Glottal characteristics of male speakers: Acoustic correlates and comparison with female data[a)]

Helen M. Hanson[b)]
Sensimetrics Corporation, 48 Grove St., Suite 305, Somerville, Massachusetts 02144-2500

Erika S. Chuang
Department of Electrical Engineering, Stanford University, Stanford, California 94305-9505

(Received 18 June 1998; accepted for publication 18 May 1999)

Acoustic measurements believed to reflect glottal characteristics were made on recordings collected from 21 male speakers. The waveforms and spectra of three nonhigh vowels (/æ, ʌ, ɛ/) were analyzed to obtain acoustic parameters related to first-formant bandwidth, open quotient, spectral tilt, and aspiration noise. Comparisons were made with previous results obtained for 22 female speakers [H. M. Hanson, J. Acoust. Soc. Am. **101**, 466–481 (1997)]. While there is considerable overlap across gender, the male data show lower average values and less interspeaker variation for all measures. In particular, the amplitude of the first harmonic relative to that of the third formant is 9.6 dB lower for the male speakers than for the female speakers, suggesting that spectral tilt is an especially significant parameter for differentiating male and female speech. These findings are consistent with fiberscopic studies which have shown that males tend to have a more complete glottal closure, leading to less energy loss at the glottis and less spectral tilt. Observations of the speech waveforms and spectra suggest the presence of a second glottal excitation within a glottal period for some of the male speakers. Possible causes and acoustic consequences of these second excitations are discussed.
[S0001-4966(99)06208-6]

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

INTRODUCTION

The work reported in this paper focuses on individual variations in glottal configuration and glottal-source waveform characteristics. This research has relevance for several areas of speech research and applications. It adds to studies that seek to establish quantitative ranges of voice-source characteristics of normal speakers. Such studies are used to understand the production of normal voice and to evaluate pathological voice (Holmberg *et al.*, 1988). They may also be useful for modeling the variation among speakers encountered by speech recognition and speaker recognition systems. Likewise, it may be necessary to include models of such variations in voice in speech-synthesis systems.

In addition to individual variations, we address gender differences in voice production. In the past, much of the literature that described acoustic differences between male and female speech concentrated on differences in fundamental frequency and formant frequencies (for a review, see Karlsson, 1992b). These features can be used to automatically distinguish male speakers from female (Childers and Wu, 1991). However, it is noteworthy that while most speech-analysis tools and speech applications are also based on those features, they are generally more successful for male speech. A particular problem has been synthesis of female speech, which tends to sound less natural than synthesized male speech (Karlsson, 1992a). Thus, it would seem that features of the acoustic sound-pressure waveform other than fundamental frequency and formant frequencies contribute to gender characteristics of speech. Recent studies have looked more closely at female speech and the ways in which it differs from male speech, especially in regard to voice-source characteristics. (See, for example, Titze, 1989; Klatt and Klatt, 1990; Karlsson, 1992a.) Observation of the glottis during phonation has suggested that the presence of a posterior glottal opening that persists throughout a vibratory cycle is common for female speakers, but occurs much less frequently among male speakers (Södersten and Lindestad, 1990). Holmberg *et al.* (1988, 1989) have found differences in glottal-waveform characteristics that may affect perceived voice quality; for example, female speakers tended to have larger open quotients and more gradual rises and falls in glottal flow than male speakers. Klatt and Klatt (1990) and Hanson (1995b) have shown that careful control of glottal characteristics improves the naturalness of synthesized female speech. Thus, glottal configuration and its effects on voice-source characteristics may play a significant role in the perception of gender from speech.

Quantitative estimates of the characteristics of the glottal waveform are difficult to obtain. Inverse filtering of the acoustic sound pressure or oral-airflow signal are the most common methods of analysis, but both are sensitive to experimental error, and require strict conditions and special equipment during data acquisition. Thus, while voice-source characteristics may prove useful for improving applications such as speech synthesis, speech recognition, and speaker recognition, analysis techniques that are robust and easily

[a)]The work reported in this paper was completed while the authors were at the M.I.T. Research Laboratory of Electronics, Speech Communication Group.
[b)]Electronic mail: hanson@sens.com

automated are necessary in order to make their use practical.

One alternative to inverse filtering is to make measurements directly on the acoustic sound-pressure waveform and spectrum. These measurements require only simple microphone recordings, and have the potential to be easily automated. Such measures were described and used in a study of 22 female speakers (Hanson, 1995a, 1997). Preliminary evidence based on breathiness ratings and on fiberscopic images collected from a subset of the subjects suggested that the acoustic measures could be used to categorize the speakers by glottal configuration.

In the current paper, we describe an extension of the earlier work (Hanson, 1995a, 1997) to include male speakers. Acoustic data were collected from 21 males, interpreted in terms of the theoretical models presented in Hanson (1995a, 1997), and compared with the data from females. Based on previous research comparing glottal characteristics of males and females (see above), we expected that if these acoustic parameters did represent characteristics of the glottal source, we would find significant differences in the mean values for the two genders. In addition, because the size of a posterior glottal opening can be considered to provide an additional degree of variability in the acoustic parameters considered, and females are more likely than males to have these openings, we expected that we might find a smaller degree of interspeaker variation among males than among females. Note that there can be other factors that lead to variations among members of a gender group, and therefore gender differences in degree of variation are not necessarily due only to differences in posterior glottal openings.

The outline of the paper is as follows. Section I begins with a brief summary of the theoretical background of the acoustic measures and the results of the previous study. Expected means, maxima, and minima of some of the acoustic measurements are then calculated for male speakers. In Sec. II we describe the experimental procedure, and the methods of making the acoustic measures of the speech waveform and speech spectrum. Section III begins by describing the results of our analysis of the male data. The results are then compared to the female data.

I. THEORETICAL BACKGROUND AND PREDICTIONS

A. Summary of previous study with female speakers

During vowel production, the configuration of the vocal folds may vary in several ways. Four types of glottal configuration were considered by Hanson (1995a, 1997): (1) the arytenoids are approximated and the membranous part of the folds close abruptly; (2) the arytenoids are approximated, but the membranous folds close from front to back along the length of the folds; (3) there is a posterior glottal opening at the arytenoids that persists throughout the glottal cycle, and the folds close abruptly; (4) a posterior glottal opening extends into the membranous portion of the folds throughout the glottal cycle, forcing the folds to close from front to back. Theoretical background was given for the manner in which these various configurations affect the glottal vibration pattern and volume-velocity waveform, and how such effects are manifested in the speech spectrum or waveform. A set of acoustic descriptors was presented as possible correlates of different types of glottal configuration.

When there is complete glottal closure at some time during a cycle of vibration, the glottal waveform can show several kinds of differences. For example, there can be differences in the open quotient (OQ, expressed as a percent). In that case, if all else remains the same, the spectrum of the glottal source is only influenced at the low frequencies. Experiments have shown that changes in the relative amplitudes of the first and second harmonics reflect changes in the closed quotient (Holmberg *et al.*, 1995). (The closed quotient is 100–OQ.)

When the derivative of the glottal-airflow signal has a discontinuity at glottal closure, its spectrum drops off at 6 dB per octave at high frequencies (Stevens, 1998). Sometimes, however, the glottis closes nonsimultaneously; that is, the glottal closure begins at the anterior end of the vocal folds and proceeds back to the posterior end. This type of closure leads to a gradual, rather than abrupt, cutoff of glottal airflow, and thus the derivative of the glottal airflow does not have a discontinuity. As a result, the high-frequency content of the glottal waveform is reduced. If this gradual cutoff is approximated as an exponential, the closing time can be used to approximate the time constant of the effective low-pass filter, and thus the cutoff frequency. According to this model, the spectral tilt of the glottal source above the cutoff frequency increases from 6 to 12 dB per octave.

The presence of a posterior glottal opening (glottal chink) throughout a glottal cycle introduces additional modifications to the spectrum. First, formant bandwidths, particularly that of the first formant, are increased due to additional energy loss at the glottis. Given the size of the glottal chink, one can estimate the additional bandwidth introduced at the frequencies of the formants (Stevens, 1998). The second consequence of the glottal chink is the introduction of additional spectral tilt. This additional tilt is due to the fact that the airflow through the glottal chink cannot undergo a discontinuous change because of the acoustic mass of the moving air in the system. The change in airflow at glottal closure has a time constant which can be used to calculate the cutoff frequency of the effective low-pass filter. Above this frequency, the spectral tilt increases by 6 dB per octave. The third acoustic consequence of a glottal opening is the generation of turbulence noise in the vicinity of the glottis. When the glottal opening is large, the spectral amplitude of the noise becomes comparable to the spectral amplitude of the periodic source at high frequencies.

The following list summarizes the relevant acoustic measurements, made directly on the speech spectrum or waveform, that were used by Hanson (1995a, 1997). These measurements are illustrated by the speech waveforms and spectrum (collected from female speakers) shown in Fig. 1.

1. First-formant bandwidth (B1)

A formant oscillation can be modeled as a damped sinusoid of the form $e^{-\alpha_i t}\cos 2\pi f_i t$, where f_i is the frequency of the ith formant and the constant α_i is the exponential decay rate. By applying a bandpass filter to the speech waveform

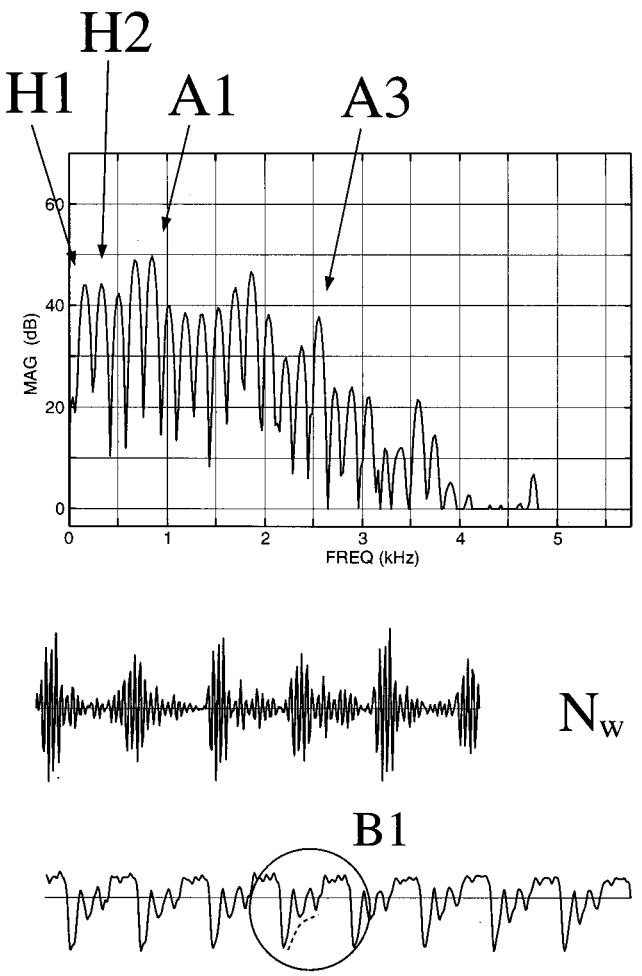

FIG. 1. Speech waveforms and a vowel spectrum produced by female speakers. The acoustic parameters labeled in the spectrum are the amplitudes of the first harmonic ($H1$), second harmonic ($H2$), first formant ($A1$), and third formant ($A3$). The top waveform, N_w, was obtained by bandpass filtering a sound-pressure waveform at the third-formant frequency region. The bottom waveform illustrates the decay of the first-formant oscillation, used to calculate an estimated bandwidth, $B1$. (Please note that each of these three examples is from a different speaker.)

centered at the first-formant frequency ($F1$) and measuring the decay rate of the resulting waveform (as shown in the waveform marked $B1$ in Fig. 1), the bandwidth of $F1$ can be approximated.

2. H1*–H2*

The amplitude of the first harmonic ($H1$) relative to that of the second ($H2$) is used as an indication of open quotient (OQ), the ratio of the open phase of the glottal cycle to the total period. The asterisks indicate that $H1$ and $H2$ have been corrected to remove the amount by which the vocal-tract transfer function, particularly the first formant, ''boosts'' the amplitudes of these harmonics. The effect of this correction is that the amplitudes of the harmonics more closely approximate those of the actual source spectrum. It is particularly important to make this correction when comparing this acoustic measure across vowels, for which the first-formant frequency can vary greatly. The correction factor used is given in Hanson (1997, footnote 5; 1995a, Appendix A). It assumes that the first-formant bandwidth is small. This assumption is valid unless the harmonic frequencies are too close to the first-formant frequency; that is, less than about 100 Hz away.

3. H1*–A1

The amplitude of the first harmonic ($H1$) relative to that of the first-formant prominence in the spectral domain ($A1$) reflects the bandwidth of $F1$, and may also be affected by source spectral tilt. Hence, this measure is an indication of the presence of a posterior glottal chink. Again, the asterisk indicates that H1 is corrected for the effect of the first formant.

4. H1*–A3*

The amplitude of the first harmonic ($H1$) relative to that of the third-formant spectral peak ($A3$) reflects the source spectral tilt. The mid- to high-frequency components are mostly influenced by how abruptly the flow is cut off when the glottis closes. $H1$ is corrected for the effect of the first formant (see above), and $A3$ is corrected for the contribution of the first and second formants on the amplitude of the third-formant prominence. As for the correction to $H1$ and $H2$, the correction to $A3$ is particularly important when comparing across vowels, for which the first two formant frequencies can vary greatly. This correction factor is given in Hanson (1997, footnote 6; 1995, Appendix A). It assumes that the bandwidths of the first two formants are small, which is valid as long as the third formant is not too close to the second. For the vowels considered in this study, the second and third formants are typically well separated.

5. Noise rating N_w

A posterior glottal opening that persists during phonation introduces an increase in turbulence noise generated in the vicinity of the glottis. The noise becomes prominent at the high-frequency range because of its increased amplitude, together with an increase in spectral tilt. Because the energy of the first and second formants is relatively strong, evidence of aspiration noise can usually not be seen when viewing the vowel waveform. Klatt and Klatt (1990) introduced a method of estimating degree of aspiration noise in a vowel waveform. Judgments of waveform irregularity are made on a four-point scale by viewing the vowel waveform after bandpass filtering at the third-formant frequency, where aspiration noise should be visible. An example of such a filtered waveform is given in Fig. 1.

Using acoustic models, Hanson (1995a, 1997) made estimates of maximum, minimum, and average values for some of the acoustic measures for female speakers. These predictions from models were based in part on analysis of the KL-GLOTT88 glottal-source model (Klatt and Klatt, 1990), together with minimum-airflow data collected during vowel production (Holmberg *et al.*, 1988). Data collected from 22 female speakers were in agreement with the model-based estimates, and substantial differences among individual speakers were found. There were strong correlations between several of the acoustic measures. In particular, the source

TABLE I. Average dimensions of the glottis, trachea, and vocal tract for male speakers.

Length of vocal tract	17 cm
Vertical length of glottis	0.4 cm
Length of trachea	12 cm
Cross-sectional area of vocal tract	4 cm^2
Cross-sectional area of trachea	2.5 cm^2

spectral tilt indicator ($H1^*-A3^*$) was found to be highly correlated with that of the first-formant prominence ($H1^*-A1$) and the noise judgment (N_w). According to the theoretical models, all three measurements are related to the existence and size of a glottal chink. Analysis of the acoustic measurements suggested that the subjects could be classified into two groups. Speakers having relatively low values of $H1^*-A3^*$ and $H1^*-A1$, indicating strong high-frequency content and prominent $F1$ spectral peaks, were classified as members of group 1. The noise ratings were also lower for this group, suggesting little aspiration noise in the vowels. Speakers in this group were hypothesized to have abrupt glottal closure, with posterior glottal chinks that vary in size (including zero). Speakers in group 2 had higher values of $H1^*-A3^*$ and $H1^*-A1$, indicating much less high-frequency energy and weaker $F1$ peaks. The higher noise ratings for these speakers suggested more aspiration noise in the speech. Therefore, speakers in this group were hypothesized to have relatively large posterior glottal chinks extending beyond the vocal processes, with nonsimultaneous glottal closure.

Preliminary results of fiberscopic images collected from four of the 22 subjects supported this hypothesis (Hanson, 1995a). Moreover, a listening test showed that group 2 speakers were perceived as having breathier voice quality than group 1 speakers, also suggesting that they had relatively large glottal chinks (cf. Södersten and Lindestad, 1990).

B. Current study

In the current study, we extended the earlier experiment to male speakers. Based on the theoretical models and the equations presented in Hanson (1995a, 1997), we calculated the contribution of a posterior glottal chink to the spectral tilt and the first-formant bandwidth, for a range of cross-sectional areas for the chink. For these calculations, we assumed a uniform vocal tract having the dimensions given in Table I. In addition, we assumed a subglottal pressure of 6300 dynes/cm^2 (6.4 cm H_2O), which is the average value for males found by Holmberg *et al.* (1988). For each cross-sectional area of the glottal chink A_{ch}, we also calculated U_{ch}, the minimum (dc) flow. The estimations are summarized in Table II. We see, for example, that a glottal chink of area 5 mm^2 would result in a first-formant bandwidth increase of 77 Hz and an additional tilt of 12 dB at 2500 Hz.

From analysis of the glottal-waveform model KLGLOTT88 (Klatt and Klatt, 1990) and experimentally measured formant bandwidths (Fant, 1972), we estimated minimum values for the measures $H1-A1$ and $H1-A3$. These calculations assumed a uniform vocal tract for which

TABLE II. 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 flow cutoff; resulting increment in spectral tilt at 2500 Hz (Tilt). A subglottal pressure of 6300 dynes/cm^2 (6.4 cm H_2O) (Holmberg *et al.*, 1988) is assumed. Other assumed values include: vocal-tract, glottis, and trachea dimensions given in Table I, first-formant frequency of 500 Hz, and vocal-tract losses of 73 Hz for vowel /æ/ (House and Stevens, 1958).

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	73	37	0	0	0.0
0.01	15	88	39	33	0.13	7.3
0.02	31	104	40	66	0.16	8.8
0.03	46	119	42	100	0.17	10.0
0.04	62	135	43	133	0.20	11.1
0.05	77	150	44	166	0.23	12.1
0.06	93	166	44	199	0.25	13.0
0.07	108	181	45	232	0.28	13.8
0.08	124	197	46	265	0.30	14.5
0.09	139	212	47	299	0.32	15.2
0.10	155	228	47	332	0.35	15.8

$F1=500$ Hz, $F2=1500$ Hz, …, a fundamental frequency of 100 Hz, and an open quotient of 50%. For this hypothetical case, the formants are centered on harmonics, and measuring $H1-A1$ and $H1-A3$ directly from the synthesized spectra would give us estimates of the minimum *possible* values. However, because it is often the case that formant frequencies are not centered on harmonics, we estimated the minimum *expected* values by subtracting 3 dB from the amplitude of $A1$, and 2 dB from the amplitude of $A3$, to compensate. By matching the calculated minimum airflow (U_{ch}), shown in Table II, with measured minimum (dc) airflow data collected by Holmberg *et al.* (1988) and Perkell *et al.* (1994), we estimated the average and maximum expected areas of posterior glottal chinks (A_{ch}). From these values, the expected average and maximum first-formant bandwidths were obtained by calculating the contribution of the glottal chink to the first-formant bandwidth, using an equation given in Hanson (1997), and then adding that value to minimum bandwidths measured experimentally (Fant, 1972; House and Stevens, 1958). The average and maximum values of $H1-A1$ were estimated using the bandwidth estimates. The average and maximum values of $H1-A3$ were obtained from Table II using the average and maximum values of A_{ch}.

Table III summarizes these predicted values. Because

TABLE III. Predicted average, minimum, and maximum values of acoustic measures for male speakers, assuming an $F0$ of 100 Hz, an open quotient of 50 percent, and a uniform vocal tract with resonant frequencies at odd multiples of 500 Hz. $H1-A1$ and $H1-A3$ are given in dB, and $B1$ is given in Hz. The maximum estimated value of $H1-A3$, given in parentheses, is based on the assumption of simultaneous closure of the vocal folds, and therefore, measured values could be higher. See Sec. I B of the text for additional details of the calculations.

	$H1-A1$	$H1-A3$	$B1$
Average	−7.0	13.7	119
Minimum	−11.3	4.7	73
Maximum	−2.3	(19.7)	205

additional spectral tilt due to both a posterior glottal chink and nonsimultaneous glottal closure is difficult to estimate, our estimate of the maximum value of $H1-A3$ assumes a simultaneous glottal closure. Therefore, we expected that we might see even higher values in the human data. Predicted values for the females tended to be higher than those for the males; the maximum expected values of both $B1$ and $H1^*-A1$ were higher, as were both the minimum and maximum values of $H1^*-A3^*$.

We next describe experimental data collected from male speakers on which the acoustic measures described in Sec. I A were made. As described in the Introduction, earlier work by other researchers suggests that females are more likely to have posterior glottal chinks than are males. Therefore, theory predicts that there should be significant differences between the average values of the acoustical descriptors for male and female speakers. It is also possible that males will display a smaller degree of variation than females.

II. EXPERIMENT

A. Speakers and speech material

We collected data from 21 adult male speakers between the ages of 19 and 71. Most of these speakers were MIT students. Four of the subjects had significant experience participating in speech production experiments. The speakers appeared to have no signs of voice or hearing problems, and all were native speakers of American English. The microphone was positioned to be 20 cm from a speaker. Subjects were instructed to speak in their normal tone of voice, as naturally as was possible. There was otherwise no control of intensity. 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, and the first and second formants are well separated from each other, thus simplifying the acoustic measures. Each utterance was repeated seven times, with the 21 sentences presented in random order. The first and last tokens of each vowel were discarded. The remaining 15 utterances were low-pass filtered at 4.5 kHz and digitized with a sampling rate of 10 kHz.

B. Measurements

The acoustic measurements described in Sec. I A were made in the following manner:

1. First formant bandwidth (B1)

Each repetition of the vowel /æ/ was bandpass filtered around its average $F1$ frequency using a four-pole digital Butterworth filter with a bandwidth of 600 Hz. The first-formant bandwidth was estimated from the rate of decay of the resulting waveform as determined by the peak-to-peak amplitude of the first two $F1$ oscillations. The vowel /æ/ was chosen because its $F1$ is usually high enough to allow at least two oscillations to take place during the closed part of the glottal cycle. Estimates were made on eight consecutive pitch periods during a stable section of each token. The 40 estimates were then averaged to obtain a mean value for each speaker.

2. H1–H2*, H1*–A1, H1*–A3**

Measurements were made for all repetitions of the three vowels. Spectra were obtained by applying a Hamming window to the speech signals, and computing the 512-point discrete Fourier transform (DFT). The length of the Hamming window was chosen such that a minimum of two complete pitch periods of the waveform were covered; it ranged from 32 to 50 ms, depending on the average pitch of each speaker. For the vowel /æ/, the window was centered at the initial part of the eight consecutive glottal cycles from which the first-formant bandwidth was estimated. For the vowels /ɛ/ and /ʌ/, the measurements were taken three times throughout each token at 20-ms intervals. Corrections were made to normalize transfer-function effects on the amplitude of the third formant. The effects of neighboring formants were removed, for which, on average, $A3$ was corrected by -7.8 dB for /æ/, -4.6 dB for /ɛ/, and $+2$ dB for /ʌ/. In addition, the amplitude $A3$ was further adjusted, because the average third-formant bandwidth varies by vowel. House and Stevens (1958) measured the $F3$ bandwidths for the vowels /æ, ʌ, ɛ/ at 103, 64, and 88 Hz, respectively. Thus, vowels /æ/ and /ɛ/ would, on average, have $F3$ amplitudes that are 4 and 3 dB lower, respectively, than that of the vowel /ʌ/. Because we are only interested in source effects on the amplitude of $F3$, we compensated by adding 4 dB to the measure $A3^*$ for /æ/, and 3 dB for /ɛ/.

3. Noise rating (N_w)

Each repetition of the three vowels was bandpass filtered around its average $F3$ frequency using a filter with a bandwidth of 600 Hz. Plots of the filtered waveforms were arranged randomly across vowel and speaker, and were rated on a four-point scale: (1) periodic, no visible noise; (2) periodic but occasional noise intrusion; (3) weakly periodic, clear evidence of noise excitation; (4) little or no periodicity, noise is predominant (Klatt and Klatt, 1990). The ratings were made independently by two judges, who did not know which waveforms corresponded to which speaker. Their ratings were well correlated ($r=0.77$), and the average difference in the two ratings for each token was only 0.36. Their two ratings were then averaged to obtain one noise rating for each token per speaker. Figure 2 shows typical waveforms corresponding to the four rating levels.

Despite the consistency of their ratings, the judges reported feeling somewhat uncertain about them because some of the waveforms had what appeared to be second pulses appearing about halfway through a pitch period. These extra pulses gave the appearance of noise, but did not really seem to be the random noise expected for aspiration. Hence, the judges were not convinced that high noise ratings actually reflected large degrees of aspiration noise. This effect is discussed further in Sec. III B.

III. RESULTS AND DISCUSSION

A. Statistical analysis and comparison with predicted values

The mean values of the acoustic parameters for each speaker are summarized by vowel in Tables IV–VI. Mini-

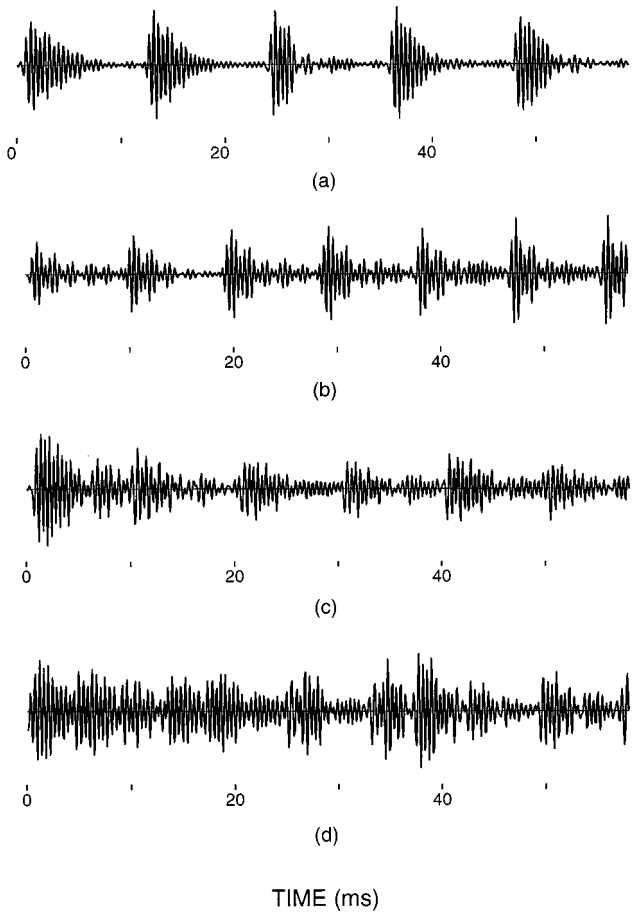

FIG. 2. Examples of the degree of aspiration noise ratings, N_w. The speech waveforms (vowel /æ/) have been bandpass filtered in the third-formant region, and their amplitudes have been scaled to fill the full available range. Ratings were made by two judges and averaged. Each example is from a different male speaker. (a) $N_w \approx 1$; periodic, little evidence of noise. (b) $N_w \approx 2$; periodic, occasional intrusion of noise. (c) $N_w \approx 3$; weakly periodic, clear evidence of noise. (d) $N_w \approx 4$; little or no periodicity, strong noise excitation.

TABLE IV. Average values of the acoustic parameters for the vowel /æ/, 21 male speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, N_w is the waveform-based noise judgment, and $B1$ is the bandwidth of the first formant, given in Hz. Numbers in boldface represent maxima or minima for each measure across speakers. Parentheses around a noise judgment N_w indicate that a speaker was judged to consistently exhibit secondary pulses for the vowel (see Sec. III B for details). The mean value, the standard deviation, and the average standard deviation per each subject are also given for each measure.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	$B1$
M1	3.7	−7.4	9.0	1.4	102
M2	−0.5	−8.0	**6.2**	(2.2)	115
M3	0.6	−6.4	20.6	1.9	103
M4	1.1	−1.2	22	1.9	213
M5	0.4	−2.3	14.6	**1.2**	65
M6	−0.2	−7.8	14.1	1.6	92
M7	−0.7	−8.6	15.8	(2.6)	101
M8	0.3	−2.3	11.1	1.3	127
M9	−0.3	−8.0	16.3	(2.3)	114
M10	−0.3	−7.7	12.6	1.7	80
M11	**−3.3**	−8.7	15.9	1.5	67
M12	0.4	−5.9	21.5	(2.8)	157
M13	−0.9	−5.5	12.7	1.4	147
M14	−0.9	−12.5	10.6	1.7	104
M15	−0.4	−10.3	16.6	1.5	178
M16	−1.3	−5.6	17.6	(2.9)	**58**
M17	−1.2	−5.5	15.2	1.6	**245**
M18	1.8	**0.4**	**24.1**	(**3.2**)	160
M19	**4.0**	−5.2	22.2	2.1	121
M20	−0.6	**−16.1**	11.1	1.8	63
M21	−0.6	−2.3	16.6	(2.2)	230
Mean	0.0	−6.5	15.5	1.9	126
s.d.	1.6	3.8	4.7	0.4	55
Mean s.d.	1.3	1.4	2.5	0.4	20.5

mum and maximum values for each measure are given in boldface, and mean values across speaker are shown at the bottom of each table. The standard deviations of the means are also given. To get some idea of the consistency of each speaker, standard deviations across token were computed for each measure. The average standard deviations across speaker are given at the bottom of Tables IV–VI. Comparing those with the range of values across speaker observed for each parameter, we can say that most speakers were generally quite consistent. However, there are only five tokens for each vowel, and it is not clear to what extent these measures of variability represent a larger sample size.

All measurements were subjected to repeated-measures analysis of variance (ANOVA) with vowel /æ, ʌ, ɛ/ as a within-subject factor (Table VII). Only $H1^*-A3^*$ showed a significant difference among vowels [$F(2,40)=11$, $p<0.01$]. Reference to Tables IV–VI shows that the mean for this measure is about 3.5 dB greater for the vowel /æ/ than for the other two vowels. Upon closer inspection, we found that for eight speakers $H1^*-A3^*$ is significantly larger for /æ/ than for /ɛ/ or /ʌ/, while for the remaining speakers there is little difference among the vowels. Recalling that /æ/ has a relatively large third-formant bandwidth because it is produced with a larger degree of mouth opening, it is possible that these speakers have particularly large third-formant bandwidths for this vowel. Although we attempted to remove filter effects, this result indicates that the measure might be sensitive to variations in articulation among speakers. However, given that the difference among vowels is due to a minority of speakers, and that compared with the total range of the measure, the difference is small, we do not consider each vowel separately in the following analysis.

The means, maxima, minima, and ranges of the measures are summarized in Table VIII, and compared with estimated values from Sec. I B where appropriate. As can be seen, there is considerable subject-to-subject variability in the measurements. For the most part, the measured values are quite close to the estimated values, in several cases within less than 1 decibel.

The minimum value of $H1^*-A1$ is about 5 dB less than that predicted, but this relatively large difference is only due to one speaker (M20) who has an unusually low first-harmonic amplitude ($H1$). It is likely that this speaker has an open quotient that is significantly lower than the 50% which was assumed for calculation of the predicted value. The prediction may also have been too conservative because we assumed that the first formant does not usually center on a harmonic, and therefore we reduced the estimated amplitude of the first-formant peak ($A1$) by 3 dB. In retrospect,

TABLE V. Average values of the acoustic parameters for the vowel /ʌ/, 21 male speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, and N_w is the waveform-based noise judgment. Numbers in boldface represent maxima or minima for each measure across speakers. Parentheses around a noise judgment N_w indicate that a speaker was judged to consistently exhibit secondary pulses for the vowel (see Sec. III B for details). The mean value, the standard deviation, and the average standard deviation per each subject are also given for each measure.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w
M1	**4.2**	−4.9	10.3	1.5
M2	−0.2	−9.8	**4.8**	1.5
M3	1.9	−6.5	17.0	2.0
M4	0.8	−3.9	**22.8**	2.2
M5	2.5	−2.9	13.0	1.5
M6	−2.3	−9.8	11.1	1.7
M7	−0.1	−5.7	18.1	2.3
M8	0.4	−5.7	11.3	1.4
M9	−1.8	−11.1	8.8	(2.3)
M10	−2.3	−6.1	11.2	1.8
M11	**−2.7**	−9.4	14.8	2.0
M12	0.3	−6.6	15.4	2.3
M13	−1.8	−3.5	10.7	1.4
M14	−1.4	−11.2	9.7	1.3
M15	0.4	−6.0	15.4	1.6
M16	−1.9	−2.5	18.8	(2.5)
M17	1.2	−7.0	9.6	**1.2**
M18	2.0	**−2.2**	15.4	(2.5)
M19	−0.1	−8.5	18.4	**2.9**
M20	−2.3	**−14.9**	7.9	1.4
M21	−1.0	−7.5	10.9	1.5
Mean	−0.2	−6.9	13.1	1.8
s.d.	1.9	3.3	4.4	0.3
Mean s.d.	0.8	1.4	2.4	0.4

TABLE VI. Average values of the acoustic parameters for the vowel /ɛ/, 21 male speakers, where $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, and N_w is the waveform-based noise judgment. Numbers in boldface represent maxima or minima for each measure across speakers. Parentheses around a noise judgment N_w indicate that a speaker was judged to consistently exhibit secondary pulses for the vowel (see Sec. III B for details). The mean value, the standard deviation, and the average standard deviation per each subject are also given for each measure.

Subject	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w
M1	**3.8**	−4.9	10.5	1.4
M2	1.2	−9.9	**5.7**	1.6
M3	0.8	−7.2	17.2	1.8
M4	−0.8	−5.9	18.8	2.1
M5	2.2	−3.7	13.5	1.4
M6	−2.3	−11.9	12.0	1.5
M7	−0.1	−7.0	12.7	1.8
M8	−0.2	−6.2	12.5	**1.3**
M9	−2.1	−12.3	7.8	(2.7)
M10	−2.3	−6.7	9.7	1.5
M11	**−3.2**	−9.6	11.7	1.7
M12	1.4	−5.1	**23.1**	(3.0)
M13	−0.7	−6.5	7.5	1.4
M14	−1.8	−12.7	8.2	1.8
M15	0.1	−7.3	15.6	1.7
M16	−1.5	−2.3	19.2	(2.8)
M17	2.6	−6.8	11.2	1.6
M18	3.5	**−1.2**	17.1	(**3.0**)
M19	3.1	−5.0	20.7	2.1
M20	−2.3	**−14.0**	6.4	1.5
M21	−0.7	−6.8	9.9	1.5
Mean	0.0	−7.3	12.9	1.9
s.d.	2.1	3.4	4.9	0.4
Mean s.d.	1.0	1.6	2.5	0.3

considering that the harmonics are closely spaced for most male speakers, perhaps we should not have made this adjustment.

The total range of $H1^*-A3^*$ is about 19 dB; that is, some speakers have strong high-frequency content in the vowel, while others do not. The maximum measured value of $H1^*-A3^*$ was 24.1 dB, about 4 dB larger than that expected for abrupt glottal closure, suggesting that some speakers may have nonsimultaneous glottal closure or relatively wide third-formant bandwidths.

The first-formant bandwidth $B1$, as estimated from the speech waveform, ranges widely from a minimum of 58 Hz to a maximum of 245 Hz, with an average of 126 Hz. The noise judgments ranged from 1.2 to 3.2, indicating that some of the speakers show little noise at high frequencies, while others show significant noise. This result is somewhat surprising because in Klatt and Klatt's (1990) male data, stressed syllables were not given noise ratings higher than 2.0. We discuss this result in more detail in the next section.

$H1^*-H2^*$ has a total range of about 7.5 dB. Holmberg *et al.* (1988) found that the open quotient for a group of 25 male subjects ranged from 46% to 77%. According to the KLGLOTT88 model (Klatt and Klatt, 1990), this range corresponds to an 8.6-dB change in the measure $H1^*-H2^*$, assuming that $F0$ is 100 Hz. Therefore, the range of $H1^*-H2^*$ found in the current experiment seems reasonable.

In Fig. 3 we compare spectra from two speakers, illustrating the variability among male speakers. Subject M20, in the upper panel, has strong harmonic structure at all frequencies, and sharp formant peaks. He has relatively little energy in the region of the first two harmonics. Subject M18, in the lower panel, has good harmonic structure only up to about 2 kHz; above that frequency, the spectrum appears noisy. His formant peaks are less well defined, especially the first. Energy in the region of the first and second harmonics is quite strong. The amplitudes of the formant peaks fall off more rapidly for M18 than for M20.

Table IX shows Pearson product moment correlation coefficients for the various measures. The correlations are moderate ($0.49 \leqslant r \leqslant 0.60$, $N=63$). One would expect that because degree of spectral tilt, high-frequency noise, and $H1^*-A1$ are related to the presence and size of a posterior glottal opening, higher correlations should be observed among these parameters. The correlation between $H1^*-A1$

TABLE VII. Results of repeated measures analyses of variance (ANOVA) performed to examine differences in the acoustic parameters across vowels. An asterisk (*) in the third column indicates statistical significance.

Measure	$F(2,40)$	p
$H1^*-H2^*$	0.4	>0.1
$H1^*-A1$	1.2	>0.1
$H1^*-A3^*$	11.0	<0.01*
Noise	0.8	>0.1

TABLE VIII. Mean, minimum, and maximum values of the measured (Meas.) acoustic parameters, compared, where appropriate, with the values estimated (Est.) in Sec. I B. $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB, and $B1$ is given in Hz. The estimated maximum of $H1^*-A3^*$ is given in parentheses to indicate that it is a lower bound on the maximum (see the text and the caption for Table III for details).

	$H1^*-H2^*$	$H1^*-A1$		$H1^*-A3^*$		N_w	$B1$	
	Meas.	Meas.	Est.	Meas.	Est.	Meas.	Meas.	Est.
Mean	0	−6.9	−7.0	13.8	13.7	1.8	126	119
Minimum	−3.3	−16.1	−11.3	4.8	4.7	1.2	58	73
Maximum	4.2	0.4	−2.3	24.1	(19.7)	3.2	245	205
Range	7.5	16.5	13.6	19.3	15	2	187	132
s.d.	1.8	3.5	n.a.	4.8	n.a.	0.4	43	n.a.

and $B1$ is also low ($r=0.44$, $N=21$), given that A1 is mostly influenced by its bandwidth. However, other factors do influence A1, including decay rate during glottal opening and source spectral tilt. In addition, as we describe in the next section, the method of approximating the $F1$ bandwidth might have been inaccurate because of interference from a second excitation pulse.

B. Complicating factors

The analyses of the data were complicated by two factors not observed for female speakers. First, for a number of the male speakers, a second pulse during a glottal period was observed in the speech waveform. Such second pulses were rarely seen for the female subjects of Hanson (1995a, 1997). Figure 4(a) and (c) show vowel waveforms from subjects M9 and M16, in which examples of such pulses are indicated with arrows. As can be seen in Fig. 4(b) and (d), the pulses are more easily identified in the signals on which the noise judgments were made; that is, the speech waveform bandpass filtered at the third formant.

Given that the initial deflection for these pulses is in the same direction as the pulses at the beginning of a period, and that it occurs about halfway through a period, it is likely that it is the result of excitation of the vocal tract at the time of glottal opening. (Note that the change in the derivative of the glottal waveform is positive following both glottal opening and closure.) As a rough approximation, then, we can consider a second pulse to be a delayed, attenuated replica of the main pulse. By Fourier analysis, these second pulses should result in an attenuation of certain harmonics in the speech spectrum. The delay of the second excitation relative to the main excitation determines which harmonics are affected, with the effect being strongest when the second excitation is delayed by about 50% of the glottal cycle. Figure 5 is the spectrum of the waveform shown in Fig. 4(a), where the second excitations occur about halfway between the main pulses. As the model predicts, the amplitude of almost every other harmonic appears to be attenuated, resulting in a noisy-looking or irregular spectrum. Preliminary experiments with speech synthesis show similar changes in the speech spectrum. Note also that multipulse excitation has been found to improve the quality of speech synthesized according to the linear predictive coding (LPC) model (Atal and Remde, 1982). Therefore, our hypothesis that the secondary pulses are caused by excitation at glottal opening seems reasonable.

In Fig. 4(b) and (d), we see that the extra pulses interfere with the perceived regularity of the waveforms, and, in fact,

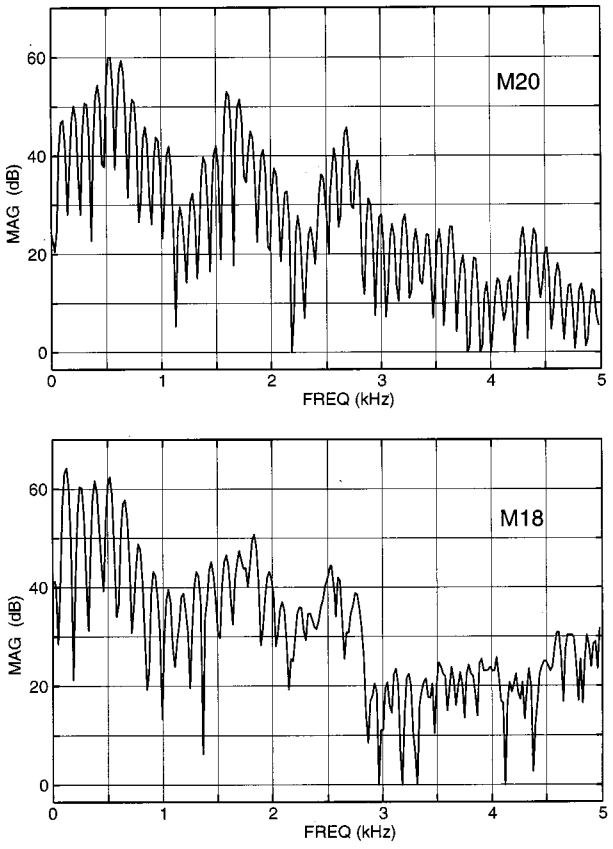

FIG. 3. Illustration of the range of spectral characteristics related to glottal configuration observed in male subjects. The vowel is /ɛ/. The spectrum for subject M18 (below) has greater spectral tilt, less well-defined formant peaks, a greater degree of noise at high frequencies, and a higher relative amplitude of the first harmonic, compared to the spectrum for subject M20 (above).

TABLE IX. Pearson product moment correlation coefficients for the acoustic parameters for the three vowels /æ, ʌ, ɛ/ combined. The notation n.s. indicates that a correlation was not significant. $N=63$, except for correlations with $B1$, for which $N=21$.

	$H1^*-H2^*$	$H1^*-A1$	$H1^*-A3^*$	N_w	$B1$ (/æ/)
$H1^*-H2^*$	1				
$H1^*-A1$	0.49	1			
$H1^*-A3^*$	n.s.	0.55	1		
N_w	n.s.	n.s.	0.60	1	
$B1$ (/æ/)	n.s.	0.44	0.33	n.s.	1

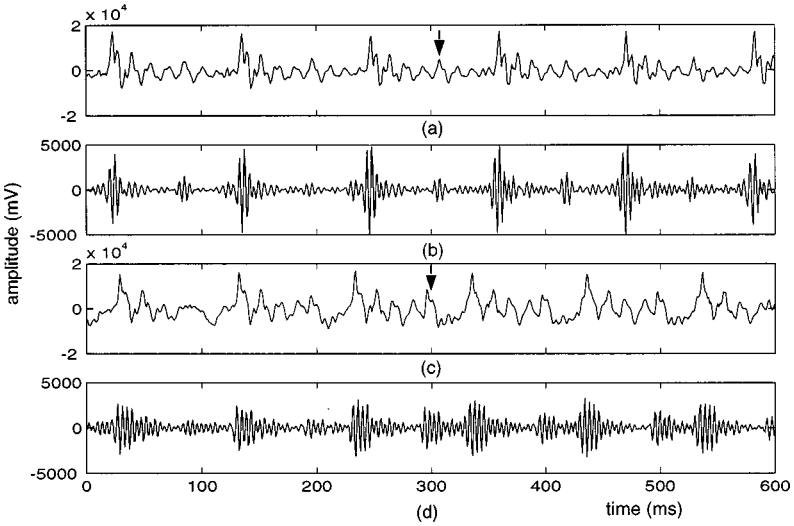

FIG. 4. Speech waveforms showing evidence of a second glottal excitation: (a) Speech waveform for vowel /æ/, subject M9. The arrow indicates an example of a second pulse; (b) The waveform of (a), following bandpass filtering in the $F3$ region. The second pulses are more easily seen. (c) Speech waveform for the vowel /ɛ/, subject M16; (d) The waveform of (c), following bandpass filtering in the $F3$ region. The second pulses show greater variation in amplitude than those of the waveform in (b), making this waveform appear relatively noisy.

we mentioned in Sec. II B that these pulses hampered the confidence of the judges in their noise ratings. To study this effect, the bandpass-filtered waveforms used for the noise judgments were presented to a third judge for evaluation of the existence of a second excitation. If extra pulses were consistently observed, that is, a second pulse was present in almost every glottal period in the waveform with the same amount of delay, the vowel token was judged to have a second excitation. If at least three out of the five tokens for a given vowel were found to have a second excitation, that vowel was labeled as having a second pulse; the remaining vowels were labeled as not having a second pulse. A plot of noise judgment (N_w) vs $H1^*-A3^*$ is shown in Fig. 6. The open circles represent vowel data judged to have a second excitation. Nearly all the vowel tokens given a noise rating much above 2 were also judged to show evidence of second pulses. Given that the male subjects of Klatt and Klatt (1990) did not have noise ratings greater than 2 for stressed syllables, this result suggests that for the group of data in question, the noise ratings do not accurately reflect the degree of aspiration noise in the signal. Although the ratings for this group of data are questionable, they are included in Tables IV–VI because this finding is the result of *post hoc* analysis and our hypothesis is not conclusive. The noise ratings that are in doubt are given in parentheses in those tables.

The question arises as to why we do not observe this phenomenon for all of the data. One possibility is that not every speaker has a second excitation at glottal opening. However, only three speakers were found to have second pulses in all three vowels, while one speaker had second pulses in two vowels and three showed second pulses only for /æ/. It seems unlikely that a speaker would have second pulses only for certain vowels. Hence, it is possible that excitation at glottal opening is common, but obvious only for vowel tokens in which the third-formant oscillations die out quickly. Evidence for the latter is that /æ/ usually has a relatively wide third-formant bandwidth and /æ/ tokens were much more likely to be judged to have a second pulse than the other two vowels.

One can also question why this problem arose for the male speakers and not for the female speakers. A possible explanation is that the lower fundamental frequency of males allows a longer time for the third formant to decay before glottal opening occurs, and thus a second excitation would be easier to see. It could also be that females are less likely to have second excitations for physiological reasons, such as less surface tension of the folds.

Details of the causes and consequences of these extra excitations will require further investigation. However, it is clear that this second excitation and its effects on the speech waveform and spectrum could have consequences for studies of male speech. As we have seen, the high-frequency noise judgments become more difficult, because the existence of the second pulses may make the filtered waveforms seem more irregular, as shown in Fig. 4(b) and (d). Spectrum-based noise measures could also be affected (Fig. 5).

A difficulty also arose for the estimation of the $F1$ bandwidth. For some speakers, the decay of the $F1$ oscillation did not appear to be exponential as expected. Some waveforms showed signs of an increase in the amplitude of the formant oscillation, possibly due to a second glottal excitation. In other cases, formant decay was truncated due to the opening of the glottis. Therefore, the accuracy of the bandwidth mea-

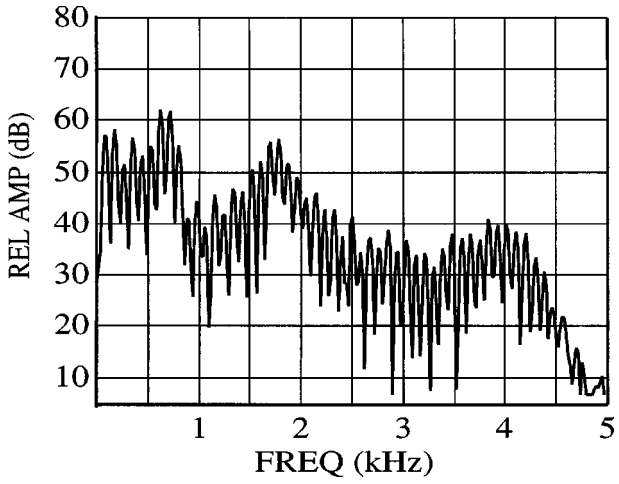

FIG. 5. Spectrum of the speech waveform in Fig. 4(a). The effect of the second pulse is evident in the attenuation of the alternating harmonics.

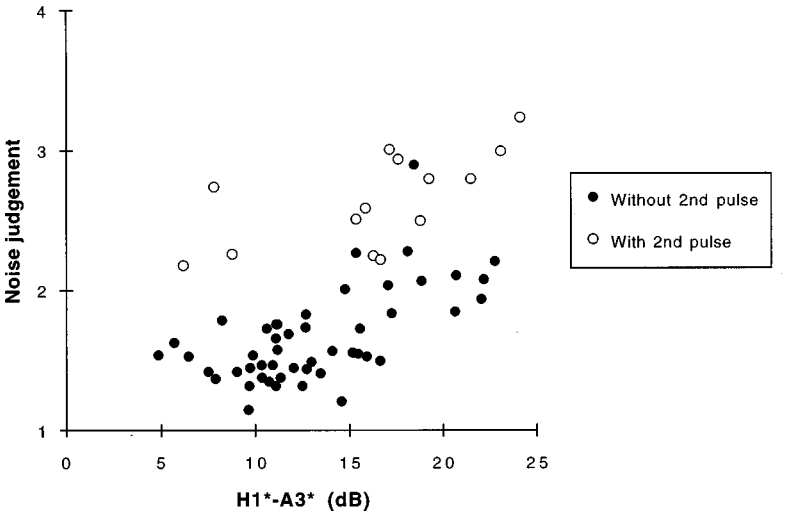

FIG. 6. Relation between the noise judgments, N_w, and the parameter $H1^*-A3^*$, for the male data. Each point represents data for one vowel of one speaker. The open circles indicate the cases that exhibited more evidence of second pulses.

surements is uncertain, which perhaps explains the low correlation between the measures $H1^*-A1$ and $B1$.

The complications described in this section suggest that the $F1$ bandwidth estimates and the waveform-based noise ratings must be interpreted with care for male speakers.

C. Comparison with female data

The acoustic measurements for male speakers were compared with the female data collected by Hanson (1995a, 1997). The maxima, minima, means, ranges, and average standard deviations of the acoustic parameters for male and female speakers are given in Table X. Histograms for each measure, illustrating the contrast between male and female speech, are shown in Fig. 7.

There are considerable differences between the mean values for female and male data for all measurements. The average $H1^*-H2^*$ for male and female speakers differs by about 3 dB. Previous studies which measured the relative amplitudes of $H1$ and $H2$ for both male and female speakers found somewhat greater differences. Klatt and Klatt (1990) found a difference of 5.7 dB between the genders, and Henton and Bladon (1985) also found that $H1-H2$ was about 6 dB greater for female speakers. However, the general trend is in agreement that female speakers have, on average, larger relative amplitudes of the first harmonic, suggesting that they have larger open quotients, as has been observed by Holmberg *et al.* (1988). Note that the range and standard deviation are also slightly larger for the female speakers, in agreement with previous data that females have a wider range of open quotient (Holmberg *et al.*, 1988). The histogram in Fig. 7(a) shows that the male and female data are fairly well separated, with only about a 4-dB overlap.

There is a highly significant difference between genders for $H1^*-A3^*$. The female speakers have an average of 23.4 dB, while the male speakers have an average of 13.8 dB, indicating that female speakers tend to have much weaker high-frequency content in the speech signal. This result is in agreement with the finding by Perkell *et al.* (1994) that males have a higher maximum flow declination rate (MFDR) than females. A 10-dB change in tilt is easily perceived, suggesting that spectral tilt may be an important parameter in differentiating male voice quality from female. In the histogram, Fig. 7(c), the female data are seen to be rather evenly spread throughout their range, while the male data are clustered around 8 to 12 dB.

For the measure $H1^*-A1$, the average difference between genders is around 3 dB, indicating that female speakers, on average, have a weaker $F1$ amplitude. In Fig. 7(b), the data are seen to be less well separated than those for $H1^*-H2^*$ and $H1^*-A3^*$. Out of a total range of 20 dB, significant overlap of the genders is about 10 dB. In Fig. 7(d), we see that females tend to have wider first-formant bandwidths than males, as was found experimentally by Fujimura and Lindqvist (1971). The male data are more tightly clustered than the female data.

The amount of aspiration noise is another acoustic cor-

TABLE X. Comparison of mean, maximum, and minimum values, and standard deviations of the acoustic parameters for male and female speakers. The measures $H1^*-H2^*$, $H1^*-A1$, and $H1^*-A3^*$ are given in dB and $B1$ is given in Hz. N_w is the waveform-based noise judgment. M indicates male data and F indicates female. [Female data from Hanson (1995a, 1997).]

	$H1^*-H2^*$		$H1^*-A1$		$H1^*-A3^*$		N_w		$B1$ (/æ/)	
	M	F	M	F	M	F	M	F	M	F
Mean	0.0	3.1	−6.9	−3.9	13.8	23.4	1.9	2.3	126	165
Minimum	−3.3	−2.6	−16.1	−12.4	4.8	8.6	1.2	1.1	53	53
Maximum	4.2	6.9	0.4	3.9	24.1	35.0	3.2	3.8	245	280
Range	7.5	9.5	16.5	16.3	19.3	26.4	2.0	2.7	192	227
s.d.	1.8	2.0	3.5	4.3	4.8	6.6	0.5	0.7	54	61

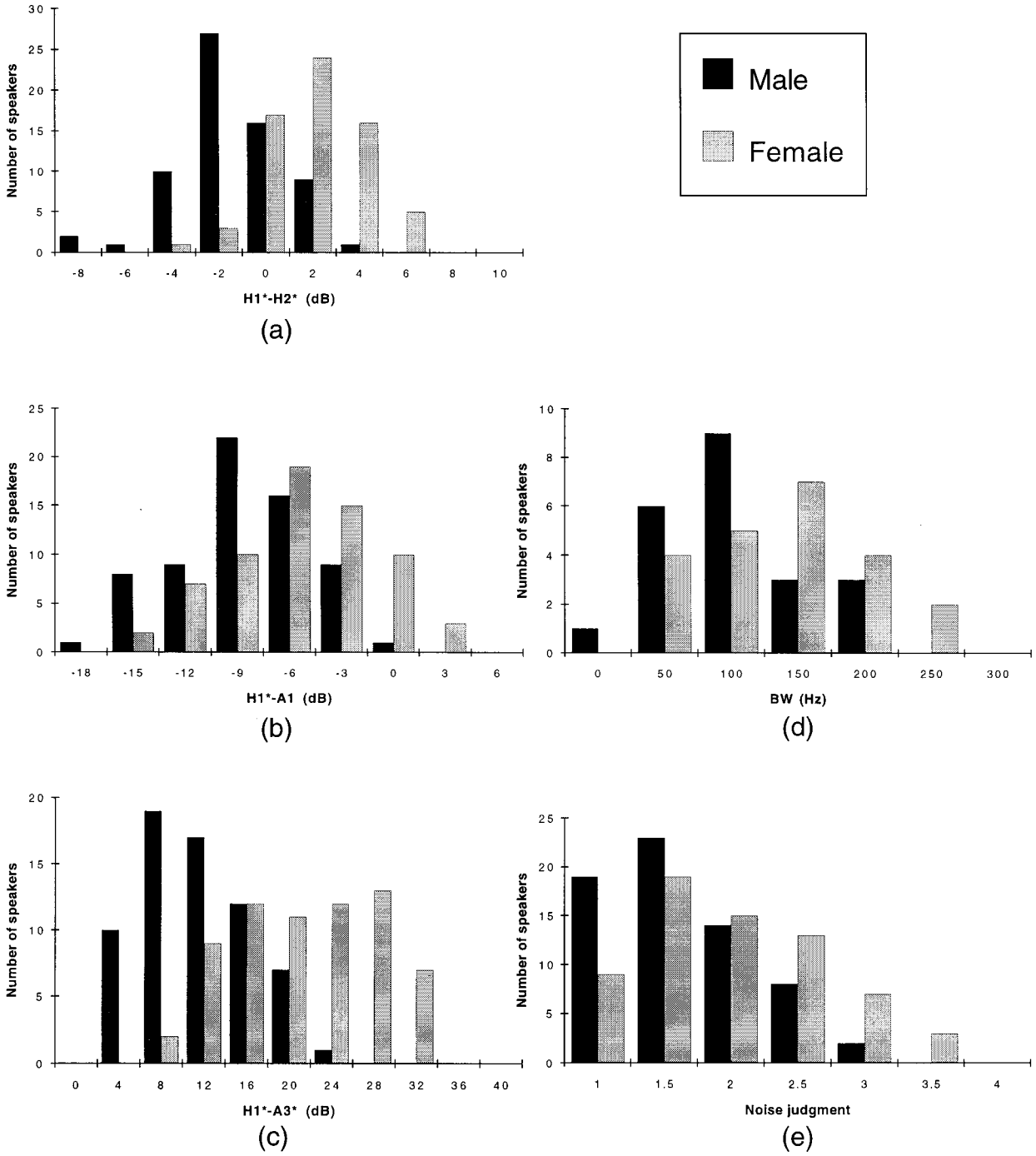

FIG. 7. Histograms of the acoustic measures for male and female speakers.

relate of the degree of breathiness of a vowel. Comparing the noise ratings of female and male speakers, we see that female speakers display more noise, on average, than male speakers, in the frequency range of the third formant. Although the average difference is not very large (1.9 for males vs 2.3 for females), as we discussed in Sec. III B it is likely that most male tokens given ratings of 2 or higher appeared to be noisy due to a second glottal excitation rather than to aspiration noise. In that case, the likelihood of females to have stronger aspiration noise than males is greater than is indicated by the data in Fig. 7(e) and Table X.

Figure 8 compares vowel spectra from male and female subjects having average values of the acoustic parameters. The gender differences discussed above are evident in these spectra: the female subject (upper plot) has less well-defined formants, steeper spectral tilt, more high-frequency noise, and a larger relative amplitude of the first harmonic.

Figure 9 plots the noise judgments against the measure $H1^*-A3^*$ for both male and female speakers. The data for the female speakers are divided in the two groups, group 1 and group 2, described in Sec. I A. The data of the male speakers are divided to indicate which tokens were judged to show evidence of having been produced with a second excitation. Most of the male data, with the exception of those given relatively higher noise ratings due to the second excitation, fall in the same range as the group 1 female data. This

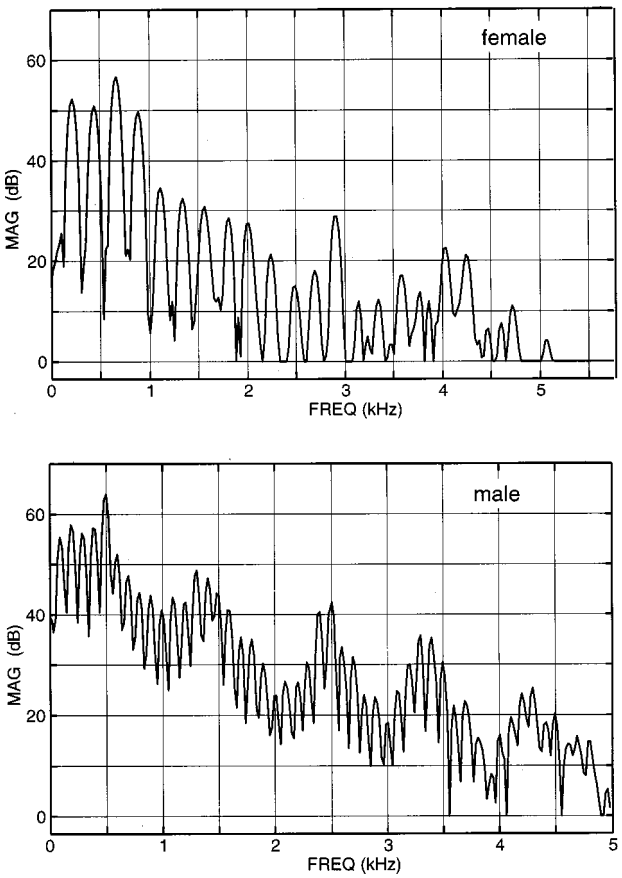

FIG. 8. Comparison of spectra of the vowel /ʌ/ for average female and male subjects. The female spectrum (upper plot) has greater spectral tilt, less well-defined formant peaks, a greater degree of noise at high frequencies, and a higher relative amplitude of the first harmonic, compared to the male speaker's vowel spectrum (lower plot).

result is not surprising if we recall that group 1 females were hypothesized to have abrupt glottal closure with relatively small posterior glottal chinks (Sec. I A). The male data that were marked by second excitations do not follow the trend set by the female data, providing further evidence that the high noise ratings for these vowel tokens do not truly reflect degree of aspiration noise.

Figure 10 shows the relationship between the measures $H1^*-A1$ and $H1^*-A3^*$. Most of the male data overlap with the group 1 female data. There is very little overlap of male data with the group 2 female data. This result agrees with the previous studies that the presence of a posterior glottal chink that persists throughout a glottal cycle is common for female speakers, while occurring much less frequently, or to a lesser degree, among male speakers (Södersten and Lindestad, 1990). Note also that, contrary to Fig. 9, the vowels judged to have been produced with a second excitation do not stand out in any way, but instead follow the trend set by the female data. This observation is further evidence that the noise judgments for these speakers are not representative of their glottal configurations.

It is well known that adult male speakers tend to have lower fundamental frequencies and formant frequencies than females. It is fair to ask how significantly these differences contributed to the gender differences reported. If the male formant frequencies are simply scaled versions of the female formants, the amplitudes of the peaks in the frequency response should remain the same. While that assumption is not entirely true, it is probably safe to assume that gender differences in formant frequencies have minimal effect on the acoustic measures considered here.

To analyze the effects of fundamental frequency differences, we consider the derivative of the glottal waveform, because that is the effective excitation; for simplicity we refer to this derivative as the source waveform $U(t)$. Let us first assume that the male source waveform $U_M(t)$ is simply the female source waveform $U_F(t)$ scaled in time, and therefore open quotient and waveform amplitude remain the same. On average, the pitch period of males is about twice that of females, and consequently we scale in time by a factor of 2, that is, $U_M(t)=U_F(t/2)$. In the frequency domain, $U_M(\omega)=2U_F(2\omega)$. There are two effects, then: a compression along the frequency axis and a scaling of spectrum amplitude. The compression in frequency and scaling in amplitude mean that a harmonic at 2500 Hz in the male spectrum will have the same amplitude as a harmonic at 5000 Hz in the female spectrum, plus 6 dB. At very low frequencies, where the source spectrum is relatively flat, the spectrum magnitude will be raised by 6 dB. At higher frequencies, the magnitude of the female spectrum falls off at 6 dB per octave; thus, at a given frequency ω,

$$20\log_{10}|U_M(\omega)| = 6 \text{ dB} + 20\log_{10}|U_F(2\omega)|$$

$$= 6 \text{ dB} + 20\log_{10}|U_F(\omega)| - 6 \text{ dB}.$$

The net effect of a simple waveform scaling, then, is that the lower-frequency harmonics increase in amplitude by 6 dB, but the amplitudes of the higher-frequency harmonics are unchanged. Therefore, all else being the same, the measure $H1^*-H2^*$ should be the same for males and females, but $H1^*-A1$ and $H1^*-A3^*$ should be 6 dB greater for males.

Our data suggest, however, that this model is not appropriate, and therefore the gender differences observed in our data are most likely due to details of the glottal configuration and waveform, and to vocal-tract losses, rather than to fundamental frequency and formant differences. In fact, we know from other experimental data that the male source waveform is not simply a time-scaled version of the female waveform. In particular, for male speakers the open quotient is smaller and the maximum flow declination rate (MFDR) is greater than for females (Holmberg *et al.*, 1988; Perkell *et al.*, 1994). Typical open quotients for males and females are 50% and 60%, respectively, leading to a gender difference of about 3 dB for the relative amplitudes of the first two harmonics, based on the KLGLOTT88 model of the glottal waveform (Klatt and Klatt, 1990). This difference is about the same as that found for our male and females subjects.

It is primarily the first harmonic that is affected by changes in open quotient, and therefore the predicted values of $H1^*-A1$ and $H1^*-A3^*$ for males relative to females are also reduced by about 3 dB. The MFDR, or negative peak of the flow derivative, mainly affects the spectrum well above the first harmonic (Fant, 1995). The higher MFDR of males should boost the amplitude of their formants, relative to those of females. Data reported in Perkell *et al.* (1994) pre-

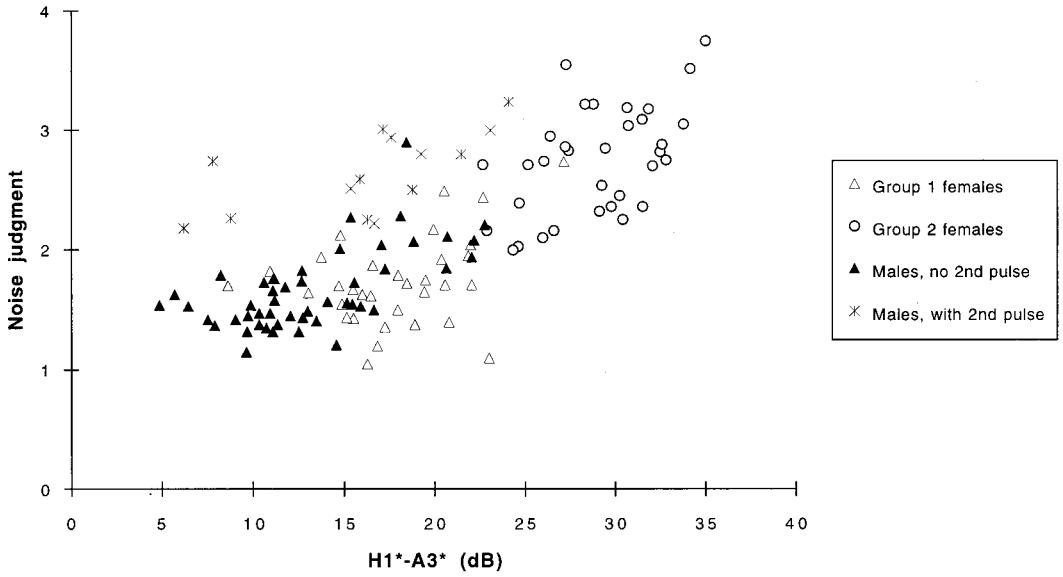

FIG. 9. Relation between noise judgment, N_w, and $H1^*-A3^*$ for the male and female data combined. Open triangles (△) represent group 1 female data, open circles (○) represent group 2 female data, filled triangles (▲) represent male data showing little evidence of second glottal excitations, and asterisks (*) represent male data showing significant evidence of second glottal excitations.

dict that the difference will be about 5.3 dB (for normal voice). The combination of the open quotient and MFDR effects with the time scaling, then, predicts that the measures $H1^*-A1$ and $H1^*-A3^*$ will be about 2.3 dB less for males than for females.

Our data show $H1^*-A1$ to be, on average, 3 dB less for males, and $H1^*-A3^*$ to be 9.6 dB less. Our first-formant bandwidth estimates (which include both glottal-chink and vocal-tract losses) would lead us to expect that $A1$ will be 2.3 dB higher for males than for females, in addition to the MFDR effect. The net effect is that $H1^*-A1$ should be 4.6 dB lower for males than for females, consistent with our finding. Third-formant bandwidth differences could increase the gender difference in $A3^*$ by 3 dB or so. It is reasonable to attribute the remaining 4.3-dB difference in the measure $H1^*-A3^*$ to the greater tendency for females to have posterior glottal chinks. Thus, we have shown that the observed gender differences are largely due to details of the glottal configuration and source waveform characteristics.

IV. SUMMARY

Vowel data were collected for 21 male speakers, and were analyzed using acoustic measures believed to reflect glottal configuration. Significant variations among the speakers were observed for all of the acoustic measures. The data were compared with female data collected in an earlier study. In agreement with predictions based on theoretical models

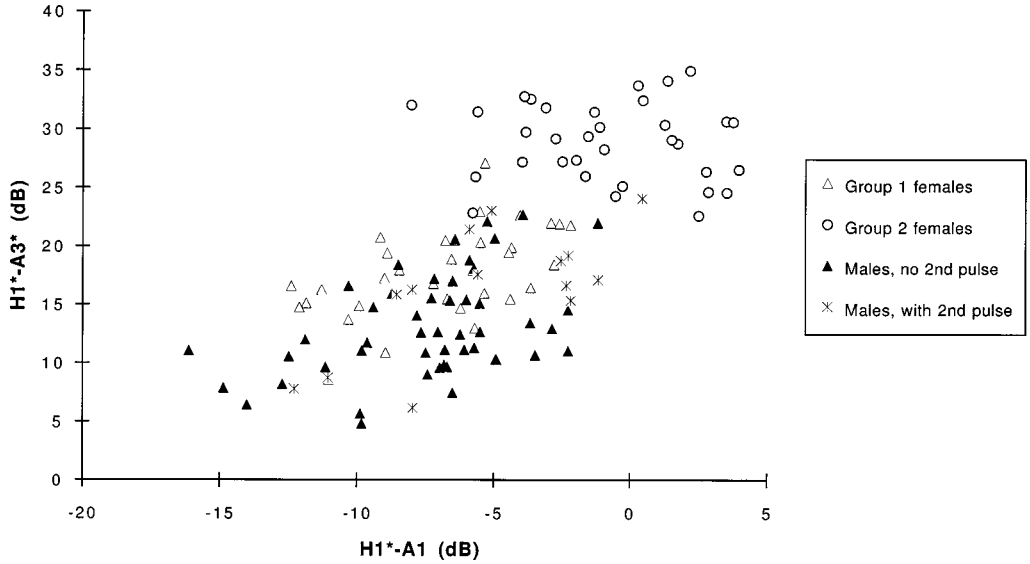

FIG. 10. Relation between $H1^*-A1$ and $H1^*-A3^*$ for the male and female data combined. Open triangles (△) represent group 1 female data, open circles (○) represent group 2 female data, filled triangles (▲) represent male data showing little evidence of second glottal excitations, and asterisks (*) represent male data showing significant evidence of second glottal excitations.

and previous reports of physiological and airflow data (Holmberg *et al.*, 1988; Södersten and Lindestad, 1990), we found that the group averages for all acoustic measures for the males were lower than those of the females. Although we found significant overlap of the male and female data, the male data tended to be more tightly clustered about the means. The greatest differences were observed for the measures $H1^*-H2^*$ and $H1^*-A3^*$, which reflect open quotient and source spectral tilt, respectively. Changes in tilt significantly affect perceived voice quality (see, for example, Klatt and Klatt, 1990), and therefore spectral tilt may greatly contribute to gender differences that we perceive in speech. Another gender difference that was observed in the speech waveforms is the presence of second excitation pulses for some of the male speakers. Such pulses were very rarely observed in the female data.

Some of the measures were found to be sensitive to second excitations during the glottal cycle, which were not apparent for all speakers. The ratings of high-frequency noise in the waveform in particular were affected, and the first-formant bandwidth estimates may also have been affected. The evidence for these secondary excitations on the speech waveform is hypothesized to be greatest for speakers with wide formant bandwidths. Therefore, some care must be taken when interpreting the acoustic measures.

Because the data illustrate normal variation of the source spectrum characteristics among individual speakers, and between males and females, they may be useful for applications such as speech synthesis, speech recognition, and speaker recognition. In addition, they may be used clinically, for example, to assess potential voice dysfunction or monitor remediation.

ACKNOWLEDGMENTS

This work was supported by Research Grants Nos. DC00075 and 1 F32 DC 00205-02 from the National Institute on Deafness and Other Communicative Disorders, National Institutes of Health. We are grateful to Kenneth N. Stevens and Eva Holmberg for their careful readings of an earlier version of the manuscript. Special thanks to Ken Stevens for help in working through the theoretical issues involved in comparing male and female acoustic measures. The comments and suggestions of Anders Löfqvist, Ronald C. Scherer, and an anonymous reviewer are also greatly appreciated.

Atal, B. S., and Remde, J. R. (**1982**). ''A new model of LPC excitation for producing natural-sounding speech at low bit rates,'' In Proceedings of the IEEE ICASSP '82, Paris, France, pp. 614–617.

Childers, D. G., and Wu, K. (**1991**). ''Gender recognition from speech. Part II: Fine analysis,'' J. Acoust. Soc. Am. **90**, 1841–1856.

Fant, G. (**1972**). ''Vocal tract wall effects, losses, and resonance bandwidths,'' Speech Trans. Lab. Q. Prog. Stat. Report 2–3, Royal Institute of Technology, Stockholm, pp. 28–52.

Fant, G. (**1995**). ''The LF-model revisited. Transformations and frequency domain analysis,'' Speech Trans. Lab. Q. Prog. Stat. Report 2–3, Royal Institute of Technology, Stockholm, pp. 119–156.

Fujimura, O., and Lindqvist, J. (**1971**). ''Sweep-tone measurements of vocal-tract characteristics,'' J. Acoust. Soc. Am. **49**, 541–558.

Hanson, H. M. (**1995a**). ''Glottal characteristics of female speakers,'' Ph.D. thesis, Harvard University, Cambridge, MA.

Hanson, H. M. (**1995b**). ''Synthesis of female speech using the Klatt synthesizer,'' Speech Communication Group Working Papers 10, Research Laboratory of Electronics, M.I.T., pp. 84–103.

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: Acoustic correlates,'' J. Acoust. Soc. Am. **101**, 466–481.

Henton, C. G., and Bladon, R. A. W. (**1985**). ''Breathiness in normal female speech: Inefficiency versus desirability,'' Lang. Commun. **5**, 221–227.

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. Plus Erratum, *ibid.* **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., 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.

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. (**1992a**). ''Analysis and synthesis of different voices with emphasis on female speech,'' Ph.D. thesis, Royal Institute of Technology, Stockholm.

Karlsson, I. (**1992b**). ''Evaluations of acoustic differences between male and female voices: A pilot study,'' Speech Trans. Lab. Q. Prog. Stat. Report 1, Royal Institute of Technology, Stockholm, pp. 19–31.

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.

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.

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.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Titze, I. R., (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.