MEASURES OF THE GLOTTAL AIRFLOW WAVEFORM, EGG, AND ACOUSTIC SPECTRAL SLOPE FOR FEMALE VOICE

Eva B. Holmberg, Robert E. Hillman, Joseph S. Perkell
Massachusetts Eye and Ear Infirmary and Massachusetts Institute of Technology

ABSTRACT

Comparisons were made among aerodynamic, electroglottographic, and acoustic spectral measures for syllable production and sustained vowel phonation in comfortable and loud voice of 20 women with normal voices. Measures differed significantly between tokens having harmonic energy versus noise in the F3 region. Spectral measures added useful information to glottal waveform data about abrupt versus gradual vocal fold closing.

INTRODUCTION

The objective of this study was to examine whether acoustic spectral measures of voice production could be used to supplement measures obtained from the inverse filtered oral flow waveform that are especially sensitive to technical difficulties [1]. We focused on measures that have been found salient for vocal intensity [2] and glottal aperture [3]. Our flow-based inverse filtering technique is not without problems: 1) An accidental air leak between the subject's face and the transducer mask lowers amplitude-based flow measures that include the DC flow. 2) Unsuccessful inverse filtering that results in formant residuals superimposed on the glottal waveform can make time-based measures unreliable. 3) Low-pass filtering, used in our inverse filtering algorithm, has the undesirable effect of rounding of waveform discontinuities, such as at the instant of vocal fold closure.

Measures made from the acoustic spectrum may also assist in an objective evaluation of voice quality, such as degree of perceived breathiness. A breathy voice is the result of incomplete vocal fold closure and increased subglottal coupling [4]. Thus, this study also examines relationships between measures of the acoustic spectral slope and glottal waveform measures that are believed to correlate with increased subglottal coupling [3]. In addition, qualitative observations were made of the energy content (noise versus harmonic energy) in the third formant frequency region [5].

A measure of an "adduction quotient" on the electroglottographic (EGG) signal [6] was also included, in order to determine whether any useful information could be obtained from the EGG-based quotient that was not available from the analogous flow-based quotient.

Finally, two different elicitation materials were used: strings of repeated /pæ/ syllables and sustained phonation of /æ/ vowels. Our intention is to combine measures from the two speech tasks.

METHODS

Detailed descriptions of recording procedures, signal processing, data extraction and analyses procedures are presented in previous publications [7, 8]. In brief: twenty American females, age 20 to 43 years, with healthy voices served as subjects. They produced two different speech tasks in comfortable and loud voice: (1) strings of five repetitions of the syllable /pæ/, and (2) the vowel /æ/, sustained for 2-3 seconds. Recordings were made of oral airflow with a "Rothenberg mask" (Glottal Enterprises); intraoral air pressure with a thin catheter connected to a differential pressure transducer (Glottal Enterprises); sound pressure, with a small microphone (Sony model ECM 50) attached at a fixed, reproducible distance of 15 cm from the subject's lips; and EGG, using a laryngograph (Glottal Enterprises). The flow signal was low-pass filtered at 1100 Hz and inverse filtered to remove effects of the first formant. The EGG signal was low-pass filtered at 1710 Hz. Appropriate calibration signals were recorded for air pressure, airflow and intensity. The recorded signals were sampled at different rates, digitized simultaneously, demultiplexed and processed further in software. Measures were extracted algorithmically, with interactive monitoring, at a vowel midpoint location.

Estimates were made of *average transglottal air pressure*, (the driving force for phonation, cm H_2O). *SPL* was calculated from the RMS of the speech signal. *Glottal airflow waveform measurements* were made of: *DC flow* (the unmodulated flow, l/sec), and *flow-adduction quotient* (closed time/T) using a 30% amplitude criterion level [9]. *EGG-adduction quotient* (vocal fold contact time/T), measured at an (arbitrary) 65% criterion amplitude level. *Amplitude differences (dB) were calculated from the acoustic spectra* between: the first two harmonics (AH1-AH2); the first harmonic and the peak harmonic in the first formant (AH1-AF1); the first harmonic and the spectral peak of the third formant (AH1-AF3); the peak harmonic of the first formant and the spectral peak of the third formant (AF1-AF3). *Qualitative observations were made of the energy content in the frequency region of F3*, whether the spectrum consisted predominantly of harmonics, noise, or a mixture of harmonics and noise. *Statistical analyses* were performed to examine: differences between /pæ/ and /æ/; pairwise linear relationships between parameters; and the extent to which the parameters differed between tokens with F3 harmonic energy and those with F3 noise.

RESULTS

SPL was higher for the vowel in the syllable strings than in the sustained phonation. Analysis of covariance (ANACOVA, $p<0.05$, Bonferroni corrections, $p=0.0045$) with SPL as the covariate showed that there were no significant differences in other parameters between /pæ/ and /æ/ productions after adjustment for SPL.

Relationships between Flow- and EGG-Adduction Quotients.

Pearson product moment correlations calculated between the flow- and EGG adduction quotients, showed strong relationships ($r>0.70$) for individual speakers, whose signals were strong and noise free. The results suggest that the quotients measured at the amplitude levels of 30% (flow) and 65% (EGG) were highly related.

Relationships between Glottal Airflow Measures and the Spectral Slope.

Pearson product moment correlations between measurements of glottal airflow waveforms and spectral slope showed a relatively strong relationship between flow-adduction quotient and AH1-AH2 ($r=-0.69$). The results suggest that the degree to which the glottal waveform has a sinusoidal shape, and inversely, the degree of glottal adduction, was reflected relatively well in AH1-AH2. AH1-AH2 was also relatively strongly correlated with SPL ($r=0.69$). The relationship between flow-adduction quotient and AF1-AF3 was significant for tokens with predominantly F3 noise, but non-significant for tokens with predominantly F3 harmonic energy. Other relationships were weak in the group data. The relationships between acoustic spectral measures and glottal waveform measures were examined also for each individual speaker. A majority of the individual speakers displayed strong relationships ($r>0.70$) between flow-adduction quotient and all the spectral measures, with the exception of AF1-AF3. The data suggest that the degree to which the glottal waveform had a sinusoidal shape was reflected in ratios that included the amplitude of the fundamental, but not in the ratio which included only the higher frequency region.

Relationships between F3 Spectral Energy Content and Loudness Condition.

Simple tallies were made of the number of tokens with F3 harmonic energy, tokens with F3 noise, and tokens with mixed noise and harmonic F3 energy, in comfortable and loud voice respectively. Most tokens (122 of 240) in comfortable voice displayed a mix of harmonic energy and noise in the F3 region, followed by tokens with predominantly F3 noise (84). Few tokens (34) displayed predominantly F3 harmonic energy in comfortable voice. In contrast, in loud voice, most tokens (144 of 240) displayed harmonic F3 energy, followed by tokens with a mix of F3 harmonic energy and noise (68). Few tokens (28) in loud voice displayed F3 energy with predominantly noise.

Differences in Acoustic and Glottal Waveform Measures between Tokens with F3 Spectral Energy and F3 Noise.
Analyses of variance ($p<0.001$) showed that tokens with predominantly noise in the F3 region were associated with significantly: lower SPL; larger values of AH1-AH2, AH1-AF3, and AF1-AF3; smaller values of AH1-AF1; lower subglottal air pressure; smaller adduction quotients (flow and EGG); and higher DC flow.

DISCUSSION

The finding of higher SPL for /æ/ in the /pæ/ syllable strings than for /æ/ in sustained phonation was most likely due to differences in location of the mid-vowel data extraction; in the syllables the mid-vowel point occurs shortly after initiation of vocalization, while in sustained phonation mid-vowel occurs well into the vowel, at a point where SPL was stabilized and somewhat lower than at the beginning of the sustained vowel.

The literature has suggested that the EGG waveform contained a number of interesting features and events that could be useful for a better understanding of the underlying vocal-fold vibration pattern [6]. However, we seldom find such clear events in the EGG waveforms. In addition, we have experienced difficulties in recording EGG, for example weak and noisy signals for women, and intermittent disruptions of the signal due to gross movements of larynx that accompany changes in vocal effort. However, for speakers with strong EGG signals and clean glottal waveforms (without formant residuals), the flow and EGG quotients, measured at 30% and 65% amplitude criteria levels respectively, were highly correlated ($r>0.85$). This finding suggests that quotients from clean samples from one signal can complement the other if necessary.

A particular focus of this study was relationships between flow and acoustic parameters that have been found reflective of glottal aperture, with the goal of cross validating these measures. Gradual closing movements of a somewhat abducted membranous portion of the vocal folds should result in relatively sinusoidal glottal waveforms and small adduction quotients [3]. The result of relatively strong (negative) correlation between flow-adduction quotient and AH1-AH2 suggests that measurements of adduction quotient at the 30% level criterion was sensitive enough to differentiate among waveforms with gradual and abrupt closing, and that AH1-AH2 could be used as a substitute for adduction quotient in case of unsuccessful inverse filtering. The strong (negative) relationships between flow-adduction quotient on the one hand and AH1-AF1, AH1-AF3, and SPL on the other for the individual speakers suggest that gradual vocal fold closures resulted in an increased amplitude of the first harmonic, reduced amplitude of the first formant, steeper overall spectral slope, and reduced SPL, in agreement with previous research [2].

Ideally, high vocal fold closing velocities and abrupt reduction of the airflow should result in glottal waveforms with sharp corners between the closing and closed portions. However, a detrimental effect of low-pass filtering at 1100 Hz is a "rounding" of waveform discontinuities, which could have an influence on waveforms associated with high vocal fold closing velocities. Waveforms which result from more gradual closing movements (therefore already rounded) would be relatively uninfluenced by the low-pass filtering rounding effect. These waveforms, with F3 excited by noise, have a significant relationship between flow-adduction quotient and AF1-AF3. In contrast, waveforms with F3 excited by harmonic energy have a non-significant relationship between flow-adduction quotient and AF1-AF3. The filter-induced rounding of waveforms with more abrupt closures may account for the lack of significant correlation between adduction quotient and AF1-AF3. In other words, the sharp corners in those waveforms were obscured, and reliable adduction quotient measurements were precluded by the low-pass filtering at 1100 Hz. However, the effects of the sharp discontinuities were preserved in the AF1-AF3 value derived from the full-bandwidth acoustic spectra. These results suggest that the spectral measurement of AF1-AF3 may serve as a useful complement to the flow-based adduction quotient, especially when there are high vocal fold closing velocities [10].

In normal phonation, DC flow in the glottal waveform has been assumed to reflect airflow that passes through a posterior glottal "chink". A large chink would increase the subglottal coupling with reduced high frequency energy and reduced SPL. However, neither the relationship between underlying physiology and the DC flow [11], nor the acoustic effect of the DC flow is completely understood. None of the acoustic measures varied systematically with DC flow. Thus, DC flow data must be interpreted with caution.

CONCLUSIONS

The following conclusions could be drawn from the results of this study:

1) Comparisons between measures obtained from the vowel in /pæ/ syllables and those obtained from sustained /æ/ phonation can be made, as long as SPL differences are controlled for.

2) Adduction quotient, measured at a 30% amplitude level on the glottal waveform is sensitive enough to differentiate among waveforms with gradual and abrupt closing portions in data for individual subjects.

3) Measurements of the amplitude difference between the two first harmonics (AH1-AH2) may be used as a substitute measure for flow-adduction quotient, in cases of unsuccessful inverse filtering that make measurements of adduction quotient unreliable.

4) The flow- and EGG adduction quotients, measured at 30% and 65% levels respectively, may serve to complement one another.

5) AF1-AF3 may serve as a useful complement to measurements of maximum flow declination rate, especially in voices with very high closing vocal fold velocities that cannot be reflected accurately in a flow waveform that is low-pass filtered at 1100 Hz.

ACKNOWLEDGMENTS

This study was supported by a grant from The National Institutes of Health.

REFERENCES

[1] Holmberg, E.B., Hillman, R.E., Perkell, J.S., Guiod, P.C., and Goldman, S.L. (in press), "Comparisons among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice", J. Speech and Hear. Res.

[2] Fant, G. (1979), "Glottal source and excitation analysis", Quarterly Report, 1, Speech Transmission Laboratory and Music Acoustics, The Royal Institute of Technology, Stockholm, pp. 85-107.

[3] Stevens, K.N. (1977), "Physics of laryngeal behavior and larynx modes", Phonetica, 34, pp. 264-279.

[4] Södersten, M., & Lindestad, P-A. (1990), "Glottal closure and perceived breathiness during phonation in normally speaking subjects", J. Speech and Hear. Res., 33, pp. 601-611.

[5] Klatt, D.H., & Klatt, L.C. (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am., 87, pp. 820-857.

[6] Childers, D.G., Hicks, D.M., Moore, G.P., Eskenazi, L., & Lalwani, A.L. (1990) "Electroglottography and vocal fold physiology", J. Speech and Hear. Res., 33, pp. 245-254.

[7] Holmberg, E.B. Hillman, R.E., & Perkell, J.S. (1988), "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", J. Acoust. Soc. Am. 84, pp. 511-529.

[8] Perkell, J.S., Holmberg, E.B., & Hillman, R.E. (1991), "A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production", J. Acoust. Soc. Am., 54, pp. 1777-1781.

[9] Colton, R.H., Brewer, D.W., & Rothenberg, M. (1985), "Vibratory characteristics of patients with voice disorders", Oral presentation, Symposium and voice Acoustics and Dysphonia, Gotland, Sweden, 1985.

[10] Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M, & Vaughan, C. (1989), Objective assessment of vocal hyperfunction." J. Speech and Hear. Res., 32, pp. 373-392.

[11] Hertegård, S., Gauffin, J. (1995), "Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography", J. Speech and Hear. Res., 38, pp. 85-100.