Shippensburg University Interlibrary Loan

ILLiad TN: 54461

Borrower: GDC

Lending String: EXK,*SQP,LFM,PBB,EXH

Patron: Amith, Jonathan

Journal Title: The Journal of the Acoustical Society of America.

Volume: 81 **Issue:**
Month/Year: 1987**Pages:** 495-504

Article Author:

Article Title: Huffman, MK; Measures of phonation type in Hmong

Imprint: [New York, etc.] American Institute of P

ILL Number: 68720503

Call #: Periodicals

Location: Compact Shelving - Lower Level, Left Available

Charge
Maxcost: $25IFM

Shipping Address:
Gettysburg College
300 N. Washington St
Interlibrary Loan
Gettysburg, PA 17325-1493
IDS #132

Fax:
Ariel: 138.234.152.5

The sound pressure at the medial end of the ear canal is always greater than the sound pressure near the outer edge of the obliquely oriented tympanic membrane, and is greater than the maximum sound pressure in the standing wave pattern within the ear canal lateral to the tympanic membrane. Furthermore, if the source drives a coupler that has a larger diameter than the ear canal and the sound pressure is measured at the wall of the coupler, then a still greater discrepancy is obtained between the sound pressure at the medial end of the ear canal and the "coupler" sound pressure. Thus it is not unexpected that calibration procedures involving direct measurements of sound pressure in the ear canal or measurements with couplers that are too large or are not suitably terminated will lead to threshold sound pressures that are lower than those found in the present study.

Our experience in this research has highlighted the difficulties in delivering known sound pressures to the medial end of some ear canals for frequencies at the upper end of the range 8–20 kHz. The cross dimensions near the entrance to some ear canals are sufficiently great that sound propagation within the ear canal in the principal mode cannot be guaranteed. The existence of higher modes in the ear canal precludes the accurate determination of the sound pressure at the tympanic membrane whether the sound is delivered through a coupling tube, as in the present study, or through a more conventional headphone. Consequently, there will be a substantial number of subjects for whom threshold determinations at the higher frequencies are likely to be in error, possibly by 10 dB or more. One way to circumvent this problem and to extend the frequency range over which reliable calibrations can be obtained would be to introduce a gas or liquid, for which the velocity of propagation is greater than that in air, into the ear canal and the coupling tube. The greater velocity of sound propagation would allow accurate high-frequency measurements to be made for ear canals having larger cross dimensions than is possible using air. Ear canals having shapes that deviate markedly from circular are more likely to have a maximum cross dimension that can support cross modes at the high frequencies. These ear canals that are noncircular also present problems in obtaining a proper fit of the earpiece to the ear canal. Further [illegible] the system, using a less homogeneous group of subjects, [illegible] be needed to assess the degree and severity of these limitations.

ACKNOWLEDGMENTS

This research was sponsored by a contract from the National Institute of Neurological and Communicative Disorders and Stroke. We wish to thank Robert Berkovitz, Michael Krasner, Alan Derr, Robert Pyle, and John Stanley for the various skills that they contributed to this research.

ANSI (**1969**). ANSI S3.6-1969, "Specification for audiometers" (American National Standards Institute, New York).

Bezold, F. (**1882**) *Die Corrosions-Anatomie des Ohres*, cited in E. G. Wever and M. Lawrence, *Physiological Acoustics* (Princeton U.P., Princeton, NJ, 1954).

Dreschler, W. A., van der Hulst, R. J. A. M., and Tange, R. A. ([illegible]). "Ototoxicity and the role of high-frequency audiometry," J. Acoust. Soc. Am. Suppl. 1 **76**, S74.

Fausti, S. A., Frey, R. H., Erickson, D. A., Rappaport, D. Z., Cleary, E. L., and Brummett, R. E. (**1979**). "A system for evaluating auditory function from 8000–20 000 Hz," J. Acoust. Soc. Am. **66**, 1713–1718.

Fausti, S. A., Erickson, D. A., Frey, R. H., Rappaport, B. Z., and Schechter, M. A. (**1981**). "The effects of noise upon human hearing sensitivity from 8000 to 20 000 Hz," J. Acoust. Soc. Am. **69**, 1343–1349.

Harris, J. D., and Myers, C. K. (**1971**). "Tentative audiometric threshold level standards from 8 through 18 kHz," J. Acoust. Soc. Am. **49**, 600–601.

Northern, J. L., Downs, M. P., Rudmose, W., Glorig, A., and Fletcher, J. L. (**1972**). "Recommended high-frequency audiometric threshold levels (8000–18 000 Hz)," J. Acoust. Soc. Am. **52**, 585–595.

Stelmachowicz, P. G., Gorga, M. P., and Cullen, J. K. (**1982**). "A calibration procedure for the assessment of thresholds above 8000 Hz," J. Speech Hear. Res. **25**, 618–623.

Stevens, K. N., Berkovitz, R., Kidd, G., Jr., and Green, D. M. (**1987**). "Calibration of ear canals for audiometry at high frequencies," J. Acoust. Soc. Am. **81**, 470–484.

Stinson, M. R., and Shaw, E. A. G. (**1982**). "Wave effects and pressure distribution in the ear canal," J. Acoust. Soc. Am. Suppl. 1 **71**, S88.

Measures of phonation type in Hmong[a)]

Marie K. Huffman
Phonetics Laboratory, University of California, Los Angeles, California 90024

(Received 3 December 1985; accepted for publication 27 October 1986)

This study examines measures of glottal flow for vowels of Hmong, a Southeast Asian language which uses breathy and normal phonation contrastively. A software inverse filter was used to recover glottal airflow from oral airflow recordings. Properties of glottal flow measured in the time domain were glottal pulse symmetry and relative closed-phase duration. In the frequency domain, measures of spectral tilt and the amplitude difference between $F0$ and $H2$ were applied to discrete Fourier transforms (DFTs) of the glottal flow waveforms. Spectral tilt could not be reliably measured for many tokens. For the other measures, values were available for all tokens and were compared across phonation types. Flow pulse symmetry is not significantly different for breathy and normal-voice vowels. On the other hand, prominence of the fundamental relative to the second harmonic is a very significant correlate of the breathy/normal distinction, as is the relative closed-phase duration. These results are considered in light of an existing model of the voice source.

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Hs

INTRODUCTION

Individual speakers of any language may use breathy and creaky phonation types in addition to "normal" phonation. Certain languages also employ these phonation types distinctively. In recent years, phoneticians and speech scientists have been interested in determining the acoustic and physiological properties that are distinctive for phonation types. Two interrelated lines of research are being pursued. One is concerned with identifying distinctive characteristics of phonation types which might be compared across speakers and languages (Fischer-Jørgensen, 1967; Laver, 1980; Bickley, 1982; Ladefoged, 1983). The other aims to ascertain the best means of quantifying different aspects of phonation, from vocal-fold vibration to the acoustic signal (Fourcin, 1974, 1981; Rothenberg, 1981; Javkin and Maddieson, 1983; Kirk *et al.*, 1984). Data on movement of laryngeal structures are not easily obtained because the larynx is delicate and somewhat inaccessible. However, there are indirect indicators of laryngeal behavior, an important one being the glottal airflow. The present study examines glottal airflow properties of phonation types in Hmong, a Southeast Asian language which uses breathy and normal phonation.[1] In particular, we will focus on measures of glottal airflow in the time and frequency domains and will consider how well they distinguish Hmong phonation types. In this way, we follow both lines of inquiry: We add another language to the still small database of linguistically significant phonation types, and we quantify properties of an interface between physiological and acoustic models of phonation, namely, glottal flow.

Our time domain measures of glottal flow were chosen with an eye to their possible physiological correlates. Rothenberg (1981) reports, for example, that physiological properties such as vocal-fold closing speed and closure duration are reflected in the glottal flow waveform. Thus variations in these properties can be inferred from measures of the glottal waveform. What differences of this type might we expect between breathy and normal phonation? Van den Berg (1958) described the basic aerodynamic and physiological conditions necessary for vocal-fold vibration. These general requirements have since been elaborated somewhat [Broad (1973) and Stevens (1977) provide useful discussions], as in the two-mass models of the vocal folds discussed by Ishizaka and Matsudaira (1968, 1972). The essential aerodynamic requirement for vocal-fold vibration is that there be a sizable pressure drop across the glottis. Given this pressure drop and suitable adduction of the vocal folds, vibration is a consequence of the differences in air pressure acting on the superior and inferior sections of the vocal folds, in combination with the mechanical coupling between these two sections, and the stiffness of the folds, which tends to restore them to their initial position from the extremes of the vibratory cycle. Breathy voice is generally considered to be generated by vocal folds that are laxer and less closely adducted than for normal voice. Since tissue stiffness is one of the factors acting to move the vocal folds inward during the closing portion of the vibratory cycle, laxer folds could contribute to a less abrupt closure. In addition, if the folds are further apart than they are for normal voice, then it is possible that a smaller portion of the folds will make contact at closure, producing a shorter closed phase. Furthermore, if the folds are very far apart, there may be no closure at all.

Some properties of the glottal flow waveform reportedly vary under conditions in which we believe these physiological factors to be varying. Bickley (1982) observes, for example, that in Gujarati, glottal flow pulses of breathy vowels have shorter closed phases, and are more symmetrical during the open phase, than are the pulses of normally voiced vowels; flow pulse closure duration is indicative of vocal-fold closure duration. In addition, the pulse symmetry difference Bickley noted may have its physiological counterpart in laryngographic studies of breathy and normal phonation such as those of Fourcin (1974), who reports that although vocal-

---

[a)] Portions of this paper were presented at the 111th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. Suppl. 1 **79**, S37 (1986)].

fold closure is always more rapid than release, the difference between closure and release is less in breathy voice than in normal voice. Javkin and Maddieson (1983) discuss some quantitative measures for these properties of glottal flow waveforms and apply them to phonation contrasts in Burmese. Among other measures, they characterized flow pulse shape by determining the slope of lines between endpoints of the rising and falling branches of the glottal pulse. They report that creaky and normal-voice tokens of Burmese were distinguished by the "slope" of the rising branch of the pulse. Sometimes the comparable measure of the falling branch of the pulse was also distinctive. Here, we apply similar quantification of glottal flow waveforms to the breathy versus normal phonation contrast in Hmong, including also a measure of closure duration.

Our frequency domain measures were chosen to complement spectral measures of phonation type which have already had some success in analyses of normal audio (pressure) recordings. Several investigators (Fischer-Jørgensen, 1967; Bickley, 1982; Ladefoged, 1983; Kirk *et al.*, 1984) have found that measures of the amplitude of some higher spectral component in relation to the amplitude of the fundamental was a distinctive acoustic difference between contrastive phonation types. For example, Fischer-Jørgensen (for Gujarati vowels) and Ladefoged (for !Xóõ vowels) found that the difference in intensity of the fundamental and the first formant or lower harmonics is a fairly reliable correlate of the breathy/normal phonation contrast. Bickley (1982) confirmed these findings and presented results of a perception test which support the notion that listeners are sensitive to this sort of spectral balance when they judge degrees of breathiness. Finally, Kirk *et al.*'s (1984) study of the three-way phonation type distinction in Jalapa Mazatec found that the relative difference between the intensity of the fundamental and the first formant was successful in distinguishing creaky, modal (normal), and breathy vowels. However, though these measures are taken as indicators of differences in laryngeal behavior, they are vulnerable to influences from the supraglottal cavity. For example, Maddieson and Ladefoged (1985) report that in comparing the modal and more tense /e/ vowels of one speaker of Yi, the influence of slightly different locations of the first formant on the second harmonic of high-pitched vowels was strong enough that the relation of harmonic amplitudes went in the opposite direction of what we would expect. That is, the difference in amplitude of the fundamental and the second harmonic was greater for the tense vowel than for the modal vowel. To quantify spectral differences across phonation types of speakers or languages, we clearly need measures which minimize such supraglottal interference. A logical place to look, then, is in frequency characteristics of glottal flow.

Discussions of glottal spectra often take the glottal flow waveform as their starting point. Thus van den Berg (1958) describes the properties of the glottal spectrum with reference to glottal pulse shape. Although he is talking about differences in mode of singing voice or changes in intensity of normal voice, rather than addressing the topic of phonation types, his comments are still helpful in predicting what spectral differences to expect for different voice qualities. Van den Berg says that a "weak and relatively broad puff" (open phase) of the glottis, characteristic of falsetto and low-intensity normal voice, will produce higher harmonics of lower amplitude than those of normal voice. He contrasts this with a "strong and relatively short puff," characteristic of high-intensity normal voice, which produces stronger higher harmonics.

Fant (1980) makes similar claims about the relation of spectral properties to glottal pulse shape. The voice source model described there was developed to parametrize the relationships between glottal spectra and glottal pulse shapes observed for normal phonation produced at different intensities. He notes that pulses which have a less sharp closing, akin to van den Berg's "weak puff," will have spectra dominated by the fundamental, with weak higher harmonics. In contrast, pulses with proportionately sharper closing phases will have weaker low harmonics and stronger higher frequency harmonics. If glottal pulses of breathy-voice vowels have a less sharp closing phase as proposed earlier, we predict that the glottal spectrum will be comparable to that of van den Berg's "broad puffs," with a prominent fundamental and few strong higher harmonics compared to normal voice. We will employ two measures to characterize glottal spectra. One is a local comparison of the amplitudes of the first two harmonics; the other is a more general measure of spectral tilt. The results of our waveform and spectrum measures taken together suggest that measures of glottal flow capture significant differences between phonation types. Furthermore, we suggest that these quantities might be incorporated into existing descriptions of phonation based on normal voicing.

I. METHODS AND PROCEDURES

Studying voice quality in Hmong requires some attention to tone patterns. Syllables in Hmong have a characteristic voice quality and a particular pitch. In Hmong, there are seven such complexes, which will be referred to here as "tones." Table I illustrates this seven-way contrast. The tone labeled here as "checked" is reported by Lyman (1974) to have a glottalized or creaky quality. However, as will be seen, the data collected here support the description of this tone given by Smalley (1976), who attributes no special voice quality to it, describing vowels with this tone as shortened and "terminated by a glottal stop." Both Lyman and Smalley report a breathy-voice quality for the seventh tone in Table I.

TABLE I. The seven-way tonal contrast in Hmong. The number pairs indicate tone height and contour, where 5 is high and 1 is low.

Pitch (quality)	Word, tone transcription, and gloss	
High (normal)	tau^{55}	"pumpkin"
Rising (normal)	tau^{35}	"to dam up (water)"
Mid (normal)	tau^{33}	"to be able"
Low (normal)	tau^{22}	"axe"
Checked (normal)	tau^{ʔ31}	"bean"
Falling (normal)	tau^{42}	"sp. of grass"
Low (breathy)	tau^{32}	"to follow"

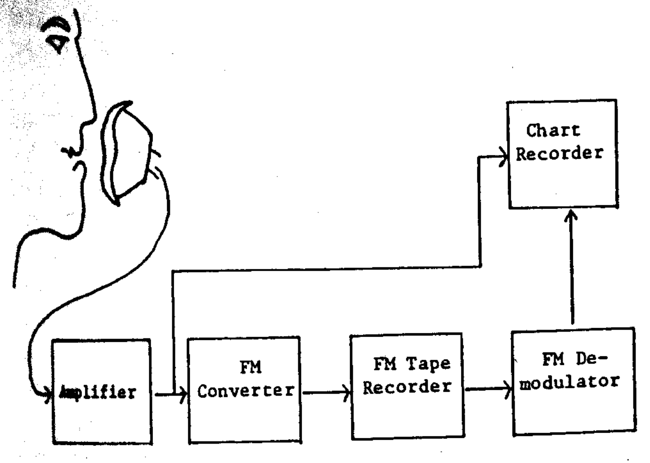

FIG. 1. Schematic diagram of the airflow recording system.

Because we can best draw conclusions about phonatory differences from measures of glottal flow properties which reflect only phonation type differences, we chose to compare "breathy" and normal voiced syllables of similar pitch range and contour. Consultation with a native speaker revealed that the falling and low normal voice tones would be appropriate for comparison with the low breathy (henceforth breathy) tone and the checked tone, the latter included because of its supposed creaky quality. A word list was developed combining each of the four tones with three syllable types of the forms /pa/, /tau/, and /taɯ/.[2] Three male speakers of Hmong between the ages of 16 and 18 were recorded reading each of these words twice.

Combined oral and nasal airflow was recorded for each speaker using a special mask of the type described by Rothenberg (1973), in conjunction with the UCLA Portable Instrumental Phonetics Station, the relevant components of which are represented in Fig. 1. The flow recording was then inverse filtered, as described below. The FM recording system had a frequency response flat (± 3 dB) from dc to 2000 Hz. The frequency response of the flow recording system was not calibrated, but, as the transducer itself has a flat frequency response up to at least 2000 Hz, the overall response is presumably limited by the characteristics of the mask. Rothenberg (1973) reports that the frequency response of the mask is flat to about 1000 Hz.

During recording, the signal from the mask was sent to a preamplifier, then through an FM system, and into a Nagra IV-L tape recorder. For each speaker, a test sample was made during which the FM signal was recorded on the Nagra, and was then displayed on the chart recorder for comparison with the unmodulated signal from the preamplifier. Thus the recording and playback system was checked for each speaker, to ensure a minimum of distortion by the FM system. A cassette tape recording was also made of most of the recording session, to facilitate identification of tokens during analysis. A final audio recording was made without the interference of the mask. At this time, the words in Table I were said once by each of the speakers, and recorded onto the Nagra using a Sennheiser condenser microphone.

The first step in analysis was digitization of the recorded flow data by sampling with an LSI 11/23 computer. After demodulation, the recorded signal was passed through a low-pass Bessel filter with rolloff beginning at about 3 kHz and reaching attenuation of − 25 dB at 9000 Hz, and then digitally sampled into the computer at a rate of 18 000 samples per second. Individual words were then identified and stored separately for further analysis. Another preliminary step was production of wide- and narrow-band spectrograms, and power spectra, from the airflow recording of each vowel token, using a Kay digital sonagraph. Wideband spectrograms were used to determine vowel duration and to observe the general formant properties of the vowel tokens. Power spectra, taken during steady-state portions identified on the wideband spectrograms, also offered insight into formant structure. Narrow-band spectrograms were used for pitch measurements.

In measuring duration from the wideband spectrograms, the beginning of the vowel was taken to be the moment of voice onset after release of the (unaspirated) stops. The end of the vowel was judged as the last point at which there was energy in two of the first three formants. To ascertain pitch from the narrow-band spectrograms, measurements of the highest well-defined harmonic (usually the fourth or the sixth) were made at vowel onset, midpoint, and offset for each token. Fundamental frequency was then calculated from these values.

As mentioned earlier, a major goal of the analysis was recovery of the glottal flow waveform from the airflow recording. The technique of inverse filtering was used, employing a computer program developed at UCLA; for a more detailed description of the procedure see Javkin *et al.* (in press). The general method of inverse filtering aims to remove the resonance effects of the vocal tract and radiation at the lips in order to determine characteristics of the acoustic signal that can be attributed to the glottal source. Lip radiation effects present in oral pressure waveforms are not a factor in volume velocity measured with the Rothenberg mask. Thus, in the present study, only the resonance effects of the vocal tract needed to be filtered out.

The inverse filtering program applies nine zeros within the frequency range 0–9000 Hz. There are four zeros below 4500 Hz; these were specified by the experimenter for each vowel analyzed. The zeros applied at higher frequencies had fixed values, set at intervals of 1000 Hz beginning with 4500 Hz. Damping by the mask and subsequent low-pass filtering during digitization made it unlikely that significant differences between tokens would be present at higher frequencies. LPC formant analysis was used primarily to identify the frequencies of the poles in the spectrum where the lowest four zeros should be applied. To facilitate determination of these formants by LPC analysis, the flow signal was differenced twice. This resulted in a preemphasis of 12 dB/oct, the negative of the rolloff characteristic commonly reported for the glottal source. This procedure has very little effect on the determination of the frequencies of the poles, although it will affect the determination of the bandwidths.

Next, linear prediction coefficients were computed for each vowel token at 10-ms intervals, using a 25.6-ms Hamming window. Trial analyses calculating formants via 14 and 16 coefficients proved unable to consistently separate the first four formants as estimated from the wideband spectrograms and power spectra mentioned earlier, so for these data 18 coefficients were calculated and corresponding formant frequencies and bandwidths were derived by solving for the roots of the LPC polynomial. Spurious formant values generated by LPC analysis were corrected, and values for formants higher than the fourth were deleted so that the files could serve as input to the inverse filtering program.

There was considerable variation in the smoothness of waveforms produced by the first pass of inverse filtering. The multiplicity of factors affecting glottal flow—oral and subglottal pressure, vocal-fold mass, tension, and symmetry, to name a few—leads us to expect gross variations in flow waveshape across speakers and phonation types as well as flow oscillations during the course of any single glottal pulse. However, inverse filtering should remove fine flow oscillations attributable to resonances of the vocal tract. Before it was possible to determine the significance of waveshape variations, it was necessary to determine whether incomplete filtering might be occurring.

A portion of each vowel token was selected for further analysis. The location within the vowels of the chosen portion was motivated by properties of the inverse filtering procedure, which assumes that the vocal tract has relatively little effect on the glottal flow, i.e., that there is no coupling between vocal-tract resonances and vibration of the vocal cords. Under this assumption, when the first-formant frequency is low and thus close to the fundamental frequency, the estimation of the first formant is influenced by the amplitude of the fundamental. In the case of a high-amplitude fundamental, the estimated first-formant frequency is probably too low.

For the Hmong vowels, this problem was avoided by using vowels with fairly high first formants, namely, the nonfront low monophthong [a] and the nonfront low onset of the diphthongs [au] and [aɯ]. A steady-state portion of six glottal pulses was selected in the middle of each vowel token by examination of formant frequency values derived by LPC analysis, and, in some cases, spectrograms as well. In the case of the monophthongs, the selected portion was near the middle of the vowel; for the diphthongs, the steady portion was at about one-third of the duration of the vowel.

Figure 2 shows a sample of a glottal flow waveform, with arrows marking oscillations possibly attributable to incomplete filtering. For this and other tokens, the frequencies of such oscillations were measured from the glottal waveforms. If these frequencies corresponded to formant frequencies for the token, then it was assumed that the formant to be canceled out by inverse filtering was of greater amplitude than originally allowed for by the bandwidths calculated by LPC formant analysis. These original bandwidths from LPC analysis of the airflow data were on the order of 120–200 Hz for $F1$, 130–300 Hz for $F2$, 300–580 Hz for $F3$, and 300–640 Hz for $F4$.

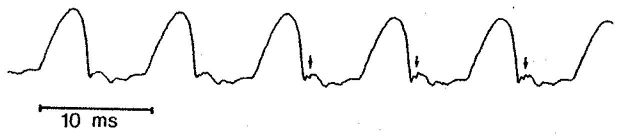

FIG. 2. Glottal flow waveform with residual oscillations possibly due to incomplete filtering.

Mask losses are surely responsible for attenuation, and therefore greater bandwidths, in the higher formants. Another factor in bandwidth determination is that LPC analysis was done within windows that included complete glottal cycles. Consequently, LPC-derived bandwidths represented neither strictly open- nor closed-phase vocal tract acoustics, but a time average. Since the amount of glottal opening has been observed to increase formant damping [Fujimura and Lindqvist (1971)], inclusion of the open phase in LPC analysis would be expected to increase bandwidths somewhat. Pitch-synchronous closed-phase LPC analysis was not possible because, while normal-voice tokens could be assumed to have a closed phase, we could not make this assumption *a priori* for the breathy tokens. On the relation of formant damping to glottal opening, it is interesting to note that many of the largest $F1$ bandwidths in our data occur in breathy-voice tokens, which would be expected to have more glottal opening than normal-voice tokens. For two of the three speakers, breathy-voice tokens have the largest bandwidths, while there are no systematic differences among the nonbreathy tokens.

To attempt to provide for more complete cancellation of formants, the bandwidths to be used for inverse filtering were decreased, and the sample was inverse filtered again. This cycle of varying bandwidths and inverse filtering again was repeated with various amounts of bandwidth reduction, until the best filtering with the least possible bandwidth reduction was achieved, as judged by the least residual high-frequency ripple in the closed phase of the glottal waveform. This procedure of judging quality of filtering by flatness of the closed phase [adopted by Rothenberg (1973) and others] might be considered a *post hoc* approximation to closed-phase bandwidths. In general, for the first and second formants, effective bandwidth reduction was on the order of 10–50 Hz, giving bandwidths for inverse filtering of 70–190 Hz and 80–290 Hz, respectively. For the third and fourth formants, reduction was on the order of 100–200 Hz, resulting in bandwidths of 100–480 Hz and 100–540 Hz, respectively. On a few occasions, examination of an incompletely filtered signal suggested that the frequency values calculated via LPC analysis for $F1$ or $F2$ might be centered slightly high. In these cases, lowering the first or second formant by 20–40 Hz improved the output of inverse filtering.

II. RESULTS

A. Duration and pitch

Before examining the results of the inverse filtering, we need to consider the duration and pitch of the vowel tokens. Duration is of interest because it might be playing a role in signaling the phonemic contrasts in question. The same concern applies to pitch. Another reason to examine pitch is that physiological adjustments made to bring about pitch changes can affect the glottal flow, and hence also influence our measures of properties of the glottal waveform.

Figure 3 shows average duration of vowels spoken with the four Hmong tones of interest to this study. Data are pooled for the three speakers. As was noted by Lyman (1974), the checked vowels are shorter than the normal (low, falling) and breathy ones. Notice also that among these latter three tones, the low tone is on average longer than both the breathy and falling tones. An analysis of variance on the pooled data found that of the four tones studied here, the checked tone had a significantly different duration from the low and breathy tones; there was no significant difference in duration between checked and falling tones or between low, falling, and breathy tones. (Unless stated otherwise, the significance level is at least 0.01, determined using Tukey's studentized range test, which controls for the experimentwise error rate for multiple comparisons.) Therefore, we conclude that checked tones are usually shorter than the other tones, but duration is not a sufficient cue to the tone differences.

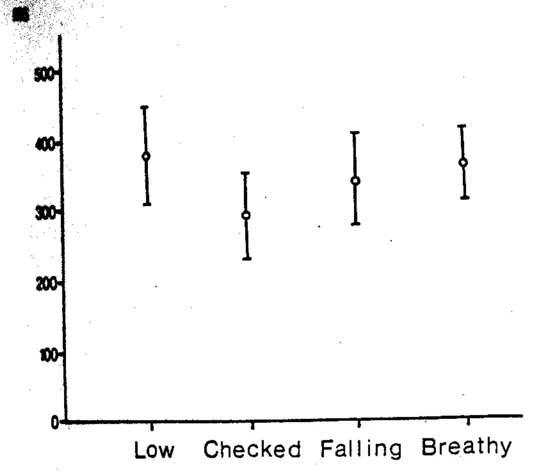

FIG. 3. Mean vowel duration for each of the four tones, pooled across the three speakers.

Figure 4 shows average onset, midpoint, and offset values of $F0$ for each tone type, for each of the three speakers. Each data point thus represents the mean of six measurements. For all speakers, the falling, breathy, and checked tones are often similar in pitch at either the onset or offset, but they usually have different values at the other endpoint, and often, at the midpoint as well. The low tone has its onset below those of the other three tones, but in falling only slightly, it generally ends with a final value higher than the others. The exception is speaker 2, who generally has less pitch differences between these four tones and whose low and checked tones have very close offset values. It should be noted that the tonal contour specifications given for these tones in Table I are only very general representations of the facts in Fig. 4. There is a considerable amount of speaker variation, both in frequency range employed, and in the degree of similarity between tonal contours.

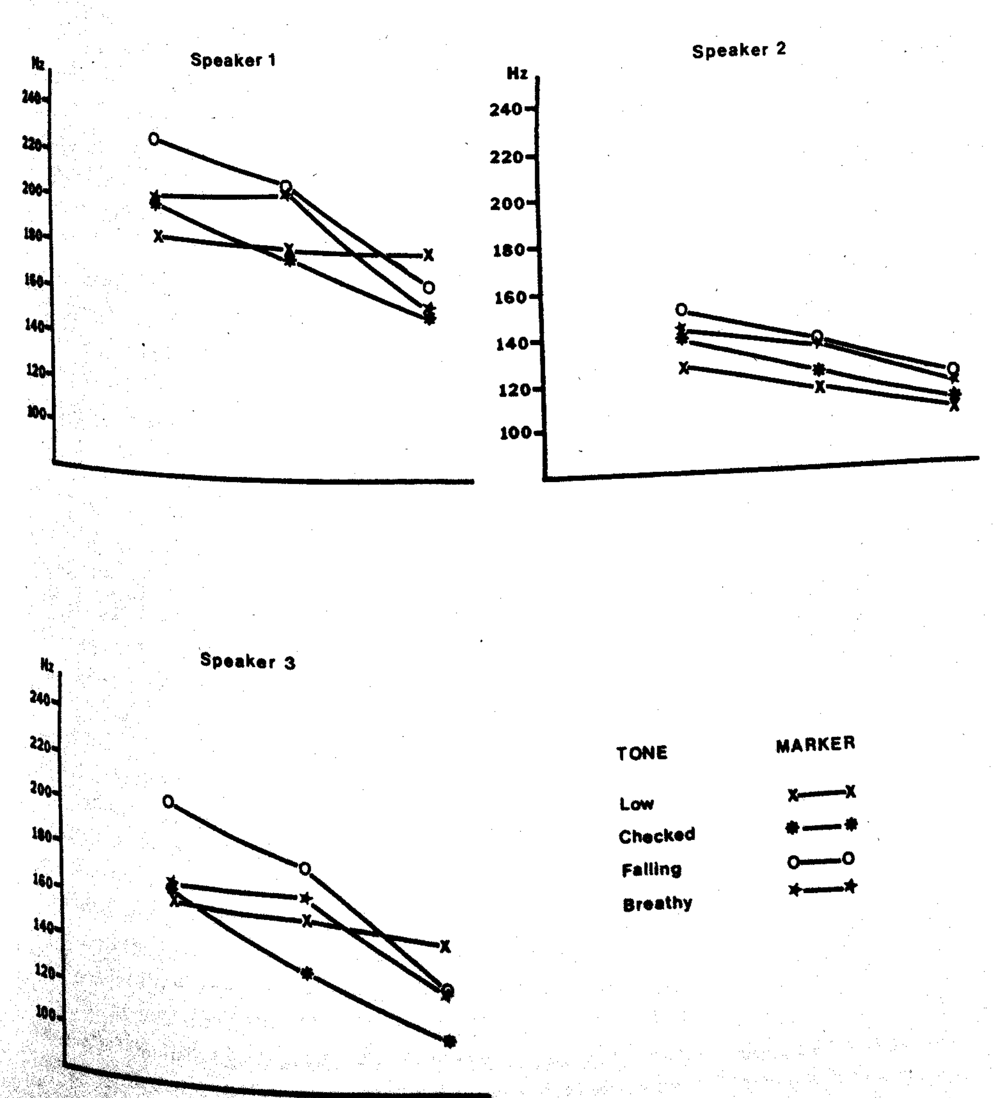

FIG. 4. The $F0$ contours for individual speakers, measured at vowel onset, midpoint, and offset.

In summary, the mean duration of the checked tone is often significantly shorter than that of the other tones, but its duration is not consistently distinctive. Also, while individual onset, offset, and midpoint values of the four tones are sometimes very similar, the pitch contours represented by these points are generally different for the four tones. If our measures of glottal airflow are able to characterize phonation type differences regardless of vowel pitch, then they can be considered that much more effective indicators of phonation type distinctions.

B. Inverse filtering

The output of the inverse filtering procedure is a waveform representing airflow through the glottis. Sample glottal flow waveforms are shown in Fig. 5. In the Introduction, we outlined ways of characterizing differences in glottal flow waveforms. The following two sections present results of analyses of glottal waveshape and spectral properties of the Hmong vowel tokens.

1. Waveshape properties

When examining glottal flow in the time domain, we considered two measures likely to reflect glottal waveshape differences attributable to phonation contrasts. The first measure used was the ratio of the duration of the closed phase of the glottal flow pulse to the total duration of the pitch period, henceforth the closure duration ratio. Bickley (1982) reports that in Gujarati normal-voiced vowels, the closed phase occupies about one-third of the total pitch period, but a smaller proportion in breathy vowels. The closure duration ratio is the complement of a quantity often used in waveform measurement, namely, the duty cycle.

The second measure considered was the ratio of slopes of the falling and rising branches of the open phase of the glottal pulse. We will call this the slope ratio. We discussed earlier how differences in vocal-fold closing speed as reflected in the flow pulse falling branch could be taken as an indirect measure of glottal stricture differences across phonation types. Ideally, then, we would want to compare falling branch slopes for different phonation types. However, Javkin and Maddieson (1983) report that amplitude variations within tokens interfered with the effectiveness of their falling branch slope measure. Therefore, to control for differences in flow pulse amplitude, we used the slope ratio. With this measure, a higher value (greater than 1) indicated a sharper falling than rising branch slope. As was mentioned earlier, open-phase symmetry is a property of glottal flow waveforms which has been reported to vary with the breathy/normal phonation contrast (Bickley, 1982). Furthermore, Fant's (1980) voice source model treats symmetry as a parameter contributing to differences in prominence of low harmonics, one spectral property reported to differ between breathy and normal phonation.

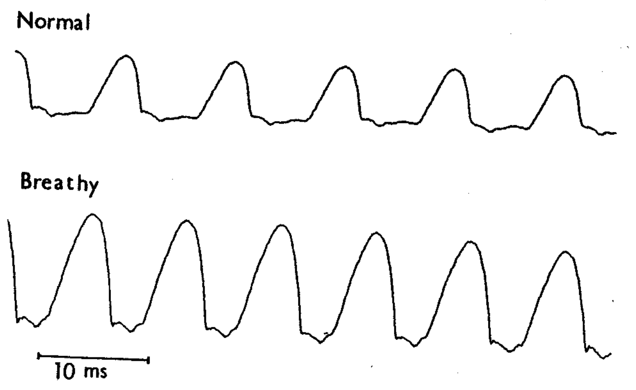

FIG. 5. Sample glottal flow waveforms from normal- and breathy-voice vowels of speaker 2.

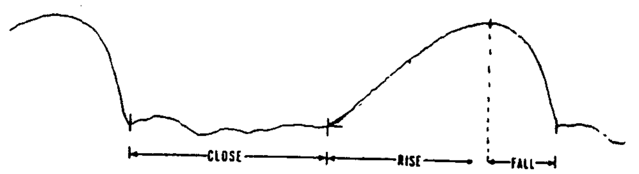

FIG. 6. Sample glottal flow pulse illustrating the closed phase, and the rising and falling branches of the open phase; after Javkin *et al.* (in press).

Glottal flow waveforms were measured with the aid of the Glotta program developed at UCLA [see Javkin *et al.* (in press) for more discussion]. First, the opening and closing points of the glottal pulse were identified for the six steady-state pulses chosen earlier. Figure 6 illustrates typical intervals for a normal-voice vowel. Next, the Glotta program calculated several parameters, including the slopes of the falling and rising branches of the open phase, and the durations of the closed phase, the rising interval, and the falling interval. The rising and falling slopes were determined from the slopes of tangents to the rising and falling branches of the flow pulse. These tangents were estimated using the average derivative over seven points in the middle of the rising branch and falling branch. The closure duration ratio and the slope ratio were calculated for five pulses of each Hmong token. From these values, mean duration and slope ratios were computed for each token.

Figure 7 shows means and standard deviations of the closure duration ratios pooled across speakers for all four tones. The nonbreathy tones are very close in value, with a closed-phase duration at about 40% of the pitch period. The breathy tone, on the other hand, has a closure duration of about half of this, or slightly less than 20% of the pitch period. Analysis of variance found the closed-phase duration ratio to be a significant indicator of differences between phonation types. Thus, while none of the nonbreathy tones was distinguishable from each other on this measure, all had values significantly different from the breathy tone. The same is true for speakers 2 and 3 taken individually. For speaker 1, while the nonbreathy tones were not distinguished from each other on this measure, only the checked and falling tones had closure durations significantly different from the breathy tone.

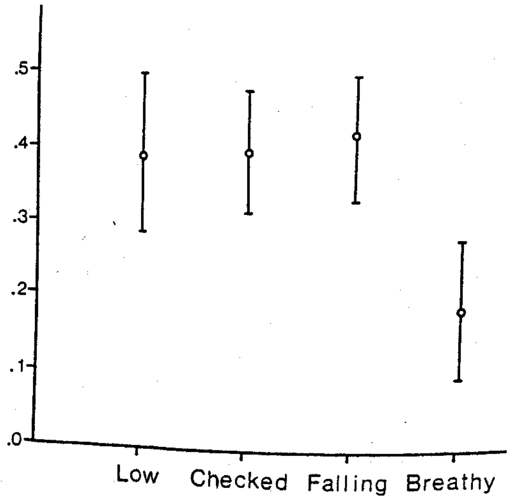

FIG. 7. Mean closure duration ratio (closed phase duration/ total pitch period) for each tone, pooled across speakers.

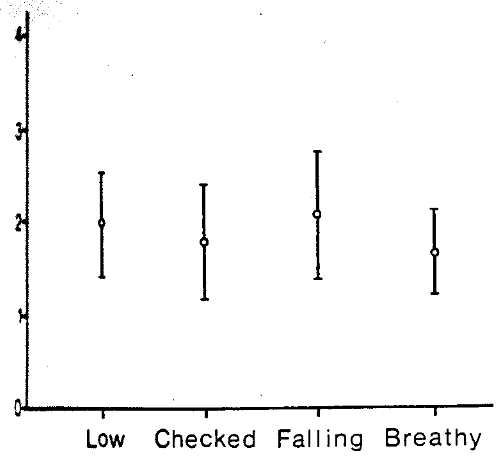

FIG. 8. Mean slope ratio (falling slope/rising slope) for each tone, pooled across speakers.

The ratio of falling to rising slopes was less effective as a diagnostic of phonation differences. Figure 8 shows mean values and standard deviations for each tone, across the three speakers. Contrary to expectations, the breathy tone has a shape that is only marginally more symmetrical than the nonbreathy tones. Analysis of variance showed that the tones were not significantly different from each other on this measure. The slope-ratio measure is equally ineffective when taken as a variable in analyses of variance for individual speakers, and thus fewer tokens. Using the same significance criterion, for no speaker did the ratio of falling to rising slopes distinguish any of the four tones.

What might these results tell us about physiological parameters of phonation contrasts? Symmetry of the glottal pulse open phase as measured here was not significantly different between the breathy and nonbreathy tones. The same lack of significant differences was evident for individual speakers as well. Insofar as pulse symmetry reflects vocal-fold closing speed, it would appear that this parameter is not significantly different for breathy and normal voice in Hmong. Note that this measure also does not reflect differences in pitch because it does not distinguish the normal-voice falling and low tones, which are quite different in pitch at the point at which the samples for inverse filtering were taken.

On the other hand, the ratio of closed-to-total glottal pulse duration has been shown to be successful in distinguishing Hmong breathy and nonbreathy vowels. By its nature, the measure controls for several other linguistic attributes of the vowels that a speaker could have varied independently of phonation type such as vowel quality, pitch, and duration. As with other measures of the inverse filtered waveforms, differences in vowel formants are assumed to have been removed. Pitch differences as reflected in duration of the total pitch period are controlled for by the fact that the measure is a ratio, of closed to total duration. Furthermore, overall vowel duration differences are irrelevant to this measure.

To the extent that flow pulse closure corresponds to vocal-fold closure, the success of the duration ratio measure is interesting because the differences between phonation types are too small to be controlled directly. Recall that the ratio of closed-to-total duration was about 0.40 for normal voice and 0.20 for breathy voice. At a low fundamental frequency of 100 Hz, this would mean a difference in closure duration of approximately 2 ms. According to van den Berg (1958), about 5 ms are required for a complete chain of motor and sensory impulses to pass between the brain and the larynx. Clearly, a speaker cannot be controlling the duration of the closed phase directly. However, a speaker could be controlling one or more physiological variables which would produce this effect. As was mentioned earlier, changes in glottal aperture or vocal-fold tension could contribute to such differences in closure duration of the vocal folds.

It should be noted further that the Hmong checked tone was not distinguished from the other three tones by either the pulse symmetry measure or the closure duration ratio. With respect to these measures, the checked tone acts like the other nonbreathy tones, which are produced with normal voice. In addition to this fact, there is spectrographic evidence that we are justified in considering the checked tone to be produced with normal phonation. Examination of wideband spectrograms reveals that checked-tone syllables have final glottalization. Recall that Smalley (1976) described the checked tone as ending in a glottal stop. This leads us to expect the first part of the vowel to have normal voice quality, with only the end of the vowel being glottalized, i.e., characterized by low-frequency, often irregular taps of the vocal folds as they come together for the glottal stop. This is, in fact, how the checked tone sounds in the pronunciations of our three speakers. Spectrograms of the 18 checked-tone tokens show 14 of them ending with 2 to 6 glottalized pulses, where "glottalized" describes the widely and often irregularly spaced voicing striations. In contrast, only 4 of the 54 nonchecked tokens showed even slight glottalization, of 1 or 2 pulses at the end of the vowel. Thus, in the following, the checked tone will be considered with the nonbreathy tones, without further discussion.

2. Spectral properties

In previous sections, we have discussed how degree of glottal stricture might affect glottal closure duration and sharpness of the glottal closing gesture. In addition, we argued that sharpness of the glottal closing gesture should be reflected in the glottal flow spectrum. If Hmong breathy and nonbreathy vowels differ in glottal stricture, then we would expect differences in spectral balance to be detectable in their glottal spectra. In particular, we would predict that, compared to normal voice, breathy-voice tokens would have a more prominent fundamental and weaker higher harmonics.

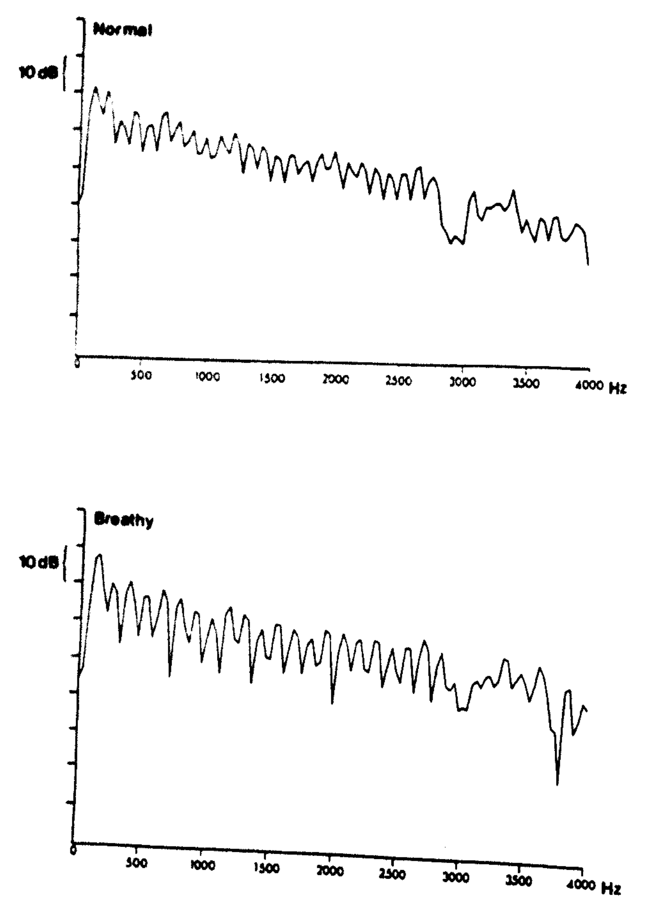

FIG. 9. Spectra of the normal and breathy glottal flow waveforms in Fig. 5.

Two kinds of measures were chosen for analysis of glottal spectra. The first was a general measure of spectral tilt. Calculated as the slope of a line fitted to harmonic peaks, spectral tilt was measured for two frequency intervals: 0 to 1000 Hz and 1000 to 2000 Hz. The second measure was a more localized indication of spectral structure, namely, the difference in relative amplitude of the fundamental and the second harmonic. The first step in spectral analysis was to calculate discrete Fourier transforms (DFTs) for the glottal waveforms of all of the Hmong vowel tokens. A Hamming window 25.6 ms long was fitted over the middle of the six glottal pulses chosen earlier. Care was taken to position the window over the pulses so as to minimize the difference in the amplitude of the signal at the beginning and end of the window (i.e., to minimize discontinuities in the series of pulses). Because of differences in fundamental frequency, the number of pulses analyzed varied from three to six. Figure 9 shows sample spectra of a normal- and a breathy-voice token.

Our general spectral measure refers to spectral tilt between 0 and 2000 Hz. Given the low range of flat frequency response of the Rothenberg mask, it is unlikely that spectral shape determined for frequencies much above 2000 Hz can be considered accurate. A linear regression routine was used to fit lines to harmonic peaks within the two 1000-Hz intervals chosen. Since the literature (for example, Flanagan, 1958; van den Berg, 1958; Fant, 1980) suggests that the spectrum of the glottal source exhibits logarithmic rolloff, lines fitted to harmonic peaks were calculated with frequency on a logarithmic scale. For comparison, the regression was also done with frequency on a linear scale. A two-tailed t test for linear relation was computed for each of these lines. The t tests indicated that many spectra could not be characterized reliably by a line fitted to harmonic peaks within the frequency intervals considered, whether the frequency scale was linear or logarithmic. These results remind us that care should be taken in making comparisons between results of modeling studies and measures of real speech. Measures of spectral tilt are considered further in Jackson *et al.* (1985).

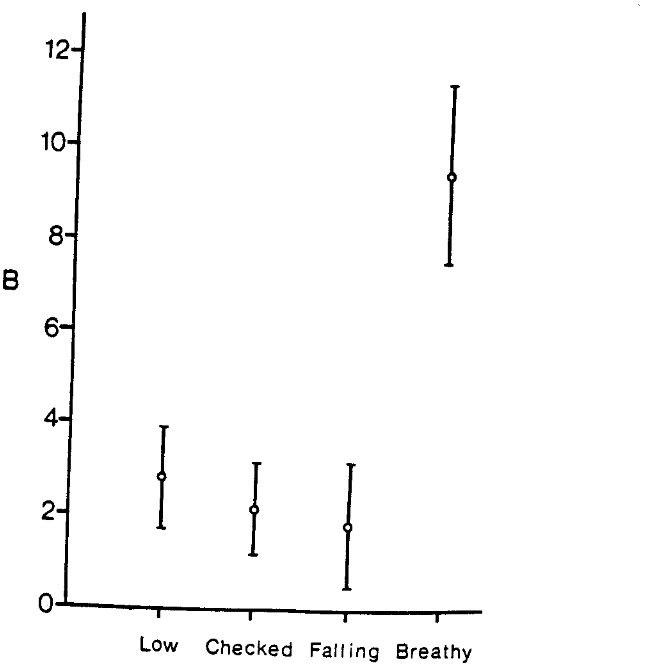

FIG. 10. Mean difference between $F0$ and $H2$ for each tone, pooled across speakers.

Figure 10 shows mean values and standard deviations averaged over speakers for another measure of spectral balance, the difference in amplitude between the first and second harmonics ($F0 - H2$). The nonbreathy tones have very similar values for this difference, all clustering around slightly over 2 dB. The breathy tone shows a much greater difference between $F0$ and $H2$, with a mean of 9.48 dB. An analysis of variance found the $F0 - H2$ measure to be as successful at distinguishing phonation differences as the closure duration ratio. While $F0 - H2$ was not significantly different among the nonbreathy tones, the breathy tone was distinguished from each of the nonbreathy tones on this measure. Analyses on this variable had the same result for each individual speaker as well. Like the closure duration ratio, the $F0 - H2$ measure is not affected by pitch, duration, or quality of the vowel tokens analyzed. Pitch, meaning here the fundamental frequency, or rate of vibration of the vocal folds, affects this measure directly only to the extent that it influences the number of glottal pulses analyzed with the 25.6-ms window for DFT analysis. If pitch were having an influence, we should see a differentiation between low and higher pitch tones spoken in normal voice. The $F0 - H2$ measure does not show any such effect. Instead, the categories of $F0 - H2$ values correspond exactly to the breathy/normal phonation contrast within the Hmong tokens.

II. DISCUSSION

To adequately characterize a phonetic phenomenon, we need to relate its acoustic attributes to perception and production by native speakers. For Hmong phonation types, we have investigated the relationship between acoustic properties and laryngeal gestures via examination of glottal flow. Although we have not considered perception of phonation types here, it should be observed that our results are consistent with Bickley's (1982) finding that an "enhanced first harmonic...was important in the perception of breathy vowels" (p. 80).

In summarizing our Hmong results, we will first discuss the relationship between acoustic properties of the speech signal and the glottal flow waveform. Previous studies of linguistic phonation contrasts identified spectral prominence of the fundamental in the radiated speech signal as an important correlate of phonation type. The Hmong data show that, for the breathy/normal voice contrast, there are quantifiable and significant differences in prominence of the fundamental in glottal spectra as well. This confirms previous claims that source differences, not vocal-tract differences, are responsible for the breathy/normal distinction. These differences are represented in the $F0 - H2$ measure. On the other hand, contrary to theoretical predictions, measures of spectral tilt over several harmonics could not reliably characterize glottal spectra of different phonation types. The spectral contrasts found co-occur with significant differences in flow pulse closure duration, such that, when closure duration is smaller, the fundamental is more prominent. While a shorter closure duration does not directly cause the spectral difference, the relationship established between flow pulse closure duration and prominence of the fundamental relative to the second harmonic is a first step to quantifying the link between the glottal flow waveform and acoustic attributes of phonation types in a natural human language.

Recall that there is already in the literature a model of the waveshape and spectral properties of glottal flow for normal voicing. Fant's (1980) voice source model was developed to describe relationships between glottal waveforms and spectra based on evidence from normal phonation at different intensity levels. We will briefly discuss this model in light of the Hmong data on waveshape and spectral properties of breathy and normal phonation.

Fant's model is based on three parameters: peak flow, glottal frequency (equivalent to the inverse of twice the duration of the rising branch of the flow pulse), and k, a parameter reflecting symmetry between closing and opening speeds of the vocal folds. Quantities such as closure duration fall out from the interaction among these other variables. Variation in the three parameters changes the relative strength of different spectral components. So, for example, if glottal frequency and peak flow are held constant, a lower value of k (making the flow pulse more symmetrical, i.e., less skewed to the right) results in a less sharp falling branch, shorter closure duration, and weak higher harmonics.

To derive the results we reported for Hmong, a model like Fant's needs to be able to produce amplitude differences between $F0$ and $H2$, while generating the appropriate closure durations for the two phonation types. We assume that amplitude differences are not a result of changes in flow pulse symmetry because symmetry appears to be nondistinctive across tones in Hmong. So, for example, varying only k as just described would not be an appropriate way of accounting for all of the Hmong facts.

In Fant's model, there are ways of generating differences in spectral balance without essential reference to symmetry. Low-frequency harmonics can be amplified by increasing the "total air volume of the vocal pulse" (Fant, 1980). As an illustration, Fant suggests that a high-amplitude (peak flow) pulse that is also more symmetrical (thus broader) is consistent with a relative spectral dominance of the fundamental. Describing the Hmong data in similar terms, we might say that a relatively high-amplitude pulse that is also proportionately broad (but not necessarily more symmetrical) has a dominant fundamental, and that this dominance holds not only in comparison to higher harmonics, but also in comparison to the second harmonic. In other words, varying the peak flow and glottal frequency parameters should produce spectra similar to the Hmong spectra. However, the literature holds that it is the sharpness of the falling branch that determines spectral differences most strongly. A better understanding of glottal pulse shape differences (amplitude, rising and closing slope, average "volume") among phonation types and of the physiological factors which produce these differences would help in identification of the primary acoustic correlates of phonation types.

What can be said about the relationship between flow waveshape properties and laryngeal gestures? Fant's model is of little direct assistance in this respect in that the parameters used are arbitrary properties of waveshape that cannot be related simply to physiological quantities. Furthermore, the physiological dimension(s) underlying flow pulse closure variations are not yet known. In this regard, the Hmong results are of interest because, though the differences in duration ratio for breathy and normal voice vowels in Hmong are statistically significant, they are of an order of magnitude so small that they could not be controlled directly. It is possible that, as has been discussed in the literature, variations in glottal stricture are contributing to these closure duration differences. Modeling studies of vocal-fold vibration using physiologically motivated models (e.g., Titze, 1984) may clarify the effect which glottal stricture and other factors such as vocal-fold tension can have on flow pulse closure in phonation contrasts such as those in Hmong. In summary, the Hmong data demonstrate that investigation of glottal flow by inverse filtering can help direct research towards integrating acoustically and physiologically based models of phonation. Moreover, our results suggest that flow pulse volume differences may be more important than differences in pulse symmetry in the normal/breathy contrast.

ACKNOWLEDGMENTS

This work was supported by NIH Grant 18163-02 to Peter Ladefoged. Members of the UCLA Phonetics Laboratory Group have been of assistance on many fronts; particular thanks to Peter Ladefoged, Ian Maddieson, and Patricia

Keating. Thanks also to Corine Bickley for many valuable comments on a draft of this paper.

---

[1] Hmong is a Miao language belonging to the Miao-Yao language group of Sino-Tibetan. The majority of Hmong speakers live in Southwestern China, but many also populate northern parts of Burma, Laos, Vietnam, and Thailand. There are substantial Hmong immigrant communities in the United States.

[2] Two of the resulting items were possible but unreal words in Hmong: breathy /pa/ and falling (normal voiced) /taɯ/.

Bickley, C. (**1982**). "Acoustic Analysis and Perception of Breathy Vowels," MIT Work. Pap. Speech Commun. **1**, 71–81.

Broad, D. J. (**1973**). "Phonation," in *Normal Aspects of Speech, Hearing, and Language*, edited by F. Minifie, T. J. Hixon, and F. Williams (Prentice-Hall, Englewood Cliffs, NJ), pp. 127–167.

Fant, G. (**1980**). "Vocal Source Dynamics," Speech Transmis. Lab. Q. Prog. Stat. Rep. **2-3**, 17–37.

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

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

Fourcin, A. J. (**1974**). "Laryngographic Examination of Vocal Fold Vibration," in *Ventilatory and Phonatory Control Systems*, edited by B. Wyke (Oxford U.P., London), pp. 315–326.

Fourcin, A. J. (**1981**). "Laryngographic Assessment of Phonatory Function," in *Proceedings of the Conference on the Assessment of Vocal Fold Pathology*, edited by C. L. Ludlow and M. O. Hart (The American Speech-Language-Hearing Association, Rockville, MD), pp. 116–124.

Fujimara, O., and Lindqvist, J. (**1971**). "Sweep-Tone Measurements of Vocal-Tract Characteristics," J. Acoust. Soc. Am. **49**, 541–558.

Ishizaka, K., and Matsudaira, M. (**1968**). "What Makes the Vocal Cords Vibrate?," in *The 6th International Congress on Acoustics*, edited by Y. Kohasi (Elsevier, New York), Vol. II, pp. B-9–B-12.

Ishizaka, K., and Matsudaira, M. (**1972**). "Theory of Vocal Cord Vibrations," Rep. Univ. Electro. Comm. **23**, 107–136.

Jackson, M., Ladefoged, P., Huffman, M., and Antoñanzas-Barroso, N. (**1985**). "Automated Measures of Spectral Tilt," UCLA Work. Pap. Phonet. **62**, 72–78.

Javkin, H., Antoñanzas-Barroso, N., and Maddieson, I. (in press). "Digital Inverse Filtering for Linguistic Research," J. Speech Hear. Res. **29**.

Javkin, H., and Maddieson, I. (**1983**). "An Inverse Filtering Study of Burmese Creaky Voice," UCLA Work. Pap. Phonet. **57**, 115–125.

Kirk, P., Ladefoged, P., and Ladefoged, J. (**1984**). "Using a Spectrograph for Measures of Phonation Type in a Natural Language," UCLA Work. Pap. Phonet. **59**, 102–113.

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.

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

Lyman, T. (**1974**). *Dictionary of Mong Njua* (Mouton, The Hague).

Maddieson, I., and Ladefoged, P. (**1985**). " 'Tense' and 'lax' in four minority languages of China," J. Phonet. **13**, 433–454.

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. (**1981**). "Some Relations between Glottal Air Flow and Vocal Fold Contact Area," in *Proceedings of the Conference on the Assessment of Vocal Fold Pathology*, edited by C. L. Ludlow and M. O. Hart (The American Speech-Language-Hearing Association, Rockville, MD), pp. 88–96.

Smalley, W. A. (**1976**). "The Problems of Consonants and Tone: Hmong (Meo, Miao)," in *Phonemes and Orthography: Language Planning in Ten Minority Languages of Thailand*, edited by W. A. Smalley, Pacific Linguistics Series C, No. 43 (Canberra, Australian National University).

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

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

van den Berg, J. (**1958**). "Myoelastic-Aerodynamic Theory of Voice Production," J. Speech Hear. Res. **1**, 227–244.

Acoustic analyses of infant fricative and trill vocalizations[a)]

Harold R. Bauer
Speech and Hearing Science Section, Department of Communication, 324 Derby Hall, The Ohio State University, 154 N. Oval Mall, Columbus, Ohio 43210

Ray D. Kent
Department of Communicative Disorders, University of Wisconsin, 1975 Willow Drive, Madison, Wisconsin 53706

(Received 26 March 1986; accepted for publication 10 September 1986)

Closants, or consonantlike sounds in infant vocalizations, were described acoustically using 16-kHz spectrograms and LPC or FFT analyses based on waveforms sampled at 20 or 40 kHz. The two major closant types studied were fricatives and trills. Compared to similar fricative sounds in adult speech, the fricative sounds of the 3-, 6-, 9-, and 12-month-old infants had primary spectral components at higher frequencies, i.e., to and above 14 kHz. Trill rate varied from 16–180 Hz with a mean of about 100, approximately four times the mean trill rate reported for adult talkers. Acoustic features are described for various places of articulation for fricatives and trills. The discussion of the data emphasizes (1) dimensions of acoustic contrast that appear in infant vocalizations during the first year of life, and (2) implications of the spectral data for auditory and motor self-stimulation by normal-hearing and hearing-impaired infants.

PACS numbers: 43.70.Bk, 43.70.Fq, 43.70.Jt

INTRODUCTION

The human vocal tract anatomy is drastically remodeled in the first few months of life (Wind, 1970; Bosma, 1975; Laitman and Crelin, 1976; Sasaki *et al.*, 1977). Because of this remodeling, the infant's sound production capabilities and proclivities may differ in important respects from those of the adult speaker and require study to provide a lifespan perspective on phonetic capability (e.g., Chapin *et al.*, 1982). Many of the anatomic differences are easily appreciated by inspection of Fig. 1. Among the most conspicuous differences are the neonate's shorter vocal tract, more anterior tongue position, broader and flatter oral cavity, relatively shorter pharyngeal cavity, relatively high larynx position (high enough to create engagement of the larynx and nasopharynx), and lack of dentition (thus permitting ready access of the tongue tip into the labial space). An overall geometric difference is that the neonate has a gradually sloping oropharyngeal channel. In comparison, the adult has a nearly right-angle bend between these two channels in a configuration that DuBrul (1977) called the "distinctively human craniovertebral angle." The peculiar features of the infant's vocal tract during the first few months of life should predispose to the production of front vowels (because of the anterior tongue carriage); velar, pharyngeal, and epiglottic consonants (because of the gradual slope of the oropharyngeal channel, and the opposition of epiglottis and soft palate); and nasalized vocalizations (by virtue of the engagement of larynx and nasopharynx). Indeed, studies of infants' sound preferences generally accord with these phonetic characteristics (Chen and Irwin, 1946; Irwin, 1947a,b, 1948; Irwin and Chen, 1946; Locke, 1983; Mowrer, 1980).

The infant's short vocal tract (about 7–9 cm from glottis to lips, compared to 17 cm for an adult male vocal tract) gives rise to vocal tract resonances that are proportionately higher in frequency than those of the adult. Consider that the center of the infant's acoustic vowel space, defined by observed values of $F1$, $F2$, and $F3$ (first three formant frequencies), is at about 1, 3, and 5 kHz (Kent and Murray, 1982). The corresponding frequencies of the midcentral vowel of an adult male are approximately 0.5, 1.5, and 2.5 kHz, respectively. These higher formant frequencies, in combination with the infant's higher fundamental frequency, averaging about 300–600 Hz with a range of about 50–2000 Hz (Kent, 1976; Murry and Murry, 1980), cause infant vocalizations to sound high in pitch to the adult ear. Very little information has been published on the spectral properties of infant fricatives and trills, though it is expected that infant fricative

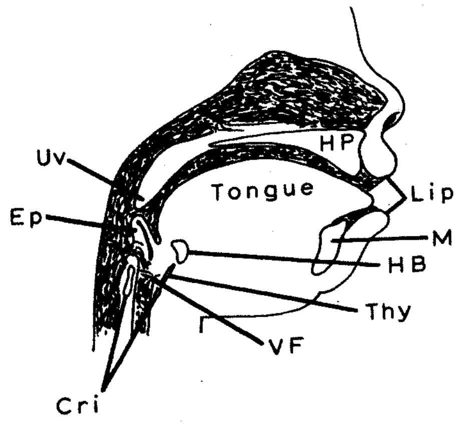

FIG. 1. Infant vocal tract showing the hard palate (HP), mandible (Man), uvula (Uv), hyoid bone (HB), epiglottis (Ep), vocal folds (VF), thyroid (Thy), and cricoid (Cri) cartilages (redrawn from Bosma, 1975).

[a)] An earlier version of this paper was presented at the meeting of the Acoustical Society of America in Chicago on 27 April 1982 [H. R. Bauer and R. D. Kent, J. Acoust. Soc. Am. Suppl. 1 **71**, S21 (1982)].

0001-4966/87/020505-07$00.80

SHIPPENSBURG
UNIVERSITY

Notice of Copyright Restrictions:

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is **not to be "used for any purpose other than private study, scholarship, or research."** If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use, that user may be liable for copyright infringement.

For more information, visit: http://www.copyright.gov/

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Problem Report:

If you have a problem with the delivery of the requested item(s), please contact us with the following information:

Missing Pages: ______________________________

Edges Cut Off: ______________________________

Unable to Read: __________ Dark __________ Blurry __________ Light

Other: ______________________________

Contact Information

Shippensburg University of Pennsylvania
Lender SQP

Contact Name: Diane Kalathas (Borrowing)	Contact Name: Teresa Strayer (Lending)
Phone: 717-477-1462	Phone: 717-477-1123 Ext. 3353
Fax: 717-477-1389	Fax: 717-477-1389
Email:dmkala@ship.edu	Email: tmstra@ship.edu