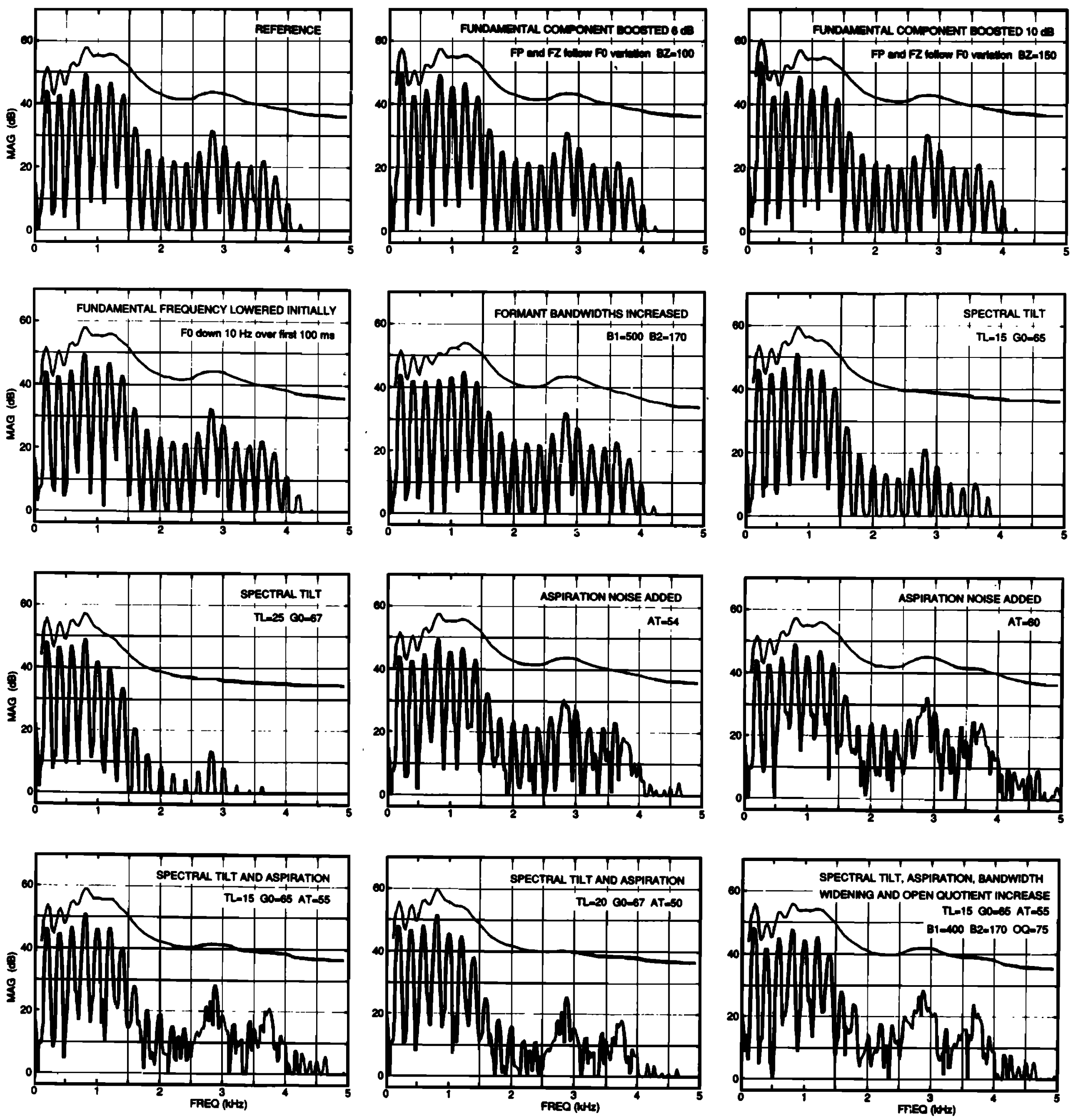

FIG. 18. Spectral cross sections (no pre-emphasis) for each of the synthetic stimuli used in the breathiness perception test.

and second vowel) to 5 (maximal increase in breathiness). If a vowel actually sounds less breathy that the reference, a negative number, such as − 1, can be used as a response. Try not to downgrade ratings of breathiness for unnatural stimuli; simply wait until the second playing of the tape to express dissatisfaction with the stimulus. The first ten trials will be treated as practice trials in order to allow you to hear the range of stimuli to be encountered; nevertheless, please respond and write down a practice answer for these trials.

Instructions before second playing: The naturalness rating scale to be used will go from 0 to 5, with 5 being perfectly natural, and zero being very unnatural or machinelike (not possible for a human to imitate through any natural speech production process, probably produced by artificial means). Remember that we would also like you to star any item that sounds nasalized."

Subjects listened over TDH model HD-420 earphones to a randomized tape recording that included a practice run of 14 trials to familiarize subjects with the expected range of breathiness, and five blocks of trials that were scored and combined to obtain average ratings of breathiness for each stimulus.

Five subjects participated individually in the perception test. The results are summarized in Table XV, which gives the average breathiness rating response of each listener, the average naturalness rating, and the fraction of times a listener indicated nasality was present. Also included in the table are group averages. These experienced listeners were re-

TABLE XV. Listener ratings of: (1) change in breathiness relative to the reference stimulus, (2) naturalness, and (3) nasality for each of the 11 synthetic stimuli that were contrasted with the reference vowel in the breathiness perception test.

Change in breathiness (− 5.0 to 5.0): Condition	CB	SH	SM	SB	KS	Av
(1) Fundamental component boosted 6 dB	1.2	0.0	1.6	0.4	1.4	0.92
(2) Fundamental component boosted 10 dB	2.0	− 0.7	2.2	0.2	1.6	1.26
(3) Fundamental frequency lowered initially	0.0	0.0	− 0.4	0.0	0.6	0.04
(4) Formant bandwidths increased	0.8	− 0.1	0.6	0.4	0.6	0.46
(5) Spectral tilt down 15 dB at 3 kHz	1.8	− 1.2	2.2	0.2	2.0	1.00
(6) Spectral tilt down 25 dB at 3 kHz	2.8	− 1.8	2.0	0.6	3.2	1.36
(7) Aspiration noise of 54 dB added	0.4	1.3	2.0	0.0	2.0	1.14
(8) Aspiration noise of 60 dB added	1.8	2.4	3.0	2.6	4.4	2.88
(9) Spectral tilt of 15 dB and aspiration of 55 dB	3.4	1.3	3.2	1.4	4.2	2.70
(10) Spectral tilt of 20 dB and aspiration of 50 dB	3.8	1.4	2.8	0.8	4.4	2.64
(11) Ditto, plus bandwidth widening and OQ increase	4.6	3.0	3.4	3.0	4.8	3.76

Naturalness (0 to 5.0): Condition	CB	SH	SM	SB	KS	Av
(1) Fundamental component boosted 6 dB	4.8	5.0	2.8	4.0	2.4	3.80
(2) Fundamental component boosted 10 dB	3.8	5.0	2.6	2.4	1.6	3.08
(3) Fundamental frequency lowered initially	5.0	5.0	2.6	4.8	4.4	4.36
(4) Formant bandwidths increased	2.4	4.2	1.2	5.0	2.8	3.12
(5) Spectral tilt down 15 dB at 3 kHz	3.4	5.0	3.6	4.0	3.4	3.88
(6) Spectral tilt down 25 dB at 3 kHz	1.8	5.0	3.6	3.2	2.4	3.20
(7) Aspiration noise of 54 dB added	5.0	5.0	3.8	5.0	3.8	4.52
(8) Aspiration noise of 60 dB added	4.0	5.0	4.2	5.0	2.4	4.12
(9) Spectral tilt of 15 dB and aspiration of 55 dB	4.4	5.0	4.4	3.4	4.2	4.28
(10) Spectral tilt of 20 dB and aspiration of 50 dB	3.4	5.0	4.8	4.6	4.2	4.40
(11) Ditto, plus bandwidth widening and OQ increase	5.0	5.0	5.0	2.4	3.0	4.08

Nasality (0 to 1.0): Condition	CB	SH	SM	SB	KS	Av
(1) Fundamental component boosted 6 dB	0.4	0.8	0.8	0.2	0.6	0.56
(2) Fundamental component boosted 10 dB	0.4	1.0	1.0	0.6	0.6	0.72
(3) Fundamental frequency lowered initially	0.0	0.0	0.2	0.0	0.0	0.04
(4) Formant bandwidths increased	1.0	0.8	1.0	0.2	1.0	0.80
(5) Spectral tilt down 15 dB at 3 kHz	0.4	0.0	0.2	1.0	0.0	0.32
(6) Spectral tilt down 25 dB at 3 kHz	0.6	0.2	0.0	1.0	0.0	0.36
(7) Aspiration noise of 54 dB added	0.0	0.0	0.0	0.0	0.0	0.00
(8) Aspiration noise of 60 dB added	0.0	0.0	0.2	0.0	0.0	0.04
(9) Spectral tilt of 15 dB and aspiration of 55 dB	0.0	0.0	0.0	0.4	0.0	0.08
(10) Spectral tilt of 20 dB and aspiration of 50 dB	0.6	0.0	0.0	1.0	0.0	0.32
(11) Ditto, plus bandwidth widening and OQ increase	0.0	0.0	0.0	0.4	0.0	0.08

markably self consistent; except for SB, they rarely differed by more than 1 in assigning an integer to ratings of breathiness and naturalness of five repetitions of each condition.

While there are some intersubject differences, the pooled results point to aspiration amplitude as the dominant factor in eliciting judgments of increased breathiness. An increase in first-harmonic amplitude, by itself, does not induce the sensation of breathiness for most listeners. The reason is probably the high fundamental frequency employed in the synthesis; an increase in the spectrum at about 200 Hz is consistent with the appearance of a nasal pole indicative of nasalization. In fact, nasalization judgments are quite common for stimuli in which only the fundamental component has been increased. We tentatively conclude that either breathiness is signaled differently for men and women, or that the increases in the first harmonic observed in production data from women must be accompanied by other cues to be interpreted by the listener as cues to breathiness (see below).

An increase in the first-formant bandwidth, stimulus 4, is also by itself ineffectual in suggesting breathiness. The stimulus sounds both nasal and unnatural. Again, it is either the case that production data (indicative of increased bandwidths for breathy vowels) do not comport with perceptual strategies, or that the bandwidth increases must be accompanied by certain other cues before they are unambiguously interpreted as relating to breathiness.

A lowering of f_0 had no effect on perceived breathiness. Tilting down the spectrum to reduce the amplitudes of higher frequency harmonics increased breathiness judgments a little bit for some listeners, but was heard as unnatural or as an increase in nasality for others. By itself, spectral tilt does not appear to be a strong cue to breathiness, perhaps again because it does not occur naturally by itself, but rather only in conjunction with certain other cues to breathiness.

The addition of aspiration noise increases the number of breathiness judgments significantly for most listeners. If the harmonic spectrum is also attenuated at high frequencies by tilting the spectrum down (as is observed in natural breathy vowels), then less noise is required to achieve the same degree of perceived breathiness. Stimuli with aspiration noise were all judged to be natural.

The stimulus perceived to be most breathy was the one in which all of the various cues observed in natural stimuli were present: aspiration noise, spectral tilt, longer open quotient, and increased bandwidths of $F1$ and $F2$. The increase in first-harmonic amplitude and the increases in bandwidths that, by themselves, induced the sensation of nasality for many listeners, did not produce nasality judgments for this stimulus. It appears that nasality perception is a rather complex function of acoustic properties that are attributed to breathiness under some circumstances and attributed to nasality in others. This outcome is troublesome for many simple models of speech perception but is in agreement with the philosophy known as the motor theory of speech perception (Liberman *et al.*, 1967) and other models that permit learning of complex cue interactions (Klatt, 1986a, 1989).

V. DISCUSSION AND CONCLUSIONS

The analysis of reiterant speech from ten female and six male speakers has revealed a number of acoustic cues related to breathiness. A perception test has established the importance of aspiration noise as a component of a breathy voice quality and has shown the complexity of some cue interaction perceptual strategies. Synthesis efforts using a new version of the Klatt formant synthesizer, KLSYN88, which has a new voicing source model, verify that it is possible to mimic several female voices with an accuracy that makes it difficult to distinguish between the original recording and the synthesis. The following sections go into greater detail on each of these topics and speculate as to the implications of several of our results.

A. Acoustic analysis

A breathy voice quality is signaled by a surprisingly large number of diverse acoustic cues, all related to the presumed posterior glottal opening posture shown in panel (3A) of Fig. 1. First of all, the posterior opening leads to a dc flow component and the generation of aspiration noise throughout a period, with noise intensity perhaps increasing during the open phase. The open quotient is increased, and this leads to a relative increase in the amplitude of the first harmonic, H1, by 6 dB or more. In addition, there is nonsimultaneous closure of the folds over their length, with the posterior portion of the folds making contact somewhat later than the anterior edges; this pattern of closure leads to a reduction in the relative amplitudes of higher harmonics in the source spectrum. As a result of these two factors, aspiration noise tends to replace harmonics at frequencies above about 1.5 kHz in a breathy vowel. The posterior glottal opening increases low-frequency losses in the vocal-tract transfer function, resulting in an increased first-formant bandwidth and a less distinct first-formant peak in the spectrum. The posterior glottal opening also provides acoustic coupling to the subglottal system, resulting in the possible appearance of tracheal poles and zeros in the vocal-tract transfer function. The poles tend to appear at predictable frequency locations—about 600, 1400, and 2200 Hz for a female voice and somewhat lower for males.

The degree to which an individual vowel takes on the cues of breathiness can vary considerably over the course of an utterance. In our data, cues to breathiness tend to increase for unstressed syllables, for final syllables, and at the margins of voiceless consonants. In a stressed vowel with a relatively high fundamental frequency, the spectrum may be perfectly periodic, even for the most breathy of our speakers. It is almost certainly this time variation that contributes to naturalness and highlights a breathy voice.

Males and females differ on average in the two perceptually most important acoustic measures of breathiness—amount of aspiration noise in the $F3$ region of the spectrum and relative amplitude of the first harmonic. Females are more breathy than males to a significant degree. However, within each gender, there is much greater variation in acoustic manifestations of breathiness, with some males being more breathy than many females. In addition, it is likely that any individual is capable of adopting a fairly wide range of speaking styles that differ in degree of breathiness. Thus it is dangerous to make sweeping generalizations with regard to sex typing, as well as the behavior of particular individuals.

We have discovered evidence for a breathy-laryngealized mode of vibration that is employed by many speakers at the ends of utterances. There is increased noise in the $F3$ region of the spectrum, but the open quotient does not increase, as it normally would whenever the posterior portions of the folds are spread in preparation for a voiceless consonant or for breathing. We speculate that the arytenoid cartilages are rotated inward to facilitate the maintenance of voicing in the face of a developing posterior glottal chink, a lowering subglottal pressure in anticipation of the end of speaking, and a slack vocal-fold posture appropriate for low f_0. This breathy-laryngealized vibration pattern will have to be incorporated into representational schemes for the phonetic description of speech and may possibly require changes to the distinctive features used to represent language. For example, the feature system of Halle and Stevens (1971) might be revised to allow simultaneous plus values for the features *spread* and *constricted* to represent a breathy-laryngealized voice quality, where spread is redefined to refer to the posterior interarytenoid separation, and constricted to refer to the medial compression resulting from rotational motion of the arytenoids.

A two-mass model of vocal-fold vibrational behavior was devised by Ishizaka and Matsudaira (1968) in order to better explain the transfer of energy from static lung pressure to dynamic vibratory motions of the vocal folds. While successful in these terms (Stevens, 1977), the model does not have the flexibility to generate output volume velocity waveforms with corner rounding associated with nonsimultaneous closure. Neither does the model permit a static posterior glottal opening necessary to simulate breathy voicing. Therefore, use of the model in synthesis of speech (Flanagan *et al.*, 1975) is likely to lead to suboptimal synthetic imitations of female voices until such time as these features of laryngeal behavior are incorporated.

A new measure of the amount of aspiration noise in a voiced sound has been proposed. By filtering out a region of the spectrum in the vicinity of $F3$, it is possible to isolate a waveform that can be visually interpreted in terms of whether the source is periodic or random. While we have used subjective rating scales to quantify this judgment, further research, such as quantifying the peak-to-valley ratio in the energy contour over a period, may make it possible to automate its measurement and thereby remove the subjective component.

Formant-frequency measurement can be very difficult in a breathy vowel, particularly in a voice with a high fundamental frequency. The strong first harmonic can be confused with a formant; the first-formant bandwidth increase can make it difficult to detect a local maximum in the spectrum corresponding to $F1$, and the tracheal coupling can cause extra peaks in the spectrum that are easily confused with formants. When harmonics are widely spaced in the spectrum, each of these complications becomes more difficult to deal with because the vocal-tract transfer function is essentially sampled only at harmonic locations. The remarkable ability of the human perceptual system to deal with these problems calls into question the degree to which formants are actually perceptual dimensions (Bladon, 1982). Alternatives include whole-spectrum template-matching approaches (Klatt, 1982), and complex "spectrum-interpretive" strategies, a discussion of which goes beyond the scope of the present paper.

Vowels in natural utterances are rarely perfectly periodic. But it has been difficult to characterize exactly how individual periods differ from perfect periodicity. A popular theory has been that there is a jitter component, or Gaussian random fluctuation in individual periods, but efforts to turn this into an effective synthesis strategy have failed. Such jitter is either imperceptible if added in amount corresponding to prior literature on the measurement of jitter, or sounds like a harsh pathological voice quality when jitter is increased. Our observations on deviations from perfect periodicity in this database suggest two new methods of synthesizing natural deviations: (1) a slowly varying "pseudorandom flutter" consisting of a sum of sine waves and (2) an optional diplophonic double pulsing that occurs in certain fairly predictable situations. The perceptual importance of temporal variability and the success of our proposed synthesis strategy have yet to be established. Also, in order to better account for diplophonia in terms of the physics of larynx behavior, a more complex physiological model may be required in which the three-dimensional nature of the vibration pattern is considered (Titze, 1974; Titze and Talkin, 1979).

Our results must be qualified by the limited scope and artificial nature of the data base. We have analyzed only one vowel in a sample of reiterant speech approximating two sentences. The speech is read with regular pauses between utterances. In the future, other vowels should be employed, and the analysis techniques should be extended to more natural databases of spontaneous speech.

B. Perception of breathiness

The perception test using natural speech samples edited from the reiterant sentences revealed that females, on average, are perceived to be slightly more breathy than males. The perception data also are in agreement with the acoustic data in showing wide variation within each gender; two females are judged less breathy than the male average, and one male is judged more breathy than the female average. A correlation analysis with ten acoustic parameters related to breathiness revealed only two statistically significant correlations with the perceptual responses—one with the degree of aspiration noise seen in the $F3$ region of the spectrum, and one with the relative strength of the first harmonic H1.

The second type of perception test used synthetic speech in order to be able to determine the perceptual importance of individual acoustic cues in isolation and in combination. This test was somewhat unusual in that the standard comparison stimulus was patterned after a female voice rather than using a male voice as has been the practice so often in the past. Stimuli were judged as to degree of breathiness, naturalness, and nasality. The strongest single cue to breathiness was found to be the amplitude of the aspiration noise added to the spectrum. However, the stimulus that was most preferred in terms of breathiness and naturalness was one in which all cues (add aspiration noise, increase spectral tilt, increase open quotient to increase H1, and widen first-formant bandwidth) were present. It is as if the perceiver is aware of all of the systematic changes that go into breathy phonation and uses these expectations during perception in such a way that no single cue is as effective as all in combination.

When only the first-harmonic amplitude was increased, some subjects heard an increase in breathiness, but many others heard an increase in nasality. While this result has never been reported in previous perceptual studies of H1, it may well be due to the fact that we simulated a female voice in which the first harmonic, on average, was about 200 Hz. This is close to the frequency of the lowest pole in the transfer function of a nasalized vowel. The same increase in amplitude of the first harmonic, if accompanied by aspiration noise, is never heard as nasalized. The implication is probably not that different perceptual strategies are employed for male and female voices. Rather, it is more likely that single-cue manipulations can create somewhat unnatural stimuli that result in perceptual ambiguities. This is also observed for first-formant bandwidth increases, which by themselves increase the nasality of a stimulus, but taken in conjunction with other cues to breathiness are not heard as nasalized.

It is interesting to speculate on the detailed nature of perceptual strategies, given the ambiguity introduced by a strong first-harmonic amplitude. Is first-harmonic amplitude a perceptual cue whose interpretation involves a complex interaction with other perceptual cues such as the degree of periodicity at high frequencies, or are spectra interpreted in a wholistic fashion against templates representing nasalized or breathy versions of various speech sounds? Neither alternative is very compelling to us as a perceptual strategy. The first option seems undesirable because, from our perspective, speech perception ought to be simple and direct at the lowest levels, with little if any cognitive processing. The second option involving a template approach may not be workable for various reasons having to do with natural variability within and across speakers (Klatt, 1986a, 1989).

C. Synthesis of a female voice

The standard acoustic theory of speech production has been reviewed in light of several source–tract interactive phenomena that have been identified by Gunnar Fant and his co-workers. Based on this theoretical background and the experimental data presented here, we conclude that the old cascade/parallel formant synthesizer model described in Klatt (1980), while still useful for most perception experiments involving cues to various consonant and vowel distinctions, is not capable of mimicking a female voice very accurately. In this paper, we have identified certain cues related to a breathy voice quality that must be modeled in order to closely mimic most female voices. The new version of the Klatt cascade/parallel synthesizer, described here, has been augmented with: (1) a new voicing source model that has control parameters **F0, AV, OQ, TL, AH, FL, DI**, (2) an ability to change first-formant bandwidth dynamically over a period, using **DB1** to simulate the rapid change in glottal losses as the glottis opens and closes, and (3) an additional tracheal pole-zero pair with control parameters **FTP, FTZ, BTP**, and **BTZ**. This new version, KLSYN88, has been used to copy reiterant utterances from several female and male speakers with very good perceptual fidelity.

Experience with the new model suggests that the ability to make dynamic changes to the degree of aspiration noise intruding in the $F3$ region of the spectrum through use of the **AH** and **TL** parameters is the most important aspect of the model for improving the quality of female voice synthesis, but that all of the new parameters are useful in optimizing the match to individual spectra. If analysis by synthesis is used, and a good initial match to the f_0 contour is achieved, then successful matching of spectra sampled throughout an utterance has always resulted in perceptually successful synthesis using the new model. General rules for synthesizing breathiness variations over a sentence have been proposed based on the voice analysis and voice-matching efforts reported on here. Cues to breathiness are strongest at the end of an utterance, in unstressed syllables, and at the margins of voiceless consonants. Utilization of these rules may lead to improved naturalness of both male and female voices in synthesis-by-rule programs for text-to-speech devices such as DECtalk.

Variation in the timing of glottal pulses, using the new voicing source control parameters **FL** and **DI** may also increase the naturalness of synthetic speech, but at this point, we have not conducted the appropriate perceptual experiments to determine the importance of the timing variation informally observed in the reiterant corpus. Another random variation that might be added to synthesis by rule is a small random change in successive values of the **TL** parameter to simulate the observed tendency for rapid period-to-period variation in harmonic strengths above 2 kHz, even in spectra that do not have a strong noise component in this frequency region.

The KLSYN88 cascade/parallel formant synthesizer has been programmed in floating point in C on a Digital Equipment Corporation Microvax-II. This software will be made available to the speech community at cost.[19] It is hoped that the synthesizer can serve as a standard in preparing stimuli for perceptual experiments, and the fact that the algorithms are documented here will make it easier for researchers to replicate and extend the findings of others.

ACKNOWLEDGMENTS

This research was supported by Grant NS04332 from the National Institutes of Health. We are very grateful to Kenneth Stevens for many helpful comments and editorial assistance.

[1] Unfortunately, there is at present no standard terminology or agreement on the meaning of many of these terms.

[2] Should creak be defined to begin at a higher threshold for females, say 1.7 times 60 Hz, or about 100 Hz, and, if so, would they show as much creak as males? The decision depends on whether the definition of creak is perceptual (such as "picket fence" percept) or physiological (so much below normal pitch). In this paper, we will assume that creak refers to the absolutely low pitch sensation where individual pulses are audible, and laryngealization, on the other hand, occurs at some percentage change below a speaker's normal f_0 range (or perhaps when the glottal pulse becomes very narrow such that H1 is reduced in amplitude).

[3] The amplitude of the strongest harmonic near $F1$ was used to define A1, the level of $F1$, for reference purposes. This definition may underestimate the level of the vocal-tract transfer function at $F1$ by as much as 6 to 9 dB when two harmonics straddle $F1$ (Klatt, 1986a).

[4] In order to be able to compare breathiness of vowels with different phonetic quality, and thus different $F1$ values, the authors decided to abandon the use of A1, the level of $F1$, as a reference, in spite of the fact that it should tend to somewhat accentuate the difference between breathy and nonbreathy phonation due to the increase in first-formant bandwidth associated with a partially open glottis.

[5] There are two problems with the noise measure chosen: (1) Lower formants dominate waveform characteristics because they are more intense, but aspiration noise, whose presence is to be detected, tends to be restricted to higher formants and (2) when sampling speech, frequency components near the high-frequency sampling limit are represented by only a few sample points per cycle so that, if the fundamental frequency is not an exact multiple of the sampling interval, even a perfectly periodic waveform will appear to be different from period to period, and thus contaminated with noise according to this measure. The first problem could be solved by highpass filtering at about 1.5 kHz, but the second problem is less amenable to a straightforward solution other than to significantly increase the sampling rate and thus decrease the granularity with which one compares periods.

[6] The analysis data that we will present suggest that aspiration noise might be a much more effective perceptual cue if the higher harmonics are simultaneously attenuated, as occurs for nonsimultaneous closure of the vocal folds along their length.

[7] In a spectrogram, the vertical striations indicative of periodic excitation would be replaced by a more random pattern if the excitation were exclusively turbulence noise. However, for relatively high-pitched female voices, the vertical striations are not as evident as for a male voice even when excitation is completely periodic.

[8] The amplitude of the second harmonic depends, in part, on locations of zeroes in the source spectrum and thus is not entirely satisfactory for reference purposes. However, the problems with the other alternatives seem more serious. The amplitude of the first formant (i.e., the peak in the underlying vocal-tract transfer function) is hard to determine from a harmonic spectrum because measured values depend to a considerable extent on whether a harmonic is centered on the formant frequency or two harmonics straddle the formant (Klatt, 1986a). Since the overall rms amplitude depends primarily on first-formant amplitude, it too suffers from this source of unpredictable variability, which can be up to 6–9 dB.

[9] Extreme changes to the amplitude of the fundamental component may be due to factors other than simply the open quotient. For example, a general reduction in the tilt of the spectrum would result if the vocal-fold closure event became nonsimultaneous due to glottal abduction. This effect can be simulated in a speech synthesizer by either the spectral tilt parameter TL

or the open quotient parameter OQ, as will be discussed in Section II below.

[10] Other measures described below will rate TW as even more breathy.

[11] Alternatively, one might use a two-pole inverse filter, but this requires fairly precise detection of the frequency and bandwidth of the third formant versus time. Such precision is not as critical when employing a 600-Hz bandwidth bandpass filter.

[12] Fant and Lin (1987) have attributed the extra pole between $F2$ and $F3$ to nonlinear superposition effects, but it seems to us to be more likely that the pole is due to tracheal resonance coupling, especially since it is seen both in aspiration and voiced excitation.

[13] Theoretically, the tracheal poles and zeros should come in pairs, where the zero is lower in frequency than the corresponding pole. However, it is difficult to detect evidence of a zero below the first tracheal pole due to the rapid falloff in energy at low frequencies.

[14] Detailed documentation of the characteristics of this source will not be given here since it was not used in the experiments to be described later in Secs. III and IV. The characteristics of the LF source are described in documentation available in the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology.

[15] The fundamental frequency extraction algorithm that was employed to produce this contour is of the harmonic sieve type, which probably averages out some rapid period-to-period changes within the 25-ms analysis window rather than accentuating them.

[16] In an analogous fashion, one can use the newly defined tracheal pole-zero pair to simulate any observed second nasal resonance in a nasalized vowel.

[17] The f_0 is specified in tenths of a hertz in order to minimize perceptual "staircase" effects for slowly changing fundamental frequency contours. The period that is computed is quantized into quarter-of-a-sample increments, again to avoid "staircase" effects; this is accomplished by running the glottal source code at four times the regular sampling rate of 10 000 samples per second, and then low-pass filtering and downsampling the resulting voicing waveform.

[18] For these utterances, the update interval was set to 10 ms, but internal to the synthesizer, AV is reset only at the beginning of each pitch period so as to avoid waveform discontinuities.

[19] Information about the availability of this software can be obtained from Kenneth Stevens, Room 36-517, Massachusetts Institute of Technology, Cambridge, MA 02139.

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Edinburgh U.P., Edinburgh).

Allen, D. R., and Strong, W. J. (**1985**). "A Model for the Synthesis of Natural Sounding Vowels," J. Acoust. Soc. Am. **78**, 58–69.

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., and Fant, G. (**1982**). "Calculation of True Glottal Flow and Its Components," Speech Commun. **1**, 167–184.

Askenfelt, A., and Hammarberg, B. (**1981**). "Speech Waveform Perturbation Analysis Revisited," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 49–68.

Askenfelt, A., and Hammarberg, B. (**1986**). "Speech Waveform Perturbation Analysis: A Perceptual-Acoustical Comparison of Seven Measures," J. Speech Hear. Res. **29**, 50–64.

Baer, T. (**1978**). "Effect of Single-Motor-Unit Firings on Fundamental Frequency of Phonation," J. Acoust. Soc. Am. Suppl. 1, **64**, S90.

Berg, J. W. van den (**1960**). "An Electrical Analog of the Trachea, Lungs and Tissues," Acta Physiol. Parmacol. Neerandica **9**, 1–24.

Bickley, C. (**1982**). "Acoustic Analysis and Perception of Breathy Vowels," Speech Commun. Group Work. Papers I, Research Laboratory of Electronics, MIT, Cambridge, MA, 71–82.

Bickley, C., and Stevens, K. N. (**1986**). "Effect of a Vocal Tract Constriction on the Glottal Source: Experimental and Modeling Studies," J. Phon. **14**, 373–382.

Bladon, R. A. W. (**1982**). "Arguments against Formants in the Auditory Representation of Speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 95–102.

Bless, D. M., Biever, D., and Shaikh, A. (**1986**). "Comparisons of Vibratory Characteristics of Young Adult Males and Females," Proceedings of International Conference on Voice, Kurume, Japan, Vol. 2, 46–54.

Brend, R. M. (**1975**). "Male–Female Intonation Patterns in American English," in *Language and Sex: Difference and Dominance*, edited by B. Thorn and N. Henley (Newbury House, Rowley, MA), pp. 84–87.

Catford, J. C. (**1964**). "Phonation Types: The Classification of Some Laryngeal Components of Speech Production," in *In Honour of Daniel Jones*, edited by D. Abercrombie, D. B. Fry, P. A. D. McCarthy, N. C. Scott, and J. L. M. Trim (Longmans, London), pp. 26–37.

Catford, J. C. (**1977**). *Fundamental Problems in Phonetics* (Indiana U.P., Bloomington, IN).

Chapin-Ringo, C. (**1988**). "Enhanced Amplitude of the First Harmonic as a Correlate of Voicelessness in Aspirated Consonants," J. Acoust. Soc. Am. Suppl. 1 **83**, S71.

Chasaide, A. (**1987**). "Glottal Control of Aspiration and of Voicelessness," Proceedings of Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia, Vol. 6, 28–31.

Chasaide, A., and Gobl, C. (**1987**). "Cross Language Study of the Effects of Voiced/Voiceless Consonants on the Vowel Voice Source Characteristics," J. Acoust. Soc. Am. Suppl. 1 **82**, S116.

Cleveland, T., and Sundberg, J. (**1983**). "Acoustic Analysis of Three Male Voices of Different Quality," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 27–38.

Cooper, W. E., and Sorenson, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer, New York).

Cranen, B., and Boves, L. (**1987**). "On Subglottal Formant Analysis," J. Acoust. Soc. Am. **81**, 734–746.

Dixit, R. P. (**1987**). "In Defense of the Phonetic Adequacy of the Traditional Term 'Voiced Aspirated,' " Proceedings of the Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia, Vol. 2, 145–148.

Dolansky, L., and Tjernlund, P. (**1968**). "On Certain Irregularities of Voiced Speech Waveforms," IEEE Trans. Audio Electroacoust. **AU-16**, 51–56.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of Pitch in Speech: An Implementation of Goldstein's Theory of Pitch Perception," J. Acoust. Soc. Am. **71**, 1568–1580.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands).

Fant, G. (**1975**). "Non-Uniform Vowel Normalization," Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 1–19.

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. (**1980**). "Voice Source Dynamics," Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 17–37.

Fant, G. (**1982a**). "Preliminaries to Analysis of the Human Voice Source," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 1–25.

Fant, G. (**1982b**). "The Voice Source: Acoustic Modeling," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 28–48.

Fant, G. (**1985**). "The Voice Source: Theory and Acoustic Modeling," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (The Denver Center for the Performing Arts, Denver, CO), pp. 453–464.

Fant, G. (**1986**). "Glottal Flow: Models and Interaction," J. Phon. **14**, 393–400.

Fant, G., and Ananthapadmanabha, T. V. (**1982**). "Truncation and Superposition," Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 1–17.

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (**1972**). "Subglottal Formants," Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 85–107.

Fant, G., Liljencrants, J., and Lin, Q. G. (**1985**). "A Four-Parameter Model of Glottal Flow," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 1–13.

Fant, G., and Lin, Q. G. (**1987**). "Glottal Source—Vocal Tract Acoustic Interaction," Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 13–27.

Fant, G., Lin, Q. G., and Gobl, C. (**1985**). "Notes on Glottal Flow Interaction," Speech Trans. Lab. Q. Prog. Stat. Rep. **2**, Royal Institute of Technology, Stockholm, 18–24.

Fant, G., and Mártony, J. (**1963**). "Speech Analysis," Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 1–5.

Farnsworth, D. W. (**1940**). "High Speed Motion Pictures of the Human Vocal Cords," Bell Lab. Rec. **18**, 203–208.

Fischer-Jorgensen, E. (**1967**). "Phonetic Analysis of Breathy (Murmured) Vowels in Gujarati," Indian Linguistics **28**, 71–139.

Flanagan, J. L. (1958). "Some Properties of the Glottal Sound Source," J. Speech Hear. Res. 1, 99–116.

Flanagan, J. L. (1972). *Speech Analysis, Synthesis and Perception* (Springer, New York).

Flanagan, J. L., Ishizaka, K., and Shipley, K. L. (1975). "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract," Bell Sys. Tech. J. 54, 485–506.

Fourcin, A. J. (1981). "Laryngographic Assessment of Phonatory Function," in *Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports 11*, edited by C. L. Ludlow and M. O. Hart (American Speech and Hearing Association, Rockville, MD); reprinted in J. Phon. 14, 435–442.

Fujimura, O. (1968). "Approximation to Voice Aperiodicity," IEEE Trans. Audio Electroacoust. AU-16, 68–73.

Fujisaki, H., and Ljungqvist, M. (1986). "Proposal and Evaluation of Models for the Glottal Source Waveform," Proc. Int. Conf. Acoust. Speech Signal Process. ICASSP-86, 1605–1608.

Goldstein, U. (1980). "An Articulatory Model for the Vocal Tracts of Growing Children," unpublished Sc.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Gunzburger, D. (1987). "Duality in Vocal Gender Roles," Prog. Rep. Institute Phon., Utrecht 12, (2), 1–10.

Halle, M. and Stevens, K. N. (1971). "A Note on Laryngeal Features," Res. Lab. Electron. Q. Prog. Rep. 101, MIT, Cambridge, MA, 198–213.

Hawkins, S., and Stevens, K. N. (1985). "Acoustic and Perceptual Correlates of the Non-Nasal/Nasal Distinction for Vowels," J. Acoust. Soc. Am. 77, 1560–1575.

Henton, C. G., and Bladon, R. A. W. (1985). "Breathiness in Normal Female Speech: Inefficiency Versus Desirability," Lang. Commun. 5, 221–227.

Henton, C. G., and Bladon, R. A. W.(1987). "Creak as a Sociophonetic Marker," in *Language, Speech and Mind: Studies in Honor of Victoria Fromkin*, edited by L. Hyman and C. N. Li (Routledge, London), pp. 3–29.

Hollien, H. (1974). "On Vocal Registers," J. Phon. 2, 125–143.

Hollien, H., Michel, J., and Doherty, E. T. (1973). "A Method for Analyzing Vocal Jitter in Sustained Phonation," J. Phon. 1, 85–91.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1988). "Glottal Air Flow and Pressure Measurements for Soft, Normal and Loud Voice by Male and Female Speakers," J. Acoust. Soc. Am. 84, 511–529.

Holmes, J. N. (1961). "Research on Speech Synthesis," Joint Speech Research Unit Report JU 11–4, British Post Office, Eastcote, England.

Holmes, J. N. (1973). "Influence of Glottal Waveform on the Naturalness of Speech from a Parallel Formant Synthesizer," IEEE Trans. Audio Electroacoust. AU-21, 298–305.

Horii, Y. (1979). "Fundamental Frequency Perturbation Observed in Sustained Phonation," J. Speech Hear. Res. 22, 5–19.

Horii, Y. (1980). "Vocal Shimmer in Sustained Phonation," J. Speech Hear. Res. 23, 202–209.

Huffman, M. K. (1987). "Measures of Phonation Type in Hmong," J. Acoust. Soc. Am. 81, 495–504.

Hunt, M. J. (1987). "Studies of Glottal Excitation Using Inverse Filtering and an Electroglottograph," Proceedings of Eleventh International Congress of Phonetic Sciences, Tallinn, Estonia, Vol. 3, 23–26.

Ishizaka, K., and Matsudaira, M. (1968). "What Makes the Vocal Cords Vibrate?," in *The Sixth International Congress on Acoustics, Vol. II*, edited by Y. Kohasi (Elsevier, New York), B9–B12.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (1976). "Input Acoustic Impedance Measurements of the Subglottal System," J. Acoust. Soc. Am. 60, 190–197.

Javkin, H. R., and Maddieson, I. (1983). "An Inverse Filtering Study of Burmese Creaky Voice," Work. Papers Phon. 57, U. California at Los Angeles, 115–125.

Kahn, M. (1975). "Arabic Emphatics: The Evidence for Cultural Determinants of Phonetic Sex-Typing," Phonetica 31, 38–50.

Karlsson, I. (1985). "Glottal Waveforms for Normal Female Speakers," Speech Trans. Lab. Q. Prog. Stat. Rep. 1, Royal Institute of Technology, Stockholm, 31–36.

Karlsson, I. (1987). "Sex Differentiation Cues in the Voices of Young Children of Different Language Backgrounds," J. Acoust. Soc. Am. Suppl. 1 81, S68.

Kasuya, H., and Ogawa, S. (1986). "Normalized Noise Energy as an Acoustic Measure to Evaluate Pathologic Voice," J. Acoust. Soc. Am. 80, 1329–1334.

Kato, Y., Ochiai, K., Fujimura, O., and Maeda, S. (1967). "A Vocoder Excitation with Dynamically Controlled Voicedness," 1967 Conference on Speech Communication and Processing, Cambridge, MA, 288–291.

Kirk, P., Ladefoged, P., and Ladefoged, J. (1984). "The Linguistic Use of Different Phonation Types," Work. Papers Phon. 59, U. California at Los Angeles, 102–113.

Klatt, D. H. (1980). "Software for a Cascade/Parallel Formant Synthesizer," J. Acoust. Soc. Am. 67, 971–995.

Klatt, D. H. (1982). "Prediction of Perceived Phonetic Distance from Critical-Band Spectra: A First Step," Proceedings of the International Conference of Acoustics on Speech and Signal Processing, ICASSP-82, 1278–1281.

Klatt, D. H. (1984). "The New MIT Speechvax Computer Facililty," Speech Communication Group Working Papers IV, Research Laboratory of Electronics, MIT, Cambridge, MA, 73–82.

Klatt, D. H. (1986a). "Representation of the First Formant in Speech Recognition and in Models of the Auditory Periphery," in *Proceedings of Montreal Symposium on Speech Recognition*, edited by P. Mermelstein (McGill University, Montreal, 1986), pp. 5–7.

Klatt, D. H. (1986b). "Detailed Spectral Analysis of a Female Voice," J. Acoust. Soc. Am. Suppl. 1 81, S80.

Klatt, D. H. (1987a). "Acoustic Correlates of Breathiness: First Harmonic Amplitude, Turbulence Noise, and Tracheal Coupling," J. Acoust. Soc. Am. Suppl. 1, 82, S91.

Klatt, D. H. (1987b). "Review of Text-to-Speech Conversion for English," J. Acoust. Soc. Am. 82, 737–793.

Klatt, D. H. (1989). "Review of Selected Models of Speech Perception," in *Lexical Representation and Process*, edited by W. Marslen-Wilson (MIT, Cambridge, MA), pp. 169–226.

Klatt, D. H., and Stevens, K. N. (1969). "Pharyngeal Consonants," Res. Lab. of Electron. Q. Prog. Rep. 93, MIT, Cambridge, MA, 207–216.

Koopmans-van Beinum, F. J. (1980). "Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions," Ph.D. dissertation, Academic, Amsterdam.

Ladefoged, P. (1973). "The Features of the Larynx," J. Phonetics 1, 73–83.

Ladefoged, P. (1983). "The Linguistic Use of Different Phonation Types," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. Bless and J. Abbs (College Hill, San Diego), pp. 351–360.

Ladefoged, P., and Antoñanzas-Barroso, N. (1985). "Computer Measures of Breathy Phonation," Work. Papers Phon. 61, U. California at Los Angeles, 79–86.

Laver, J. (1980). "*The Phonetic Description of Voice Quality* (Cambridge U.P., Cambridge).

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., and Studdert-Kennedy, M. (1967). "Perception of the Speech Code," Psychol. Rev. 74, 431–461.

Lieberman, P. (1961). "Perturbation in Vocal Pitch," J. Acoust. Soc. Am. 33, 597–603.

Lieberman, P. (1963). "Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges," J. Acoust. Soc. Am. 35, 344–353.

Lieberman, P. (1967). *Intonation, Perception and Language* (MIT, Cambridge, MA).

Makhoul, J., Vishwanathan, R., Schwartz, R., and Huggins, A. W. F. (1978). "A Mixed-Source Model for Speech Compression and Synthesis," J. Acoust. Soc. Am. 64, 1577–1581.

Margulies, M. K. (1979). "Male–Female Differences in Speaker Intelligibility: Normal versus Hearing Impaired Listeners," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. J. Wolf and D. H. Klatt (Acoustical Society of America, New York), pp. 363–366.

Meditch, A. (1975). "The Development of Sex-Specific Speech Patterns in Young Children," Anthropol. Linguistics 17, 421–465.

Milenkovic, P. (1987). "Least Mean Square Measures of Voice Perturbation," J. Speech Hear. Res. 30, 529–538.

Monsen, R. B., and Engebretson, A. M. (1977). "Study of Variations in the Male and Females Glottal Wave," J. Acoust. Soc. Am. 62, 981–993.

Nord, L., Ananthapadmanabha, T. V., and Fant, G. (1986). "Signal Analysis and Perceptual Tests of Vowel Responses with an Interactive Source-Filter Model," J. Phon. 14, 401–404.

Pandit, P. B. (1957). "Nasalization, Aspiration and Murmur in Gujarati," Indian Linguistics 17, 165–172.

Peterson, G. E., and Barney, H. L. (1952). "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24, 175–184.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). "Speaking Clearly for the Hearing Impaired I: Intelligibility Differences between Clear and Conversational Speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking Clearly for the Hearing Impaired II: Acoustic Characteristics of Clear and Conversational Speech," J. Speech Hear. Res. **29**, 434–449.

Pollack, I. (**1971**). "Amplitude and Time Jitter Thresholds for Rectangular Wave Trains," J. Acoust. Soc. Am. **50**, 1133–1142.

Robinson, D. W., and Dadson, M. A. (**1956**). "A Redetermination of Equal-Loudness Relations for Pure Tones," Br. J. Appl. Phys. **7**, 166–181.

Rosenberg, A. (**1968**). "Effect of Pitch Averaging on the Quality of Natural Vowels," J. Acoust. Soc. Am. **44**, 1592–1595.

Rosenberg, A. (**1971**). "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," J. Acoust. Soc. Am. **49**, 583–590.

Rothenberg, M. (**1973**). "A New Inverse Filtering Technique for Deriving the Glottal Air Flow Waveform during Voicing," J. Acoust. Soc. Am. **53**, 1632–1645.

Rothenberg, M. (**1974**). "Glottal Noise During Speech," Speech Trans. Lab. Q. Prog. Stat. Rep. **2–3**, Royal Institute of Technology, Stockholm, 1–10.

Rothenberg, M. (**1985**). "Source-Tract Acoustic Interactions in Breathy Voice," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), pp. 155–165.

Rothenberg, M., Carlson, R., Granstrom, B., and Lindqvist-Gauffin, J. (**1975**). "A Three-Parameter Voice Source for Speech Synthesis," in *Speech Communication*, edited by G. Fant (Almqvist and Wiksell, Uppsala, Sweden), Vol. 2, pp. 235–243.

Rozsypal, A. J., and Millar, B. F. (**1979**). "Perception of Jitter and Shimmer in Synthetic Vowels," J. Phon. **7**, 343–355.

Ryalls, J., and Lieberman, P. (**1982**). "Fundamental Frequency and Vowel Perception," J. Acoust. Soc. Am. **72**, 1631–1634.

Sachs, J., Lieberman, P., and Erickson, D. (**1973**). "Anatomical and Cultural Determinants of Male and Female Speech," in *Language Attitudes: Current Trends and Prospects*, edited by R. W. Shuy and R. W. Fasold (Georgetown U.P., Washington, DC).

Shadle, C. (**1987**). "The Acoustics of Fricative Consonants," Ph.D. thesis, MIT, Cambridge, MA.

Sondhi, M. M. (**1975**). "Measurement of the Glottal Waveform," J. Acoust. Soc. Am. **57**, 228–232.

Sonesson, B. (**1960**). "On the Anatomy and Vibratory Pattern of Human Vocal Folds," Acta. Oto-Laryngol. Suppl. 156.

Stevens, K. N. (**1971**). "Airflow and Turbulence Noise for Fricative and Stop Consonants," J. Acoust. Soc. Am. **50**, 1180–1192.

Stevens, K. N. (**1977**). "Physics of Larynx Behavior and Larynx Modes," Phonetica **34**, 264–279.

Stevens, K. N. (**1981**). "Vibration Modes in Relation to Model Parameters," in *Vocal-Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo, Tokyo), pp. 291–301.

Stevens, K. N., and Klatt, D. H. (**1974**). "The Role of Formant Transitions in the Voiced–Voiceless Distinction for Stops," J. Acoust. Soc. Am. **55**, 653–659.

Sundberg, J., and Gauffin, J. (**1979**). "Waveform and Spectrum of the Glottal Voice Source," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, New York), pp. 301–322.

Thorne, B., Kramearae, C., and Henley, B. (Eds.) (**1983**). *Language, Gender and Society* (Newbury House, Rowley, MA).

Timke, R., von Leden, H., and Moore, P. (**1959**). "Laryngeal Vibrations: Measurements of the Glottic Wave. Part II: Physiological Considerations," A. M. A. Arch. Otolaryng. **69**, 438–444.

Titze, I. R. (**1974**). "The Human Vocal Cords: A Mathematical Model," Phonetica **29**, 1–21.

Titze, I. R. (**1984**). "Parametrization of the Glottal Area, Glottal Flow, and Vocal Fold Contact Areas," J. Acoust. Soc. Am. **75**, 570–580.

Titze, I. R., and Talkin, D. (**1979**). "A Theoretical Study of the Effects of the Various Laryngeal Configurations on the Acoustics of Phonation," J. Acoust. Soc. Am. **66**, 60–74.

Ward, P. H., Sanders, J. W., Goldman, R., and Moore, G. P. (**1969**). "Diplophonia," Ann. Otol., Rhinol., and Laryngol. **78**, 771 –777.

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.