansas; 8–12 Missouri
F22	49	0–3 Syracuse, NY; 4–12 Philadelphia, PA

noise in a spectral representation, such as those of Fig. 6(a) and (b). The observer makes estimates of the amount of noise on a scale from 1 to 4, where 1 means there is essentially no evidence of noise interference and 4 means that there is little evidence of periodicity. Separate estimates are made from the waveform and from the high-frequency part of the spectrum.

II. EXPERIMENTAL DATA

Based on the theoretical discussion of Sec. I, we suggested several measures that can be made directly on the spectra and waveforms of natural vowels and that may give some indication of the vocal-fold and glottal configuration during vowel production. The theory predicts relationships between these measures in some cases, particularly under conditions where the glottis does not close completely during some part of the vibration cycle. For example, we see in Table I that as the area of the glottal chink increases, both the $F1$ bandwidth and the spectral tilt are expected to increase, and we also expect the strength of the noise source to increase. In this section we describe data collected from a group of female speakers, and we attempt to interpret these data in terms of the theoretical models, using the proposed acoustic parameters.

A. Speakers and speech material

We collected recordings of a number of utterances from 22 adult female subjects. Subjects were recruited from the Speech Communication Group at MIT and a group of speech pathology students at Massachusetts General Hospital. The age and dialect history of each subject are listed in Table II. The speakers showed no evidence of voice or hearing problems, and all were native speakers of American English. Eleven speakers had experience being subjects for speech production experiments. The speakers had a wide range of dialects, but nearly half of the group grew up in New England. Subjectively, the vowel qualities produced were quite similar to one another, with a few exceptions for /æ/.

The utterances consisted of three nonhigh vowels, /æ, ɛ, ʌ/, embedded in the carrier phrase ‘‘Say bVd again.’’ These vowels were chosen because the first formant is well separated from the first harmonic, simplifying the acoustic measures to be described below. The recordings were made in a sound-isolated chamber. The subjects were instructed to speak naturally and to put emphasis on the word /bVd/. Each utterance was repeated five times, with the 15 sentences presented in random order during a single session. All the utterances were low-pass filtered at 4.5 kHz, digitized with a sampling rate of 11.4 kHz, and stored for further analysis.

B. Measurements

The acoustic measurements described in Sec. I were extracted from these utterances in the following manner:

First-formant bandwidths. For all repetitions of the vowel /æ/ the first-formant bandwidth during the initial part of the glottal cycle was estimated from the rate of decay of the waveform. Analysis was restricted to /æ/ because its first formant is usually high enough to get at least two oscillations during the closed part of the glottal cycle, and the second formant is well separated from the first. The rate of decay was determined from the change in the peak-to-peak amplitude in the first two cycles of the $F1$ oscillation. Estimates were made for eight consecutive pitch periods in a relatively stable portion of the vowel, generally at the middle. To reduce interference by the second formant, the waveforms were bandpass filtered with a filter having a bandwidth of 600 Hz centered at the first formant frequency. These 40 estimates were then averaged to obtain a mean value for each speaker.

$H1^*-H2^*$. The difference between the amplitudes of the first and second harmonics was measured for all repetitions of all three vowels. For /æ/, $H1-H2$ was measured from the spectrum obtained by centering a 22.3-ms Hamming window during the initial part of the glottal cycle, at the eight points where the $F1$ bandwidth was estimated. For /ʌ/ and /ɛ/, the measurements were taken at three points in midvowel, 20 ms apart, where the formants were relatively stable. Corrections were made for the amounts by which $H1$ and $H2$ are ‘‘boosted’’ by the effect of the first formant on the vocal-tract transfer function,[5] yielding the measure $H1^*-H2^*$. This corrected measure can be compared across vowels and across speakers. The values for each repetition were averaged to obtain a mean value for each vowel for each speaker.

$H1^*-A1$. The difference between the (corrected) amplitude of the first harmonic and the amplitude of the first formant peak ($A1$) was measured. $A1$ was estimated by measuring the amplitude of the strongest harmonic of the $F1$ peak. The measurements were taken at the same points as those for $H1^*-H2^*$, and similarly, average values were computed for the three vowels for each speaker.

$H1^*-A3^*$. The difference between the amplitudes of the first harmonic and the third formant peak ($A3$) was mea-

TABLE III. Average values of the acoustic parameters for the vowel /æ/, 22 female speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, N_w and N_s are the waveform- and spectra-based noise judgments, and $B1$ is the bandwidth of the first formant, given in Hz. Numbers in boldface represent maxima or minima for each measure across speakers. The mean values and average standard deviations across speaker are also given.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	N_s	$B1$
F1	3.5	−0.2	30.7	3.0	2.8	194
F2	1.7	0.4	32.2	2.8	2.9	244
F3	4.4	−8.0	32.1	2.7	2.8	94
F4	1.6	−5.7	**13.0**	1.6	1.6	209
F5	5.4	2.2	**35.0**	**3.8**	2.7	245
F6	2.4	−5.5	23.0	**1.1**	**1.1**	153
F7	3.8	−1.3	31.3	3.1	**3.1**	150
F8	2.1	−3.7	32.6	2.9	2.7	97
F9	2.8	−7.2	16.8	1.2	1.2	104
F10	5.0	**3.9**	26.4	2.2	2.6	184
F11	4.5	−4.4	19.5	1.8	2.1	158
F12	0.7	−5.6	31.3	2.4	2.2	217
F13	3.8	−8.9	19.4	1.7	1.2	**53**
F14	5.2	**−11.3**	16.3	**1.1**	1.2	78
F15	6.2	0.3	33.7	3.1	2.4	256
F16	**6.8**	1.2	30.4	2.3	2.5	132
F17	1.6	−2.6	22.0	2.0	1.8	**280**
F18	4.5	−2.2	21.8	2.0	2.5	163
F19	5.4	−0.5	24.3	2.0	2.0	166
F20	0.9	−6.2	14.7	1.7	1.6	178
F21	0.8	−8.5	17.9	1.5	1.4	124
F22	**0.6**	−9.2	20.8	1.4	1.2	149
Mean	3.4	−4.2	24.1	2.1	2.1	165
Mean s.d.	1.4	2.3	3.4	0.5	0.5	34

TABLE IV. Average values of the acoustic parameters for the vowel /ʌ/, 22 female speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, and N_w and N_s are the waveform- and spectra-based noise judgments. Numbers in boldface represent maxima and minima for each measure across speaker. The mean values and average standard deviations across speaker are also given.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	N_s
F1	4.8	2.8	26.4	3.0	2.8
F2	1.2	−0.3	25.2	2.7	2.9
F3	3.6	−1.7	26.0	2.7	2.7
F4	**−0.7**	−9.0	**10.9**	1.8	1.3
F5	3.7	1.5	29.1	2.3	2.4
F6	3.0	−6.6	18.9	**1.4**	1.2
F7	1.8	−1.0	28.3	3.2	**3.5**
F8	3.0	−2.7	29.2	2.5	2.3
F9	1.5	−6.4	20.6	1.7	1.8
F10	3.1	2.8	24.7	2.4	2.3
F11	3.9	−2.9	22.0	1.7	2.1
F12	2.2	−5.8	22.9	2.2	1.9
F13	2.7	−4.4	15.5	**1.4**	**1.1**
F14	5.1	−11.9	15.1	**1.4**	1.3
F15	3.6	−4.0	27.2	2.9	2.3
F16	**5.8**	**3.5**	24.6	2.0	2.3
F17	1.5	−4.0	22.7	2.4	1.7
F18	3.5	−2.8	18.5	1.7	2.0
F19	5.0	1.3	**34.1**	**3.5**	3.2
F20	−0.2	−9.9	14.9	1.6	1.7
F21	0.1	−6.8	20.5	2.5	1.6
F22	0.3	**−12.1**	14.8	2.1	1.2
Mean	2.6	−4.1	22.0	2.2	2.0
Mean s.d.	1.1	1.7	3.3	0.6	0.6

sured. As was done for $A1$, $A3$ was estimated using the strongest harmonic of the $F3$ peak. $H1$ was corrected as above, and $A3$ was corrected for the effect of $F1$ and $F2$ on the spectrum amplitude of the third formant.[6] For this normalization neutral first and second formant frequencies were set to 555 and 1665 Hz, respectively, based on the average third-formant frequency measured for all speakers. As mentioned in Sec. I A, $A3$ is also dependent on the bandwidth of $F3$. House and Stevens (1958) measured $F3$ bandwidths of male speakers of English for /æ, ʌ, ɛ/ to be 103, 64, and 88 Hz, respectively. In dB this means that /æ/ is expected to have an $F3$ amplitude that is 4 dB less than that of /ʌ/, while that for /ɛ/ is 3 dB less. For females speakers, the bandwidth values will be higher, but because data are not available for these vowels for female speakers, we made corrections based on the male data. This use of male data should result in minimal error because the ratios of the bandwidths are used to compute the difference in dB and these ratios are not expected to differ greatly across gender. Thus, the value of $A3$ measured for each token of /æ/ and /ɛ/ was increased by 4 and 3 dB, respectively. The combination of these two corrections, for the location of $F1$ and $F2$ and for the $F3$ bandwidth, yields a normalized $H1^*-A3^*$.

Noise ratings. All repetitions of the three vowels were bandpass filtered around the third formant using a filter having a bandwidth of 600 Hz. The bandpass-filtered waveforms and the speech spectra corresponding to the speech segments used in the previously described measures were given ratings for noise, as described in Sec. I B 3 and in Klatt and Klatt (1990). These judgments were made independently by two judges, who did not know which waveforms or spectra corresponded to which speaker. Their average ratings were highly correlated ($r>0.92$) and were averaged to obtain two noise judgments for each speaker, one based on the waveforms and the other on the spectra. The waveform-based ratings were found to be well correlated with the spectrum-based ratings.

C. Results

1. Mean values

The mean values of the acoustic parameters for each speaker are summarized in Tables III–V. Minimum and maximum values for each measure across speakers are given in boldface in these tables. $H1^*-H2^*$ has a range of about 10 dB. $H1^*-A3^*$ has a range of about 26 dB, indicating a wide variation in spectral tilt among the subjects. This large range of spectral tilt is assumed to be a consequence of the presence of a glottal chink or nonsimultaneous closure along the length of the glottis, or both, for some speakers. The minimum value of tilt is 8.6 dB, about what might be expected for the case where there is complete, abrupt glottal closure during some part of the glottal cycle (see Sec. I A). The range of $H1^*-A1$ is 16 dB, as predicted earlier, and the minimum and maximum values are very close to those predicted in Sec. I B 1: −11 and 5 dB. The range of values obtained suggests that first-formant peaks vary from being very prominent for some speakers to being highly damped for others, although part of this range can be due to variation

TABLE V. Average values of the acoustic parameters for the vowel /ɛ/, 22 female speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, and N_w and N_s are the waveform- and spectra-based noise judgments. Numbers in boldface represent maxima and minima for each measure across speakers. Average values and average standard deviations across speaker are also given.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	N_s
F1	6.3	1.7	28.8	3.2	2.9
F2	1.3	−2.0	27.4	2.8	2.1
F3	3.5	−3.1	31.9	3.2	3.1
F4	0.9	−11.0	**8.6**	1.7	1.1
F5	5.4	**3.7**	30.6	3.2	3.0
F6	3.3	−9.0	17.3	1.4	**1.0**
F7	3.1	−2.5	27.3	**3.6**	**3.3**
F8	2.6	−3.8	29.8	2.4	2.2
F9	3.0	−4.3	19.9	2.2	1.5
F10	6.5	2.5	22.6	2.7	2.5
F11	4.6	−5.8	18.0	1.8	1.7
F12	1.9	−5.7	26.0	2.1	1.6
F13	3.0	−5.3	16.0	**1.6**	1.1
F14	4.0	**−12.4**	16.6	1.9	1.2
F15	4.0	−1.1	30.2	2.5	1.9
F16	**6.9**	−1.6	29.4	2.9	1.9
F17	2.4	−5.3	27.1	2.7	2.3
F18	4.2	−3.7	16.5	**1.6**	1.6
F19	5.1	−3.9	**32.8**	2.8	2.5
F20	−0.8	−10.3	13.7	1.9	1.4
F21	1.5	−5.5	20.4	1.9	1.2
F22	**−2.6**	−6.7	15.5	1.7	1.2
Mean	3.1	−4.7	22.5	2.3	1.9
Mean s.d.	1.3	1.9	3.4	0.5	0.8

in the amplitude of $H1$ and how well $F1$ is centered on a harmonic across speakers. This range of first-formant amplitudes presumably arises in part due to a range of $F1$ bandwidths and in part due to differences in the degree to which spectral tilt extends to the low-frequency harmonics.

The first-formant bandwidth estimates for /æ/ vary from 53 to 280 Hz. For the speaker with the lowest value of bandwidth (53 Hz), this estimate is about what is expected for the closed-glottis condition (Fant, 1972). For speakers with higher values of bandwidth, losses must exist at the glottis. Theoretical analysis of glottal losses indicates that a first-formant bandwidth of 280 Hz corresponds to a minimum glottal opening of about 0.09 cm^2 (see Table I), while 75 Hz corresponds to about 0.01 cm^2, so we have a range of estimated glottal chink cross-sectional areas of about 0.08 cm^2. The noise judgments range from 1.0 to 3.8; that is, some of our speakers show little to no noise in the high-frequency range, while other speakers have substantial noise.

Because there were only five tokens of each vowel per speaker, it is difficult to judge the consistency of the speakers. Standard deviations of each measure were computed for each speaker, and the average value across speakers is given in the last lines of Tables III–V. We see that the speakers were generally quite consistent, and there are only small differences in consistency across vowel. Given the range of values observed for each acoustic parameter, the variations within a speaker do not seem significant.

2. Statistical analysis

All acoustic parameters (except $B1$) were subjected to repeated measures analyses of variance (ANOVA) with vowel (/æ/ vs /ʌ/ vs /ɛ/) as a within-subject factor. The results are summarized in Table VI. There was a significant effect for the parameters $H1^*-H2^*$ and $H1^*-A3^*$ [$F(2,42)=4.04$, $p<0.05$, and $F(2,42)=5.25$, $p<0.01$, respectively]. However, referring to Tables III–V and comparing mean values across vowel for these two parameters, we see that, despite the statistical significance, there are only very small differences (1–2 dB). Thus, it would seem that the corrections made to $H1$, $H2$, and $A3$ for vowel quality (see Sec. II B) were successful in minimizing differences across vowels.

TABLE VI. Results of repeated measures analyses of variance (ANOVA) performed to examine differences in the acoustic parameters across vowels. An asterisk (*) indicates statistical significance at the 0.05 level.

Measure	$F(2,42)$	p
$H1^*-H2^*$	4.04	<0.05*
$H1^*-A1$	0.85	>0.1
$H1^*-A3^*$	5.25	<0.01*
Waveform-based noise	1.97	>0.1
Spectra-based noise	2.25	>0.1

Table VII shows Pearson product moment correlation coefficients (r) for the various measures for each vowel. In the following discussion we consider a correlation with $r\geqslant0.70$ to be strong. The strongest correlation was found between the high-frequency noise ratings and the measure $H1^*-A3^*$ ($r\geqslant0.79$, $N=22$). As mentioned earlier, this is not unexpected given that both tilt and noise are expected to increase with the area of a fixed glottal opening (see Table I and the discussion in Secs. I B 2 and I B 3). $H1^*-A1$ tends to have a strong correlation with the spectra-based noise ratings ($r>0.70$, $N=22$ for four out of six correlations). Again, this is predicted from earlier discussion (see Table I where $B1$ increases with A_{ch}). For the vowels /ʌ/ and /ɛ/, $H1^*-A3^*$ is well correlated with $H1^*-A1$ ($r\geqslant0.70$, $N=22$), but the correlation is only moderate for /æ/ ($r=0.62$, $N=22$). Finally, the correlation between $H1^*-A1$ and estimated $F1$ bandwidth for /æ/ is moderate ($r=0.61$, $N=22$).

Although one might expect a larger open quotient to lead to greater losses and noise, due to an increase in average glottal area, the difference $H1^*-H2^*$ is not well correlated with any other measure ($r<0.59$, $N=22$). This result could be interpreted to mean that $H1^*-H2^*$ is not a good indication of open quotient. However, Holmberg *et al.* (1995) found $H1^*-H2^*$ to be well correlated with the adduction quotient in simultaneous observations of airflow and acoustic spectra for female speakers. Assuming that adduction quotient is 100−open quotient, $H1^*-H2^*$ is then also well correlated with open quotient. Therefore, our results may indicate that open quotient is nearly independent of other glottal parameters. For example, a speaker may adjust her glottal configuration in such a way that a larger open quotient results while the rate of decrease of flow at glottal closure remains nearly the same. Thus, $H1^*-H2^*$ would increase while $H1^*-A3^*$ remains the same.

Table VIII shows the Pearson product moment correlation coefficients for the three vowels combined. The noise measures are strongly correlated with the tilt-related measure

TABLE VII. Pearson product moment correlation coefficients (r) for the acoustic parameters for each of the three vowels /æ, ʌ, ɛ/. Numbers in boldface represent strong correlations ($r>0.70$). The notation n.s. indicates that a correlation was not significant. $N=22$.

		$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	N_s	$B1$
/æ/	$H1^*-H2^*$	1					
	$H1^*-A1$	0.47	1				
	$H1^*-A3^*$	n.s.	0.62	1			
	N_w	n.s.	0.67	**0.87**	1		
	N_s	0.38	**0.72**	**0.82**	**0.88**	1	
	$B1$	n.s.	0.61	n.s.	0.45	n.s.	1
/ʌ/	$H1^*-H2^*$	1					
	$H1^*-A1$	0.57	1				
	$H1^*-A3^*$	0.51	**0.78**	1			
	N_w	n.s.	0.56	**0.81**	1		
	N_s	n.s.	**0.75**	**0.84**	**0.83**	1	
/ɛ/	$H1^*-H2^*$	1					
	$H1^*-A1$	0.59	1				
	$H1^*-A3^*$	0.49	**0.70**				
	N_w	0.45	**0.71**	**0.82**	1		
	N_s	0.48	**0.73**	**0.79**	**0.94**	1	

$H1^*-A3^*$, and the spectra-based noise measure is strongly correlated with the bandwidth-related measure $H1^*-A1$. In addition, $H1^*-A1$ has a good correlation with the measure $H1^*-A3^*$.

D. Interpretation of acoustic measurements

In order to gain a better understanding of the correlations reported in Table VIII, and to perhaps be able to interpret the acoustic measurements in terms of glottal configurations, we examined scatterplots of measures that were well correlated with each other.

Figure 8 plots the measure $H1^*-A3^*$ against $H1^*-A1$. Almost all of the data points with $H1^*-A1$ less than about -6 dB have an $H1^*-A3^*$ measure less than about 23 dB, while all of the data points with $H1^*-A1$ greater than about -2 dB have an $H1^*-A3^*$ measure greater than about 23 dB. Note that the highest $H1^*-A3^*$ measure expected for speakers with a posterior glottal opening and simultaneous closure of the membranous part of the folds is about 25 dB (see Sec. I B 2). Based on this observation, we divided the data points into two groups, depending on whether $H1^*-A3^*$ was less than or equal to 23 dB (group 1) or greater than 23 dB (group 2). Analysis of the two groups revealed that for 19 speakers, all three data points fell into either one group or the other, but not both. Data points for the other three speakers (F10, F12, F17) fell into both groups. Because subjects F10 and F12 had only one point each in group 1, they were assigned to group 2. Speaker F17 had only one point in group 2, so she was assigned to group 1.

TABLE VIII. Pearson product moment correlation coefficients (r) for the acoustic parameters for the three vowels /æ, ʌ, ɛ/ combined. Numbers in boldface represent strong correlations ($r>0.70$). $N=66$.

	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	N_s
$H1^*-H2^*$	1				
$H1^*-A1$	0.53	1			
$H1^*-A3^*$	0.46	0.68	1		
N_w	0.30	0.63	**0.80**	1	
N_s	0.40	**0.73**	**0.80**	**0.86**	1

In Fig. 8 data points for group 1 speakers are represented by filled circles and those for group 2 are represented by open circles. The 11 speakers in group 1 have relatively low values of $H1^*-A3^*$ and $H1^*-A1$; that is, speakers in this group have shallow spectral tilts and prominent first-formant peaks. Therefore, this group can be hypothesized to have abrupt glottal closures. Some of these speakers may also have posterior glottal chinks, leading to a range of $H1^*-A3^*$ (about 15 dB) and $H1^*-A1$ (about 11 dB).

Speakers in group 2, indicated by open circles, have much higher values of $H1^*-A3^*$, that is, steeper spectral tilts. Because these values are close to or greater than 25 dB, we surmise that the glottal closure is not simultaneous along

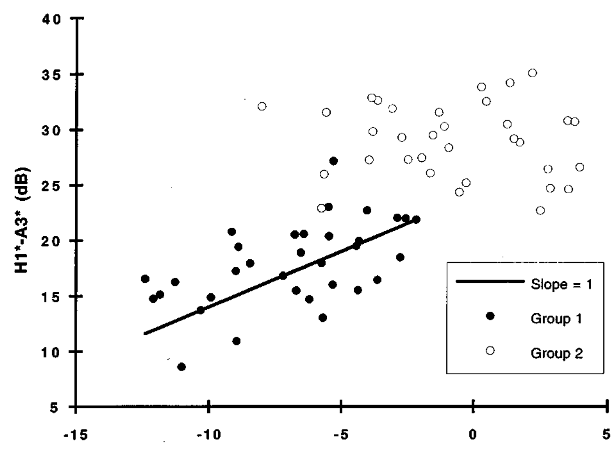

FIG. 8. Relation between $H1^*-A3^*$ and $H1^*-A1$. Each point represents data for one vowel by one speaker. Data points for group 1 are displayed as filled circles and data points for group 2 are displayed as open circles (see text). A line of slope one has been drawn through the data points for group 1, showing the theoretically predicted relationship between spectral tilt and the amplitude of the first formant. (After Stevens and Hanson, 1995.)

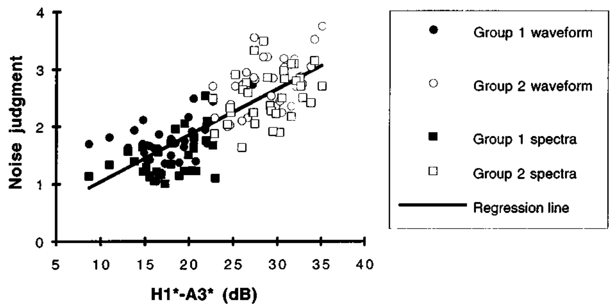

FIG. 9. Relation between noise judgments and $H1^*-A3^*$, together with a regression line ($r^2=0.62$). Points represented as circles are judgments based on waveforms and the squares are based on spectra. Filled points represent group 1 data, while open points represent group 2 data. (After Stevens and Hanson, 1995.)

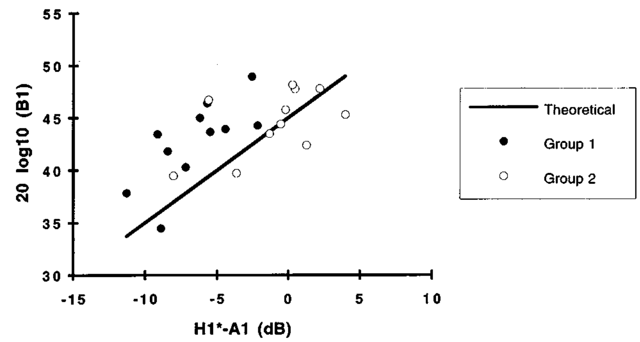

FIG. 10. Relation between $H1^*-A1$ and $F1$ bandwidth (on a log scale) as measured from the waveform. The data are from speakers producing the vowel /æ/. Data points for group 1 members are represented by filled circles, while those for group 2 members are represented by open circles. A straight line representing the slope of the theoretical relationship has been drawn through the data. (After Stevens and Hanson, 1995.)

the length of the membranous part of the vocal folds and that incomplete glottal closure also occurs. This nonsimultaneous closure is probably due to the glottis being abducted at the vocal processes, although the folds could also close nonabruptly when the vocal processes are approximated. The higher values of $H1^*-A1$ for group 2 speakers are due to two influences on $A1$: (1) the first formant has an increased bandwidth because there are greater losses associated with the glottal configuration in which the vocal processes are abducted, and (2) the spectral tilt is so steep that its influence extends down into the first-formant range. $H1^*-A3^*$ does not appear to be correlated with $H1^*-A1$ for group 2 speakers, possibly because the effect of the nonsimultaneous closure is independent of the effect of the posterior glottal opening.

From Table I we see that if the bandwidth of the first formant is expressed on a log (dB) scale, then it should increase with tilt with a slope of 1 for speakers who have abrupt glottal closure. Since $H1^*-A1$ is an indicator of bandwidth and $H1^*-A3^*$ is an indicator of tilt, a similar linear relationship should also exist between these two parameters for speakers with abrupt glottal closure. In Fig. 8 a line with slope 1 has been drawn through the data and is seen to fit nicely with the group 1 points. This result is evidence that group 1 speakers have abrupt glottal closure and posterior glottal openings that range in size across speakers, and that as first-formant bandwidth increases, $H1^*-A1$ increases.

Figure 9 shows the relation between the two types of noise judgments and the tilt parameter $H1^*-A3^*$. Recall that there was a high correlation between these quantities. This figure is also divided into the two groups of speakers of the previous figure. Speakers with greater degrees of tilt show greater amounts of noise in their speech signals, as predicted from the theoretical discussion of Sec. I. A regression line ($r^2=0.62$) has been drawn through the points in Fig. 9.

In Fig. 10 the parameter $H1^*-A1$ is plotted against $F1$ bandwidth (on a log scale) as measured in the first part of the glottal cycle ($B1$) for the 22 speakers producing the vowel /æ/. The data are presented to indicate which points belong to group 1 and group 2 speakers. A line of slope 1 is drawn through the data to represent the relationship expected based on the theoretical development. There is a trend toward a decrease in $F1$ prominence (that is, a decrease in $A1$) as the $F1$ bandwidth increases, but the correlation is only moderate ($r=0.61$, $p<0.01$). The relatively weak correlation may be due to the fact that the prominence of $A1$ depends on the entire glottal cycle, whereas the bandwidth estimate $B1$ is based only on the closed (or minimum glottal area) part of the glottal cycle. Thus, $A1$ is influenced by the open quotient and the glottal aperture during the open phase, but the bandwidth estimate $B1$ is not. In addition, other factors, such as spectral tilt, may reduce $A1$. In fact, given these influences, it is not surprising that the group 1 data in Fig. 10 appear to be better correlated than the group 2 data.

For one speaker (F13) the bandwidth is sufficiently small (53 Hz) that complete glottal closure can be assumed during a portion of the glottal cycle. This speaker is from group 1. For speakers with higher bandwidth and $H1^*-A1$ measures, it is reasonable to assume that the source of loss is an incomplete glottal closure. Two speakers from group 2 (F3 and F8) have fairly narrow bandwidths (94 and 97 Hz), although this would not be expected given our hypothesis that group 2 members have abduction at the vocal processes. The $H1^*-A1$ measure for these speakers indicates that $A1$ is indeed quite prominent, consistent with the narrow bandwidth. The findings for these speakers may indicate that their glottal closure is characterized by adducted vocal processes with no posterior glottal chink, but nonsimultaneous closure within the membranous portion. This interpretation might explain the narrow first-formant bandwidths, high first-formant amplitudes, and steep spectral tilts that these two speakers exhibit (see Tables III–V).

Our results are satisfying in that the ranges of observed values and the relationships between these values are in line with the predictions based on our theoretical development. However, our classification of the subjects by glottal configuration is only an hypothesis. It should be verified by direct observation of the vocal folds during phonation. Preliminary evidence gathered from 4 of the 22 subjects suggests that this hypothesis is correct (Hanson, 1995). These subjects were examined using fiberscopy and two subjects from group 1 were observed to have rather small glottal chinks, while two subjects from group 2 were observed to have relatively large posterior openings. For one of the group 2 speakers, the opening extended well beyond the vocal processes and closure was possibly nonsimultaneous. However, the fiberscopic images were not collected simultaneously with the sound-pressure signals, and the resolution of the images was not high enough to allow us to make judgments regarding the abruptness of closure. Therefore, a study that includes more subjects, and simultaneous recordings of image and sound-pressure data is needed to verify our classification of glottal configuration based on acoustic parameters.

III. FUTURE WORK

The wide ranges of acoustic parameter values that we have observed in this study suggest that consideration of glottal characteristics has great importance for describing female speech and, in addition to formant frequencies and fundamental frequency, should be taken into account for applications such as speech synthesis, speech recognition, and speaker recognition. The work should be extended in several ways.

As mentioned above, simultaneous collection of physiological and sound-pressure signals is necessary to validate the use of the acoustic parameters for classification of subjects according to glottal configuration. In addition, the current study was limited to female speakers without voice disorders. A study based on male subjects is currently under way and will provide comparisons with the female data (Chuang and Hanson, to appear). These acoustic parameters should also be applied to the speech of people with voice disorders.

Another aspect of these acoustic parameters that remains to be studied is the variability for a given speaker across repeated recordings. Speaking intensity has been found to be systematically related to variations in airflow-based glottal parameters (Holmberg *et al.*, 1988), and thus should also influence acoustic parameters. Such intrasubject variation may be significant (Holmberg *et al.*, 1994a; Hanson, 1995). It should also be noted that source characteristics, and hence perhaps voice quality, change constantly throughout an utterance, due to the effects of prosody and coarticulation between vowels and voiceless consonants. The possible use of the acoustic parameters presented here as a tool to examine these changes should be explored.

Finally, the relation of the proposed parameters to perceived voice quality should be examined. In a preliminary study, speakers from group 2 were perceived to be breathier than speakers from group 1, and listeners could perceive changes in the acoustic parameters obtained by manipulating synthesized speech (Hanson, 1995).

ACKNOWLEDGMENTS

This paper is based on a portion of a Ph.D. thesis submitted to the Harvard University Graduate School of Arts and Sciences. The research was carried out at the M.I.T. Research Laboratory of Electronics. I am grateful to Kenneth N. Stevens for his guidance and encouragement, to Eva Holmberg for her careful reading of the revised paper, and to Alice Turk for her assistance with the statistical analysis. The comments and suggestions of Anders Löfqvist, Ronald C. Scherer, and an anonymous reviewer on an earlier version of the paper are greatly appreciated. This work was supported by the Harvard University Division of Engineering and Applied Sciences, a U.S. Dept. of Education fellowship, and NIH Grant No. DC00075.

[1]Due to the radiation characteristic at the mouth, the radiated sound pressure can be approximated as the derivative of the volume velocity at the mouth. Thus, the derivative of the glottal waveform is the effective excitation (Fant, 1982), and may be of more interest for analysis than the glottal airflow itself.

[2]The synthesized waveforms described in this paper were produced using the KLSYN88 synthesizer (Klatt and Klatt, 1990), which contains several glottal sources, including a representation of the source proposed by Fant *et al.* (1985). Our intention in using these waveforms is simply to illustrate how changes in glottal waveform parameters are manifested in both the glottal source and speech spectra, and not to promote a particular model of the glottal waveform.

[3]This model of nonsimultaneous closure is not based on actual data. It should be verified with high-speed image data.

[4]The amplitude of the turbulence noise increases with glottal area only up to a certain point: when the glottal area becomes large enough, a drop in transglottal pressure occurs, the flow velocity decreases, and the amplitude of the turbulence noise also decreases.

[5]The quantity $20\log_{10}[F1^2/(F1^2-f^2)]$ is subtracted from $H1$ and $H2$, where f is the frequency at which the harmonic is located.

[6]The quantity

$$20\log_{10}\left(\frac{\left[1-\left(\frac{F3}{F1}\right)^2\right]\left[1-\left(\frac{F3}{F2}\right)^2\right]}{\left[1-\left(\frac{F3}{\widetilde{F1}}\right)^2\right]\left[1-\left(\frac{F3}{\widetilde{F2}}\right)^2\right]}\right)$$

is added to $A3$, where $\widetilde{F1}$ and $\widetilde{F2}$ are the first- and second-formant frequencies of a neutral vowel.

Ananthapadmanabha, T. V. (**1984**). ‘‘Acoustic analysis of voice source dynamics,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 1–24.
Ananthapadmanabha, T. V. (**1993**). ‘‘Modeling the return phase of the derivative of glottal flow,’’ Speech Communication Group Working Papers, **9**, Research Laboratory of Electronics, MIT, Cambridge, MA, 28–34.
Chuang, E. S., and Hanson, H. M. (**1996**). ‘‘Glottal characteristics of male speakers: Acoustic correlates and comparison with female data,’’ J. Acoust. Soc. Am. **100**, 2657 (A).
Cranen, B., and Schroeter, J. (**1995**). ‘‘Modeling a leaky glottis,’’ J. Phon. **23,** 165–177.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fant, G. (**1962**). ‘‘Formant bandwidth data,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 1–3.
Fant, G. (**1972**). ‘‘Vocal tract wall effects, losses, and resonance bandwidths,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 28–52.
Fant, G. (**1979**). ‘‘Glottal source and excitation analysis,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 85–107.
Fant, G. (**1982**). ‘‘The Voice source—Acoustic modeling,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 28–48.
Fant, G. (**1993**). ‘‘Some problems in voice source analysis,’’ Speech Commun. **13**, 7–22.
Fant, G., and Lin, Q. (**1988**). ‘‘Frequency domain interpretation and derivation of glottal flow parameters,’’ Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 1–21.
Fant, G., Kruckenberg, A., Liljencrants, J., and Båvegård, M. (**1994**). ‘‘Voice source parameters in continuous speech. Transformation of LF-

parameters,'' ICSLP 94 (Acoustical Society of Japan, Yokohama, Japan), 1451–1454.
Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 1–13.
Fujimura, O., and Lindqvist, J. (**1971**). ''Sweep-tone measurements of vocal-tract characteristics,'' J. Acoust. Soc. Am. **49,** 541–558.
Gauffin, J., and Sundberg, J. (**1989**). ''Spectral correlates of glottal voice source waveform characteristics,'' J. Speech Hear. Res. **32**, 556–565.
Gobl, C. (**1989**). ''A preliminary study of acoustic voice quality correlates,'' Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 9–22.
Gobl, C., and Ní Chasaide, A. (**1988**). ''The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study,'' Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 23–59.
Hanson, H. M. (**1995**). ''Glottal characteristics of female speakers,'' Ph.D. thesis, Harvard University, Cambridge, MA.
Hertegård, S., Gauffin, J., and Karlsson, I. (**1992**). ''Physiological correlates of the inverse filtered flow waveform,'' J. Voice **6**, 224–234.
Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). ''Acoustic correlates of breathy vocal quality,'' J. Speech Hear. Res. **37**, 769–778.
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84,** 511–529; Erratum **85,** 1787.
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1989**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch,'' J. Voice **4**, 294–305.
Holmberg, E. B., Hillman, R. E., Perkell, J. S., and Gress, C. (**1994a**). ''Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings,'' J. Speech Hear. Res. **37**, 484–495.
Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., and Goldman, S. L. (**1995**). ''Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice,'' J. Speech Hear. Res. **38**, 1212–1223.
Holmberg, E. B., Perkell, J. S., Hillman, R. E., and Gress, C. (**1994b**). ''Individual variation in measures of voice,'' Phonetica **51**, 30–37.
House, A. S., and Stevens, K. N. (**1958**). ''Estimation of formant band widths from measurements of transient response of the vocal tract,'' J. Speech Hear. Res. **1**, 309–315.
Karlsson, I. (**1986**). ''Glottal wave forms for normal female speakers,'' J. Phon. **14**, 415–419.
Karlsson, I. (**1988**). ''Glottal waveform parameters for different speaker types,'' Speech '88: Proceedings 7th FASE Symposium, edited by W. A. Ainsworth and J. N. Holmes (Institute of Acoustics, Edinburgh), 225–231.
Karlsson, I. (**1989**). ''A female voice for a text-to-speech system,'' *Eurospeech 89*, edited by J. P. Tubach and J. J. Mariani (CEP Consultants, Edinburgh), Vol. 1, 349–352.
Karlsson, I. (**1990**). ''Voice source dynamics for female speakers,'' ICSLP 90 Proceedings (Acoustical Society of Japan, Kobe, Japan), 69–72.
Karlsson, I. (**1991a**). ''Dynamic voice quality variations in natural female speech,'' Speech Commun. **10**, 481–490.
Karlsson, I. (**1991b**). ''Female voices in speech synthesis,'' J. Phon. **19**, 111–120.
Karlsson, I. (**1992a**). ''Analysis and synthesis of different voices with emphasis on female speech,'' Ph.D. thesis, Royal Institute of Technology, Stockholm.
Karlsson, I. (**1992b**). ''Modelling voice variations in female speech synthesis,'' Speech Commun. **11**, 1–5.
Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (**1986**). ''Normalized noise energy as an acoustic measure to evaluate pathologic voice,'' J. Acoust. Soc. Am. **80,** 1329–1334.
Klatt, D., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87,** 820–857.
Klingholz, F. (**1987**). ''The measurement of the signal-to-noise ratio (SNR) in continuous speech,'' Speech Commun. **6**, 15–26.
de Krom, G. (**1993**). ''A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals,'' J. Speech Hear. Res. **36**, 254–266.
Ladefoged, P., and Antoñanzas-Barroso, N. (**1985**). ''Computer measures of breathy voice quality,'' Work. Papers Phon. **61**, University of California at Los Angeles, 79–86.
Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge U.P., Cambridge).
Linville, S. (**1992**). ''Glottal gap configurations in two age groups of women,'' J. Speech Hear. Res. **35**, 1209–1215.
Mori, K., Blaugrund, S. M., and Yu, J. D. (**1994**). ''The turbulent noise ratio: An estimation of noise power of the breathy voice using PARCOR analysis,'' Laryngoscope **104**, 153–158.
Ní Chasaide, A., and Gobl, C. (**1993**). ''Contextual variation of the vowel voice source as a function of adjacent consonants,'' Language Speech **36**, 303–330.
Peppard, R. C., Bless, D. M., and Milenkovic, P. (**1988**). ''Comparison of young adult singers and nonsingers with vocal nodules,'' J. Voice **2**, 250–260.
Perkell, J. S., Hillman, R. E., and Holmberg, E. B. (**1994**). ''Group differences in measures of voice production and revised values of maximum airflow declination rate,'' J. Acoust. Soc. Am. **96,** 695–698.
Shadle, C. (**1985**). ''The acoustics of fricative consonants,'' Technical Report 506, Research Laboratory of Electronics, MIT, Cambridge, MA.
Siebert, W. M. (**1986**). *Circuits, Signals, and Systems* (MIT, Cambridge).
Södersten, M., and Lindestad, P.-Å. (**1990**). ''Glottal closure and perceived breathiness during phonation in normally speaking subjects,'' J. Speech Hear. Res. **33**, 601–611.
Södersten, M., Lindestad, P.-Å., and Hammarberg, B. (**1991**). ''Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers,'' in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego).
Stevens, K. N. (**1971**). ''Airflow and turbulence noise for fricative and stop consonants,'' J. Acoust. Soc. Am. **50,** 1180–1192.
Stevens, K. N. (**1993**). ''Models for the production and acoustics of stop consonants,'' Speech Commun. **13**, 367–375.
Stevens, K. N. (**to appear**). *Acoustic Phonetics* (MIT, Cambridge).
Stevens, K. N., and Hanson, H. M. (**1995**). ''Classification of glottal vibration from acoustic measurements,'' in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego).
Tize, I. R. (**1989a**). ''A four-parameter model of the glottis and vocal fold contact area,'' Speech Commun. **8**, 191–201.
Titze, I. R. (**1989b**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85,** 1699–1707.
Titze, I. R. (**1995**). ''Definitions and nomenclature related to voice quality,'' in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego).
Yumoto, E., Gould, W. J., and Baer, T. (**1982**). ''Harmonics-to-noise ratio as an index of the degree of hoarseness,'' J. Acoust. Soc. Am. **71,** 1544–1550.
Zemlin, W. R. (**1988**). *Speech and Hearing Science: Anatomy and Physiology* (Prentice-Hall, Englewood Cliffs, NJ).