Characteristics of the glottal turbulent noise source

Robert E. Hillman
Department of Communication Disorders, 48 Cummington Street, Boston University, Boston, Massachusetts 02215

Elizabeth Oesterle
Department of Audiology and Speech Science, Purdue University, West Lafayette, Indiana 47907

Lawrence L. Feth
Department of Speech-Language-Hearing: Sciences and Disorders, 290 Haworth Hall, The University of Kansas, Lawrence, Kansas 66045

(Received 27 January 1983; accepted for publication 16 May 1983)

This investigation utilized a reflectionless tube technique to obtain direct estimates of turbulent noise produced at the glottis during whispered vowels. In the past, the glottal turbulent noise source has been described theoretically as a series pressure source having a spectrum that is relatively flat for 2 or 3 oct around a center frequency [K. N. Stevens, J. Acoust. Soc Am. **50**, 1180–1192 (1971)]. Center frequency is determined primarily by the area of the constriction at which turbulence is produced with the volume velocity of air flowing through the constriction. The present results were shown to substantiate this theoretically based description of the glottal turbulent noise source. In addition, there was *no* significant difference between the glottal turbulent noise spectra of male and female speakers. The application of these findings to the synthesis of whispered vowels is discussed.

PACS numbers: 43.70.Bk, 43.70.Gr

INTRODUCTION

The primary source of vocal tract acoustic excitation for whispered vowels and aspiration (the /h/ sound) is turbulent noise generated by forcing air through a partially constricted glottis. While the nature of turbulent noise produced at the glottis during whisper or aspiration has been discussed theoretically (Fant, 1970; Stevens, 1971), to the best of our knowledge there have been no published reports of attempts to obtain direct estimates of glottal turbulent noise source characteristics. Such estimates would not only provide a test of existing theory but would also be useful to those interested in synthesizing whispered vowels (Oesterle *et al.*, 1980; Voigt *et al.*, 1981) or aspiration. Thus the goal of the present study was to obtain direct estimates of turbulent noise produced at the level of the glottis by normal speakers.

I. METHODS

This study used a reflectionless tube technique, which has been described elsewhere (Sondhi, 1975; Monsen and Engebretson, 1977; Hillman and Weinberg, 1981a). Briefly, this method requires that subjects produce neutral non-nasal vowels with their lips sealed around the mouthpiece of an acoustically reflectionless tube. Such a tube acts as a "pseudoinfinite" termination of the vocal tract resulting in significant damping of vocal tract resonances. Thus, sound pressure, sensed by a small microphone placed within the tube, is viewed as an estimate of glottal source excitation (Sondhi, 1975).

The efficacy of using a reflectionless tube to estimate the source spectra of whispered vowels is supported by the results of a previous investigation by Hillman and Weinberg (1981a). In that study, computer modeling was used to assess the effects of violating assumptions which underlie the reflectionless tube technique. One primary conclusion of that work was that the reflectionless tube technique is a valid methodology for estimating the overall slopes of glottal source spectra. Thus it appears that the reflectionless tube was well suited for obtaining estimates of the source spectra of whispered vowels.

A. Subjects and tasks

Ten normal speakers (five male, five female) ranging in age from 20 to 28 years served as subjects in this study. Subjects were required to produce three neutral /ə/, non-nasal, whispered vowels with their lips sealed around the mouthpiece of a reflectionless tube. Each subject was instructed to hold the intensity of each of their whispered productions at a constant level for at least 4 s using the VU meter of a tape recorder for visual feedback.

B. Instrumentation and analysis of data

Details concerning the construction of the reflectionless tube used in this study were provided in a previous report (Hillman and Weinberg, 1981b). The earlier work demonstrated that the reflectionless tube could be vented in a way which adequately relieved air pressure within the tube as subjects phonated and yet maintained the anechoic properties of the tube, i.e., an essentially flat ($\pm$ 1.5 dB) frequency response from 20–5000 Hz.

Sound pressure within the reflectionless tube was sensed by a $\frac{1}{4}$-in. condenser microphone (Brüel and Kjaer model 4136). As subjects produced whispered vowels in the tube, the output of the microphone was amplified and fed to a full-track magnetic tape recorder (Crown model IM-7).

A 2.4-s tape loop was made from the middle steady-state portion of each whisper sample and repeatedly played

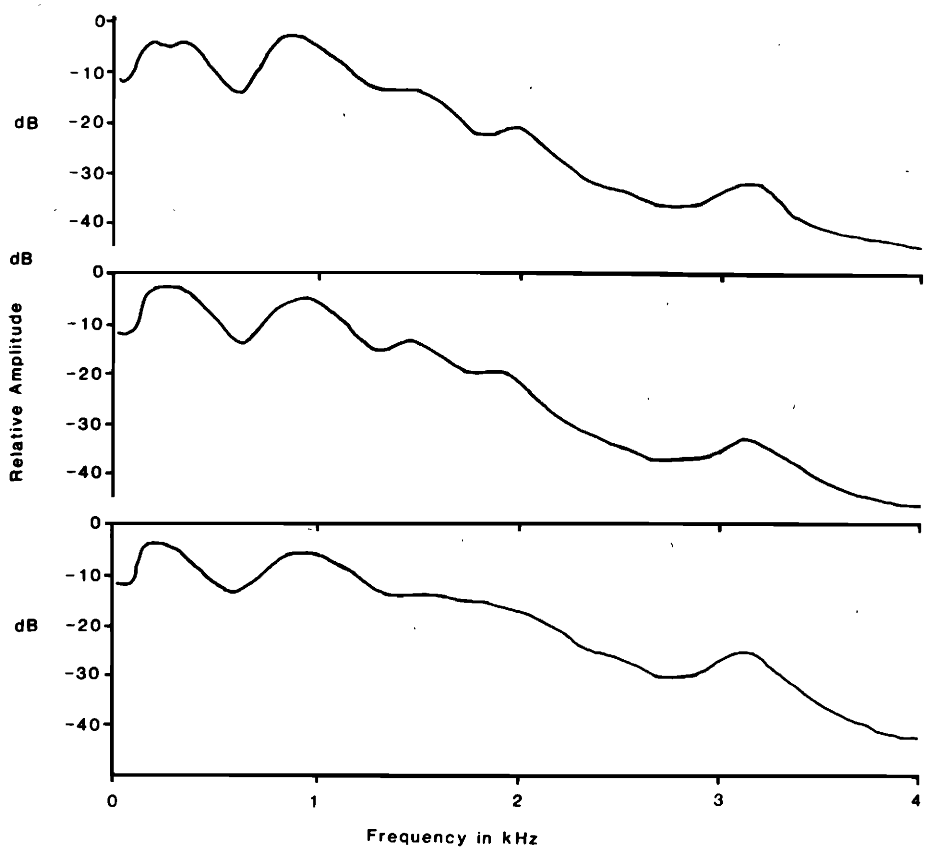

FIG. 1. Amplitude spectra of three whispered vowels produced by the same subject into the reflectionless tube.

on a sound spectrograph (Voice Identification, Inc. Series 700). The output from the spectrograph (scan-out) was fed to a wave analyzer (General Radio, model 1900) set at a bandwidth of 50 Hz. The output of the wave analyzer was displayed on a graphic level recorder (General Radio, model 1521-B; paper speed 1.5 in. min ; pen speed 1 in./s) to provide a permanent record of the amplitude spectrum of each whispered vowel sample.

II. RESULTS AND DISCUSSION

A typical example of the amplitude spectra of three whispered vowels produced by the same subject into the reflectionless tube is shown in Fig. 1. These data serve to illustrate three important points. First, the overall configurations of the three spectra in Fig. 1 are highly similar. This was the case for all of the subjects in this study; the amplitude spectra for a given subject appeared highly similar across the three whispered vowel samples produced by that subject.

Second, the spectra in Fig. 1 display evidence of residual formant shaping in the form of a slight "ripple," particularly at low frequencies. The presence of "formant ripple" was expected since previous work has shown that the reflectionless tube is not totally effective in damping out vocal tract resonances particularly in the region of the first formant (Hillman and Weinberg, 1981a). It was also shown in these earlier studies that the magnitude of ripple which appears to be present in these whisper spectra (10 to 15 dB) is not sufficient to preclude making relatively accurate estimates of overall spectral slopes.

Finally, these spectra appear to decline steadily in amplitude as frequency increases. This was the case for all the whisper spectra measured in this study.

A. Male-female comparison

Least squares linear approximations were generated for each measured spectrum as an initial step in determining

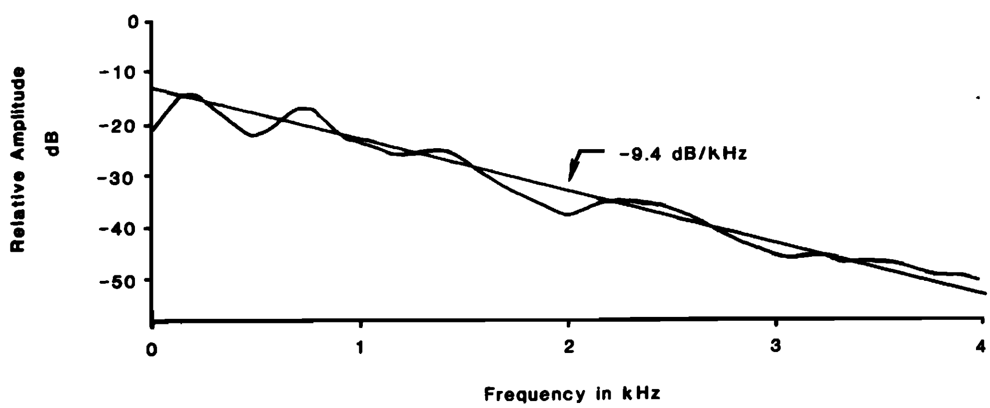

FIG. 2. Least squares linear approximation to the source spectrum of a whispered vowel.

TABLE I. Mean slopes of the three whisper source spectra obtained for each subject.

	Males					Females				
	1	2	3	4	5	1	2	3	4	5
Mean spectral slope (dB/kHz)	− 9.7	− 11.0	− 10.6	− 10.1	− 9.4	− 9.8	− 8.3	− 8.8	10.8	− 9.0

whether the spectra of whispered vowels produced by male speakers differed significantly from those produced by female speakers. Each linear approximation was based on 40 pairs of data points per spectrum which were obtained by measuring the amplitude of each spectrum at 100-Hz increments in frequency going from 100 to 4000 Hz.

An example of a typical least squares linear approximation to a whisper spectrum is shown in Fig. 2. This particular straight line approximation declines in amplitude 9.4 dB for every increase in frequency of 1000 Hz. In other words, it has a − 9.4 dB/kHz slope.

We acknowledge that it is common practice to specify spectral slopes in terms of logarithmic (dB/oct or dB/decade) rather than linear frequency. However, it was found that the best linear approximations for the measured spectra were obtained using a linear frequency scale ($r = -0.87$ to -0.98) and thus spectral slopes are also expressed in terms of linear frequency.

Table I shows a summary of the spectral slope data for male and female subjects. The values shown are in dB/kHz and represent the mean of the three spectral slopes obtained for each subject. Male whisper spectra displayed slopes which range from − 8.2 to − 11.6 dB/kHz with an overall mean slope of − 10.2 dB/kHz and a standard deviation of 2.16 dB/kHz. Female whisper spectra displayed slopes which range from − 8.0 to − 11.5 dB/kHz with an overall mean slope of − 9.3 dB/kHz and a standard deviation of 1.91 dB/kHz.

The overall mean slope of the male-whisper spectra was shown by a t-test *not* to be significantly different ($p > 0.1$) from the overall mean slope of the female-whisper spectra. The combined male and female spectral data had a mean slope of − 9.4 dB/kHz with a standard deviation of 2.04 dB/kHz.

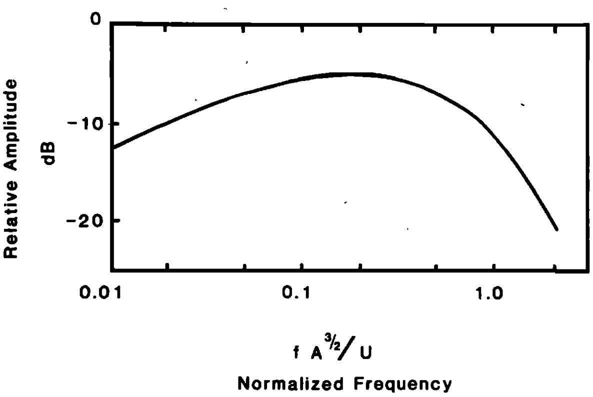

FIG. 3. Spectrum of a series pressure source based on data from aeroacoustic modeling (adapted from Stevens, 1971).

B. Theoretical considerations

The generation of turbulent noise in the vocal tract has been generally viewed as involving a pressure source rather than a volume velocity source (Fant, 1970; Stevens, 1971). Experimental data from aeroacoustic modeling woud indicate that the spectrum of such a pressure source has the form given in Fig. 3 (Stevens, 1971). A normalized frequency scale, $fA^{3/2}/U$, is plotted on the abscissa in Fig. 3, where f is frequency, A is the area of the constriction at which turbulent noise is produced, and U is volume velocity. As can be observed, the spectrum in Fig. 3 is relatively flat over a range of 2 or 3 oct centered on a frequency equal to 0.2 $U/A^{3/2}$.

The area of the glottis during the production of whispered vowels or aspiration is in the range of 0.1 to 0.5 cm^2 and instantaneous airflow rates can be as high as 1500 cm^3/s (Stevens, 1971). In order to generate a theoretically based spectrum for the glottal pressure source, intermediate values for glottal area ($A = 0.3\ cm^2$) and volume velocity ($U = 1000\ cm^3/s$) were used to yield a spectral center frequency of approximately 1200 Hz. The dotted line in Fig. 4 represents the predicted spectrum of the glottal pressure source based on this 1200-Hz center frequency.

In the case of whisper or aspiration it can be assumed that the pressure source is in series with the glottal opening. Under these conditions the glottis acts as an acoustic mass which would impose an approximately 6-dB/oct fall in the spectrum of the series pressure source (Stevens, 1971). The negative slopes which were displayed by all the spectra measured in this study would thus appear to be largely attributed to the acoustic influence of the glottal opening. In order to facilitate comparisons between measured spectra and the predicted spectrum of the glottal pressure source (dotted line in Fig. 4), the assumed influence of the glottal opening was effectively eliminated by the addition of a 6-dB/oct rise to the measured spectra. The solid line in Fig. 4 is a typical example of a measured spectrum which has been modified in this fashion. In comparing the two spectra in Fig. 4, it is clear that the predicted spectrum of the glottal pressure source (dotted line) provides a reasonably good approximation of the measured spectrum (solid line). This was the case for all comparisons made between predicted and measured spectra. Thus the direct estimates of turbulent noise obtained in this investigation appears to substantiate past theoretically based descriptions of the glottal pressure source (Fant, 1960; Stevens, 1971).

It is clear from the simple calculations above that decreasing the areas of the glottal opening would serve to increase the center frequency of the pressure source spectrum (assuming airflow rate remains constant). Since the larynx is generally smaller in females than it is in males, it is conceivable that females would display smaller areas of glottal open-

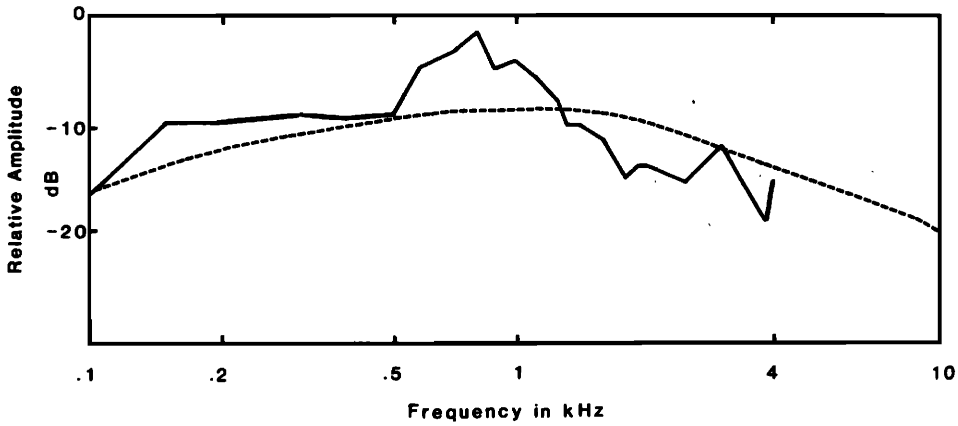

FIG. 4. Comparison of a theoretically based spectrum for the glottal pressure source (dotted line) and a measured spectrum (solid line) which has been modified to eliminate the acoustic load of the glottal opening.

ing and thus higher source spectral center frequencies than males during the production of whispered vowels or aspiration. Unfortunately, we were unable to test this notion due to the fact that residual formant ripple in the measured spectra made the accurate location of center frequencies questionable.

C. Synthesis of whispered vowels

During vowel production, the spectrum of the output at the lips is boosted by approximately 6 dB/oct due to the radiation characteristic. In synthesizing vowels it is common practice to deal with the effect of the radiation characteristic by adding a 6-dB/oct increase to the spectrum of the glottal source used to acoustically excite the vocal tract resonators. In synthesizing whispered vowels the addition of a 6-dB/oct rise to the spectrum of the glottal turbulent noise source would, in effect, cancel out the 6-dB/oct fall imposed by the acoustic mass of the glottal opening. As already demonstrated, eliminating this 6-dB/oct fall produces a source spectrum which is relatively flat over a range of 2 or 3 oct around a center frequency of approximately 1200 Hz.

It has been well established that it is the first three formants which provide the salient cues for the perception of vowels. The first three formants for all vowels are generally located below 3500 Hz which would place them within the relatively flat region of the glottal turbulent noise source spectrum. Given the broadband nature of the source spectrum across the vowel formant region, it can be argued that simple broadband white noise would suffice as the source of vocal tract acoustic excitation for synthesizing whispered vowels. It would therefore seem that the practice of some investigators of using broadband white noise as the source of acoustic excitation in synthesizing whispered vowels (Oesterle *et al.*, 1980; Voigt *et al.*, 1981) is justified by the results of this investigation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Kenneth N. Stevens for his assistance in interpreting the results of this investigation and for his helpful comments on an earlier version of this manuscript. The authors also thank Aliza Artz for her assistance in collecting the data for this investigation. This work was supported in part by grants from NINCDS and the American Speech-Language-Hearing Foundation.

Fant, G. (1970). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Hillman, R. E., and Weinberg, B. (1981a). "Estimation of Glottal Volume Velocity Waveform Properties: A Review and Study of Some Methodological Assumptions," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, New York), Vol. 6, pp. 411–470.

Hillman, R. E., and Weinberg, B. (1981b). "A New Procedure for Venting a Reflectionless Tube," J. Acoust. Soc. Am. **69**, 1449–1451.

Monsen, R., and Engebretson, A. M. (1977). "A Study of Variation in the Male and Female Glottal Wave," J. Acoust. Soc. Am. **62**, 981–933.

Oesterle, E. C., Feth, L. L., and Burns, E. M. (1980). "Pitch of Whispered Vowels," J. Acoust. Soc. Am. Suppl. 1 **67**, S50.

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

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

Voigt, H. F., Sachs, M. B., and Young, E. D. (1981). "Representation of Whispered Vowels in Temporal Patterns of Auditory-Nerve Fiber Discharges," J. Acoust. Soc. Am. Suppl. 1 **69**, S53.