SUB-GLOTTAL RESONANCES IN FEMALE SPEAKERS AND THEIR EFFECT ON VOWEL SPECTRA

Helen M. Hanson (1) and Kenneth N. Stevens (2)
(1) Div. of Applied Sciences, Harvard Univ., Cambridge MA 02138, USA
(2) Research Lab of Electronics and Dept. of Electrical Engineering and Computer Science, MIT, Cambridge MA 02139, USA

ABSTRACT

Resonances of the subglottal system often influence the acoustic characteristics of vowels. These influences are manifested as extra peaks in vowel spectra and as discontinuities in apparent formant movements. Data from vowels produced by a number of female speakers show that the magnitudes of these effects are correlated with acoustic measures indicating the degree of glottal abduction used by the speakers during phonation.

1. INTRODUCTION

We often observe prominences in the spectra of vowels that are not attributable to formants that are natural frequencies of the vocal tract [1]. An example of a vowel spectrum with these extra prominences is given in Fig. 1. In this example we see an extra spectral peak at about 1600 Hz. Adjacent to this spectral peak is a valley indicating the presence of a zero or antiresonance in the vocal-tract transfer function. These irregularities in the spectrum result from acoustic coupling, through the glottis, between the vocal tract and the trachea.

We report here some data on the prominences arising from the tracheal resonances for a number of female speakers, and we relate these data to other spectral measurements reported previously for these speakers [2][3]. We turn first to some theoretical background.

2. THEORY

The subglottal and supraglottal systems are coupled through a narrow glottal opening which has a resistance R_g and an acoustic mass M_g. As a first

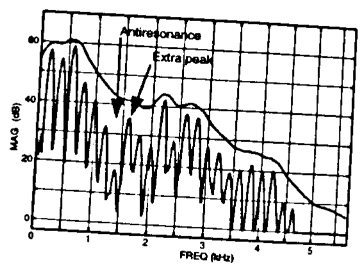

Figure 1: *Spectrum of vowel /ε/ produced by a female speaker, showing extra peak and antiresonance due to acoustic coupling to the trachea. Time window of spectrum is 22.3 ms.*

approximation we represent the acoustic source as paired volume-velocity sources U_s as shown in Fig. 2. Z_t and Z_v are the impedances looking into the trachea and vocal tract.

The transfer function U_m/U_s is characterized by poles, which are the natural frequencies of the coupled system, together with zeros at frequencies for which $Z_t = \infty$. These zeros are the natural frequencies of the subglottal system when the glottis is closed.

Measurements of the subglottal resonant frequencies have been reported by [4][5] and others. Typical values of the lowest three of these frequencies

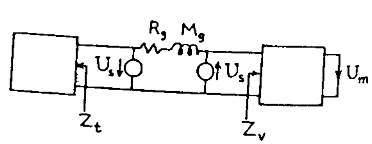

Figure 2: *Equivalent circuit showing vocal tract and subglottal system connected to sources and coupled through the glottis. See text.*

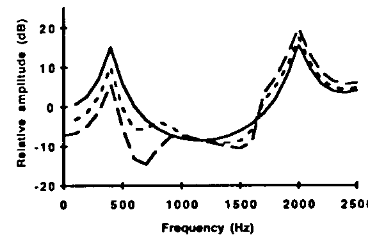

Figure 3: *Calculated transfer function U_m/U_s of the vocal tract when there is no acoustic coupling to the trachea (solid line), when $A_g \approx 0.02$ cm² (dotted line), and when $A_g \approx 0.05$ cm² (dashed line). The tracheal resonances are at 650 and 1600 Hz.*

with a closed glottis are estimated to be 700, 1700, and 2300 Hz for adult female speakers (cf. [5]). The bandwidths of these resonances are about 200 Hz [4]. The poles f_p of the transfer function due to the subglottal cavity are close to the zeros f_z noted above, and the amount of separation between a pole and a zero in a pair depends on the size of the glottal opening, i.e., the values of R_g and M_g. For different speakers, the average glottal area, and hence the average values of R_g and M_g, may be different. For some speakers, this difference is due to the fact that the glottis does not completely close during the so-called closed phase of vibration.

In order to estimate the effect of the glottal opening on the spectrum of the radiated sound, we can calculate the transfer function U_m/U_s for various values of the average glottal area A_g. We assume that the impedance looking into the vocal tract is small compared with the impedance of the glottis. This assumption is reasonable as long as the subglottal resonance is not too close to a natural frequency of the vocal tract (a formant). The frequency of the pole is estimated to be the natural frequency of the subglottal system when it is terminated by the glottal impedance.

Calculations of the vocal-tract transfer function for a typical front-vowel configuration with formants well separated from the tracheal resonances are shown in Fig. 3 for two different glottal areas. When the glottal area

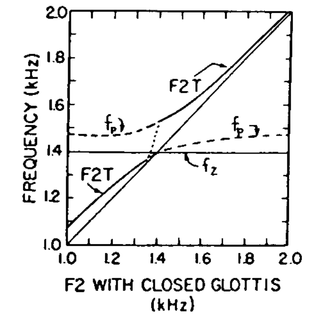

Figure 4: *Estimates of the frequencies of the two poles and zero in the vicinity of F2 when there is coupling to the trachea through a partially open glottis. The zero f_z is assumed to be fixed at 1400 Hz. F2T represents the pole corresponding to F2, shifted by the influence of the tracheal system. The solid lines indicate the most prominent spectral peak, which shows an abrupt jump in frequency (dotted line) when F2 is just below f_z. The dashed lines represent f_p which is less prominent in the spectrum.*

is larger, substantial additional prominences appear in the vowel spectrum, whereas for the smaller area the effect of the tracheal resonances is small.

If the frequency of a formant is close to a subglottal resonance, the subglottal coupling will have an influence on the spectral representation of the formant. For example, if a formant passes through the region of a subglottal resonance, interference is expected. This interference effect is illustrated in Fig. 4. The abscissa is the frequency *F2* that would exist if the glottis were closed, and the ordinate is the actual frequencies of the poles and the zero for the coupled system. The second formant frequency *F2* increases from 1100 to 1800 Hz, passing through the tracheal resonance f_z, which in this example is fixed at 1400 Hz. When *F2* is well separated from f_z, there is a small upward shift in the pole representing *F2*, and there is a pole-zero pair f_z and f_p due to tracheal coupling. When *F2* approaches f_z, the pole-zero-pole combination creates two nearby spectral peaks. When $F2 < f_z$, the lower of these peaks is dominant, but when

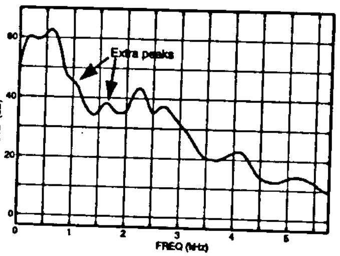

Figure 5: *Average spectrum of vowel /ɛ/ produced by the speaker represented in Fig. 1. The average was obtained over five repetitions of the vowel (in the word* bɛd*). Evidence for extra peaks due to tracheal coupling is shown.*

$F2$ passes upward through f_z the upper peak becomes dominant. Thus there is a discontinuous upward jump in the dominant spectral peak as $F2$ increases through the subglottal resonance. A similar effect occurs when $F1$ passes through the lowest tracheal resonance.

This theoretical analysis suggests, then, that there are two kinds of acoustic evidence for acoustic coupling to tracheal resonances: one is the presence of spectral prominences in addition to the prominences due to vocal-tract resonances or formants, and the other is the disruption of prominences due to formants as they pass through frequency ranges of tracheal resonances. The latter effect should be observable in diphthongs like /ai/, where $F1$ traverses downward through the lowest tracheal resonance and $F2$ follows an upward-moving trajectory through the second tracheal resonance.

Theory also predicts that tracheal resonances should be more evident in vowel spectra for individuals who phonate with a glottis that remains partially open throughout a glottal cycle. Such individuals are known to exhibit a greater high-frequency tilt in the glottal spectrum and a greater $F1$ bandwidth due to increased acoustic losses at the partially open glottis.

3. EXPERIMENTAL DATA

Two kinds of acoustic data were obtained from vowels produced by 22 speakers. From spectra of the vowels /ɛ æ ʌ/ in CVC words, estimates were

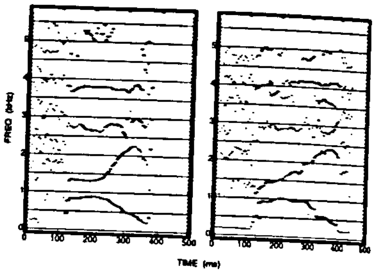

Figure 6: *Examples of trajectories produced by an LPC-based formant tracker for the word* bide *produced by two female speakers. On the left, the $F1$ and $F2$ tracks are smooth, with no discontinuities, but on the right, there are discontinuous jumps as the formants pass through tracheal resonances.*

made of the degree of perturbation by extra peaks and valleys that could be ascribed to tracheal resonances. The spectrum for each vowel was an average spectrum over the vowel portions of five repetitions of the words, using a short time window (7 ms) calculated every millisecond. An average spectrum of the vowel /ɛ/ for the speaker of Fig. 1 is shown in Fig. 5. It was thought that such an average spectrum should be effective for showing prominences (such as those due to tracheal resonances) that remain relatively fixed in frequency over time. The deviation of each vowel spectrum in terms of extra prominences was rated by two observers on a scale from 0 to 2.

A second type of acoustic data examined the tracking of the first and second formant peaks in the diphthong /ai/ in the utterance bide. Formant tracks obtained using a standard LPC algorithm are shown in Fig. 6 for two speakers. For one of the speakers, the formants are tracked smoothly, except for minor ripples due to the interaction of the fundamental frequency and the formants. For the other speaker, there is an abrupt discontinuity in both formant tracks, presumably due to the influence of tracheal resonances.

The $F1$ and $F2$ tracks for this diphthong produced five times by each speaker were examined, and cases with a significant discontinuity in either track were noted. To qualify as a discontinuity induced by a subglottal res-

onance, it must occur in the frequency range 500-1000 Hz for $F1$ and 1500-2000 Hz for $F2$. Each speaker was rated by the number of such discontinuities, ranging from 0 to 10.

These two measures — spectral deviations caused by extra prominences (EP), and discontinuities in formant tracks for a diphthong (DF) — were examined in relation to other acoustic measures. These other measures are theoretically related to the size of a fixed opening in the glottis during the "closed" phase of the glottal vibration cycle, and should increase as the cross-sectional area of the opening does [2][3]. The measures are: (1) $H1$-$A1$, the difference (in dB) between the amplitude of the first harmonic and the amplitude of the largest harmonic in the vicinity of the first formant. This difference is related to the bandwidth of the first formant. (2) $H1$-$A3$, where $A3$ is the amplitude of the third formant peak. This is a measure of spectral tilt. (3) The bandwidth $B1$ of the first formant, as determined by the rate of decay of the $F1$ waveform during the initial (most closed) part of the glottal cycle. (4) Estimates N_w and N_s of noise excitation in the $F3$ waveform and high-frequency spectrum, respectively [5].

The correlations between EP, DF and the spectral measures are summarized in Table 1. The correlations between DF and the spectral measures are all quite high, particularly DF and spectral tilt. The correlations for EP are smaller, possibly due to the subjectivity of this measure. It is clear from these correlations that when spectral measures indicate a significant glottal opening or "chink," evidence for tracheal resonances appear in the vowel spectrum. The effect of the tracheal resonances on the spectrum increases as the size of the opening increases.

4. CONCLUSION

Tracheal resonances can introduce significant modifications in the vowel spectra for some speakers. These are speakers for whom other spectral measures such as spectral tilt indicate some glottal abduction during the "closed phase" of glottal vibration. Tracheal resonances can interfere with the estimation of formants from vowel spectra

Table 1: *Correlations between two measures of the prominences of tracheal resonances and several spectral measures (see text) obtained from vowels produced by 22 female talkers.*

Measure	*EP*	*DF*
H1-A1	0.50	0.68
H1-A3	0.70	0.83
N_w	0.68	0.82
N_s	0.57	0.79
DF	0.62	1

and thus have implications for formant tracking and speech recognition systems. The effects of these resonances on both formant location and prominences can also influence vowel space and quality. Finally, our observations of these effects suggest that the simple source-filter theory may not always be adequate, even for modal phonation.

ACKNOWLEDGEMENTS

This work was supported in part by NIH Grant DC00075.

REFERENCES

[1] Fant, G. (1972), "Subglottal formants," *STL-QPSR 1*, Royal Institute of Technology, Stockholm, 1-12.
[2] Stevens, K.N. and H.M. Hanson (1995), "Classification of glottal vibration from acoustic measurements," in O. Fujimura and M. Hirano (eds.) *Vocal Fold Physiology: Voice Quality Control*, San Diego: Singular, 147-170.
[3] Hanson, H.M. (1995), "Glottal characteristics of female speakers," Ph.D. Thesis, Harvard Univ., Cambridge MA.
[4] Ishizaka, K., K.M. Matsudaira, and T. Kaneko (1976), "Input acoustic-impedance measurement of the subglottal system," *J. Acoust. Soc. Am.*, Vol. 60, 190-197.
[5] Klatt, D.H. and L.C. Klatt (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, Vol. 87, 820-857.