DISCRIMINATING SPEAKERS WITH VOCAL NODULES USING AERODYNAMIC AND ACOUSTIC FEATURES

Jeff Kuo

Bell Laboratories, Lucent Technologies
600 Mountain Ave, Murray Hill, NJ 07974
kuo@research.bell-labs.com

Eva B. Holmberg, Robert E. Hillman

Massachusetts Eye and Ear Infirmary
243 Charles Street
Boston, MA 02114

ABSTRACT

This paper demonstrates that linear discriminant analysis using aerodynamic and acoustic features is effective in discriminating speakers with vocal-fold nodules from normal speakers. Simultaneous aerodynamic and acoustic measurements of vocal function were taken of 14 women with bilateral vocal-fold nodules and 12 women with normal voice production. Features were extracted from the glottal airflow waveform and peaks in the acoustic spectrum for the vowel /æ/. Results show that the subglottal pressure, air flow, and open quotient are increased in the nodules group. Estimated first-formant bandwidths are increased, but result in minimal change in the first-formant amplitudes. There is no appreciable decrease in high frequency energy. Speakers with nodules may be compensating for the nodules by increasing the subglottal pressure, resulting in relatively good acoustics but increased air flows. The two best features for discrimination are open quotient and subglottal pressure.

1. INTRODUCTION

The modulation of air flow from the lungs by the vocal folds creates an excitation sound source during voice production. In disordered voice production, the vibratory characteristics of the vocal folds may be affected by the vocal fold pathology. We have a limited understanding of how a particular voice pathology affects the speech production. How does it influence the air flow and acoustic characteristics? By how much? For example, in many types of voice disorders such as vocal nodules, the closure of the vocal folds during vibration is not complete [1]. Theoretically, an incomplete closure leads to increased glottal losses and increased first-formant bandwidth. How big is this effect and does it affect the voice substantially?

A number of studies have reported aerodynamic measurements of speakers with vocal nodules [2, 3]. The contributions of this paper include the following. First, in addition to aerodynamic measurements, this paper also includes simultaneous acoustic measurements not previously reported. Secondly, we propose the use of these features to discriminate speakers with nodules from normal speakers using linear discriminant analysis and show which features are more effective for discrimination. Lastly, we address why increased glottal losses from the incomplete glottal closure and larger open quotient do not greatly affect the first-formant amplitude.

2. EXPERIMENT

A total of 26 women participated in the present investigation. 12 of them were normal control subjects with no history of voice problems. The other 14 women have bilateral vocal nodules, confirmed by an otolaryngologist using strobolaryngoscopy, and are judged to be mildly dysphonic.

The recording procedures, signal processing, data extraction, and analyses procedures follow those given in [4]. Each subject was seated in a sound-isolated booth and a Rothenberg mask [5] was used to measure the oral airflow during speech. A transducer fitted to the mask and inserted between the lips was used to measure the intraoral pressure during bilabial obstruents such as /p/. The subject was asked to say a sequence of [pæpæpæ...] at comfortable and loud speaking levels.

The oral airflow signal was low-pass filtered at 1100 Hz and inverse-filtered to derive the glottal airflow, from which aerodynamic features are extracted. The acoustic speech signal was also recorded simultaneously at a distance of 15 cm, and digitized at a sampling rate of 10 kHz. The SPL was calculated from the root-mean-square amplitude of the acoustic signal. All features were extracted in the middle of the vowel /æ/ except for the subglottal pressure. The subglottal pressure was inferred from interpolation of the adjacent intraoral pressures during the closure for the /p/ in the utterance [pæpæpæ...].

For each of the normal subjects, there were 7–9 tokens of [pæ] recorded under each condition of comfortable voice and loud voice. Among all the normal speakers, there were a total of 102 tokens for comfortable voice and 103 tokens for loud voice. The subjects with nodules had between 7–10 tokens recorded under each condition of comfortable voice and loud voice. Among all the speakers with nodules, there were a total of 126 tokens for comfortable voice and 120 tokens for loud voice.

3. FEATURES

3.1. Aerodynamic Features

The aerodynamic features discussed in this paper are the following, some of which are shown in Figure 1:

- **Subglottal Pressure**: (not shown in figure) the air pressure below the vocal folds which serves as the driving force for vocal fold vibration.
- **Average Flow**: (not shown) the average glottal volume velocity flow. It depends on the minimum flow, AC flow, and open quotient.

- **Minimum Flow**: minimum value of the glottal flow. A nonzero minimum flow is associated with incomplete closure of the vocal folds, and the residual area is known as the chink area.
- **AC Flow**: peak-to-peak amplitude of the glottal flow waveform, reflecting magnitude of vocal fold oscillation. It also depends on the subglottal pressure.
- **Open Quotient**: ratio of time the vocal folds are open relative to the entire period of vibration. For calculation of the open quotient, places where the flow level was 30% of the difference between peak and minimum flow were identified and used to define arbitrary times of "opening" and "closing" [6].
- **MFDR** (Maximum Flow Declination Rate): maximum negative slope of the glottal flow waveform, associated with the abrupt adducting movement of the vocal folds and related to the level of the output sound pressure.

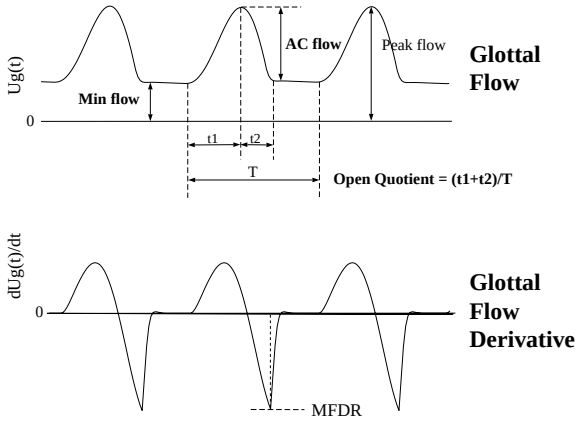

Figure 1: Definition of the aerodynamic features associated with the glottal flow waveform $U_g(t)$ and the flow derivative waveform dU_g/dt. The Maximum Flow Declination Rate (MFDR) is indicated on the flow derivative waveform.

3.2. Acoustic Features

Figure 2 shows an example of the spectrum of a speech signal of a vowel which has been multiplied by a 25.6ms Hamming window. The periodic peaks are harmonics due to the periodicity of the speech signal within vowels when the vibration of the vocal folds serves as the sound source. The various prominences in the overall shape are due to the resonances of the vocal tract or to the amplitudes of different harmonics, as shown. The acoustic features reported in this paper include the following:

- **SPL**: sound pressure level.
- **F0**: fundamental frequency.
- **H1-A1**: difference in amplitude between the first harmonic and the harmonic closest to the first formant. This quantity is a measure of the first-formant bandwidth B1 – a doubling of B1 reduces A1 by 6 dB.
- **H1-A3**: difference in amplitude between the first harmonic and the harmonic closest to the third formant. This difference is one measure of spectral tilt (amplitude of high frequency energy relative to low frequency energy.)
- **A1-A3**: difference in amplitude between the harmonic closest to the first formant and that closest to the third formant – another measure of spectral tilt.
- **H1-H2**: difference in amplitude between the first two harmonics.

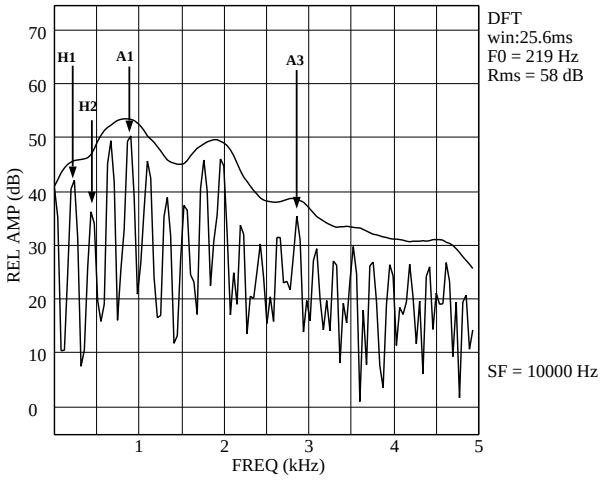

Figure 2: Example of the acoustic features associated with the speech spectrum of a vowel. The acoustic features are measured from the lower curve, which is the spectrum with a time window of 25.6 ms. The upper curve is a smoothed and offset version of the spectrum, and is not used to determine the amplitude of the labeled acoustic features.

3.3. Derived Features

From the flow and subglottal pressure measurements, we can derive glottal area and first-formant bandwidth estimates.

3.3.1. Glottal areas

We use the following equation that relates the pressure difference P across a constriction of cross-sectional area A to the volume velocity flow U [7]:

$$P = \frac{\rho U^2}{2A^2}. \tag{1}$$

Using Equation 1, the chink area A_{ch} can be obtained from the minimum flow and the subglottal pressure.

3.3.2. Bandwidth due to Glottal Losses

The contribution of glottal losses to the first-formant bandwidth during the closed phase can be approximated from the minimum flow using the following equation (assuming the vocal tract is a uniform tube of length l_t and cross-sectional area A_t) [8]:

$$B_g = \frac{\rho c^2}{\pi A_t l_t R_{ch}(1+K_m)}, R_{ch} = \frac{\sqrt{2\rho P_s}}{A_{ch}}, \tag{2}$$

where A_{ch} is the area of the glottal chink at the time that the vocal folds achieve maximum closure, estimated from Equation 1.

The factor K_m incorporates the effects of the acoustic mass and is equal to:

$$K_m = (\frac{2\pi F1 M_{ch}}{R_{ch}})^2 = (\frac{2\pi F1 \rho l_{ch}}{\sqrt{2\rho P_s}})^2, \quad (3)$$

where $M_c = \frac{\rho l_{ch}}{A_{ch}}$ and l_{ch} is the thickness of the folds. For the values of $F1 = 860$ Hz, $l_{ch} = 0.3$ cm, $P_s = 7800$ dynes/cm^2, K_m is about 0.2, so it contributes only about 20% to the denominator of the bandwidth equation.

The first-formant bandwidth $B1$ is then determined by adding the contribution due to the glottal losses to the baseline bandwidth due to the vocal tract losses, assumed to be 50 Hz [8]:

$$B1 = 50 \text{ Hz} + B_g. \quad (4)$$

Similarly, the time-averaged bandwidth $< B_g >_{av}$ over the whole glottal cycle can be approximated from the average flow. Because the open quotient of the glottal flow waveforms for speakers with nodules is larger, the average first-formant bandwidth will be larger since B_g increases dramatically at the opening of the glottis.

4. RESULTS

Tables 1 and 2 show the mean value of each feature for the nodules group and the normal group under comfortable and loud voice conditions. The tables are divided into three categories: the aerodynamic features, the acoustic features, and the derived features.

The speakers with nodules are on average using higher subglottal pressures than normal. The other aerodynamic features of average flow, AC flow, minimum flow, MFDR, and open quotient are also higher on average. The SPL for speakers with nodules are on average 2-3 dB larger. Their fundamental frequencies are not significantly different from normal. The other acoustic features are all less than 3 dB different on average. For the derived features, speakers with nodules have on average a slightly larger chink area and bigger bandwidth associated with glottal losses.

For the aerodynamic features, the mean values of the nodules group are above one standard deviation from the mean values of the normal group. In contrast, the mean values of the nodules group are in general less than one standard deviation away from the normal mean values. Thus we would expect that the aerodynamic features will be more useful in discriminating between the two groups.

Applying linear discriminant analysis [9] using different combinations of features, we obtain the discrimination results in Table 3. Using all the aerodynamic and acoustic features, we obtain an error rate of 4%, equally distributed between false positives and false negatives. Without the use of acoustic features, with only aerodynamic features, an error rate of 5% is obtained. Just using the acoustic features results in a 24% error rate, a dramatic drop in performance. We also see that the open quotient and the subglottal pressure are the two most important aerodynamic features, yielding a 6% error rate. The most important acoustic features are SPL, H1-A1, and H1-H2, giving an error rate of 24%.

Table 4 shows some cross-validation results using data of half of the speakers for training the discriminant functions and the other half for testing. We see that there is a significant rise in the error rates for many feature combinations, but the combination of open quotient and subglottal pressure seems to hold up well. The training data are probably not sufficient, since they rely on only 6-7 speakers to represent each group.

COMFORTABLE VOICE

Aerodynamic features:	Normals	Nodules
Subglottal Pressure (cm H_2O)	6.0(1.1)	10.5(3.0)
Average Flow (l/s)	0.14(0.04)	0.29(0.10)
AC Flow (l/s)	0.17(0.05)	0.28(0.10)
Minimum Flow (l/s)	0.08(0.03)	0.17(0.09)
MFDR (l/s^2)	261.6(113.8)	387.6(163.3)
Open Quotient(%)	47(6)	59(5)
Acoustic features:		
SPL (dB)	77.5(3.6)	80.1(4.4)
F0 (Hz)	203(15)	204(14)
H1-A1 (dB)	-0.9(5.1)	0.7(5.1)
H1-A3 (dB)	25.1(5.2)	24.3(5.6)
H1-H2 (dB)	5.7(2.8)	7.1(3.4)
A1-A3 (dB)	26.0(5.3)	23.5(5.8)
Derived features:		
Area_{chink} (cm^2)	0.03(0.01)	0.04(0.02)
B_g (Hz)	70.0(22.2)	85.0(46.5)
$< B_g >_{av}$ (Hz)	102.0(21.4)	130.6(40.5)

Table 1: Group means and standard deviations of aerodynamic, acoustic, and derived features for comfortable voice.

LOUD VOICE

Aerodynamic features:	Normals	Nodules
Subglottal Pressure (cm H_2O)	9.0(1.5)	14.6(4.0)
Average Flow (l/s)	0.13(0.04)	0.30(0.14)
AC Flow (l/s)	0.23(0.05)	0.41(0.14)
Minimum Flow (l/s)	0.06(0.03)	0.12(0.11)
MFDR (l/s^2)	524.4(161.2)	802.9(300.7)
Open Quotient(%)	38(4)	57(7)
Acoustic features:		
SPL (dB)	86.6(3.2)	88.8(4.1)
F0 (Hz)	232(18)	234(22)
H1-A1 (dB)	-9.3(4.3)	-7.8(3.8)
H1-A3 (dB)	13.8(7.0)	13.9(5.8)
H1-H2 (dB)	2.3(2.5)	4.1(2.6)
A1-A3 (dB)	23.2(5.0)	21.8(4.7)
Derived features:		
Area_{chink} (cm^2)	0.015(0.007)	0.024(0.021)
B_g (Hz)	34.2(14.5)	43.2(36.1)
$< B_g >_{av}$ (Hz)	65.1(16.6)	95.9(32.0)

Table 2: Group means and standard deviations of aerodynamic, acoustic, and derived features for loud voice.

Feature Combination	No. of Features	%errors	%false positive	%false negative
All	11	4	2	2
All aerodynamic	6	5	2	3
All acoustic	5	24	11	14
Aerodynamic: (OQ,Pr)	2	6	3	2
Acoustic: (SPL,H1-A1,H1-H2)	3	24	10	14

Table 3: Classification error rates using linear discriminant analysis for different combinations of features.

Feature Combination	No. of Features	%errors	%false positive	%false negative
All	11	8	3	5
All aerodynamic	6	12	7	5
All acoustic	5	41	13	28
Aerodynamic: (OQ,Pr)	2	6	4	3
Acoustic: (SPL,H1-A1,H1-H2)	3	37	12	25

Table 4: Classification error rates of a test set for different combinations of features. The discriminant functions are determined using a training set of speakers different from the test set.

5. DISCUSSION

In patients with vocal nodules, the vocal folds do not close completely during the vibration cycle, resulting in a posterior (and occasionally anterior) glottal chink at the point of maximal closure. In addition, the larger open quotient implies a larger time-averaged glottal area. The glottal chink and larger average area are expected to result in a broadening of the first-formant bandwidth and a decrease in the amplitude of the first formant. However, from Tables 1 and 2, we see that the average difference in H1-A1 for the two groups is only about 1.5 dB. Why is there only a small difference in the amplitude of the first formant between the nodules group and the normal group?

First of all, normal female speakers of American English often have posterior glottal chinks, particularly at soft voice, although the chinks are smaller than those in the group with nodules. Secondly, the contribution to the first-formant bandwidth by glottal losses depends not only on the area of the chink but also on the subglottal pressure, according to Equation 2. The increased subglottal pressures used by speakers with nodules tend to decrease B_g. Thirdly, the increase in bandwidth does not result in a significant decrease in the amplitude of the first formant. Tables 1 and 2 show that the contribution of glottal losses to the first-formant bandwidth is on average increased in the nodules group. Using Equation 4, we calculate that the first-formant bandwidth is increased 1.1-1.26 times, which translates to a decrease in the first-formant amplitude of only 0.9-2.1 dB.

H1-H2 is larger on average for the nodules group, and this difference is consistent with the larger open quotient. However, whereas there is a good correlation between the open quotient and H1-H2 for the normal speakers ($\rho = 0.77$), there is little correlation between these two features in the nodules group ($\rho = 0.02$). Therefore, although open quotient is one of the best aerodynamic features to distinguish the two groups, H1-H2 is not a good acoustic feature.

Although one might have expected a larger spectral tilt (decrease in high frequency energy, larger A1-A3) for the nodules group, A1-A3 is smaller for the nodules group than for the normal group. There are several possible reasons for this observation. First, the spectral tilt is smaller for the nodules group because the speakers are speaking louder and spectral tilt tends to decrease with loudness. Secondly, the first-formant bandwidth is larger for the nodules group, and this increased B1 will tend to lower A1 with respect to A3. The third-formant bandwidth depends more on the radiation losses than on the glottal losses and is therefore less affected by a larger glottal area.

6. SUMMARY

We have demonstrated that quantitative aerodynamic and acoustic features are effective in discriminating speakers with vocal nodules from normal speakers, using linear discriminant analysis. Aerodynamic features, in particular open quotient and subglottal pressure, are better than acoustic features for discrimination. We also showed using an acoustic model that the increased glottal areas and consequent increased first-formant bandwidths result in a decrease in the first-formant amplitude of only about 1-2 dB.

7. ACKNOWLEDGEMENTS

This work was supported in part by NIH grants DC00075 and DC00266 and was part of thesis work [10] done under Professor Ken Stevens at MIT.

8. REFERENCES

[1] R. H. Colton, P. Woo, D. W. Brewer, B. Griffin, and J. Casper, "Stroboscopic signs associated with benign lesions of the vocal folds," *J. Voice*, vol. 9, no. 3, pp. 312–325, 1995.

[2] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Objective assessment of vocal hyperfunction: An experimental framework and initial results," *J. Speech and Hearing Res.*, vol. 32, pp. 373–392, 1989.

[3] C. M. Sapienza and E. T. Stathopoulos, "Respiratory and laryngeal measures of children and women with bilateral vocal fold nodules," *J. Speech and Hearing Res.*, vol. 37, pp. 1229–1243, Dec. 1994.

[4] J. S. Perkell, E. B. Holmberg, and R. E. Hillman, "A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production," *J. Acoust. Soc. Am.*, vol. 89, pp. 1777–1781, Apr. 1991.

[5] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J. Acoust. Soc. Am.*, vol. 53, pp. 1632–1645, 1973.

[6] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *J. Speech and Hearing Res.*, vol. 38, pp. 1212–1223, Dec. 1995.

[7] S. Hertegård and J. Gauffin, "Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography," *J. Speech and Hearing Res.*, vol. 38, pp. 85–100, Feb. 1995.

[8] H. Hanson, *Glottal characteristics of female speakers*. PhD thesis, Harvard University, Cambridge, MA, 1995.

[9] B. F. J. Manly, *Multivariate Statistical Methods: A Primer*. New York: Chapman and Hall, 2nd ed., 1994.

[10] H.-K. J. Kuo, *Voice Source Modeling and Analysis of Speakers with Vocal-Fold Nodules*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1998.