On the Interdependencies between Voice Quality, Glottal Gaps, and Voice-Source related Acoustic Measures

Yen-Liang Shue, Gang Chen, and Abeer Alwan

Department of Electrical Engineering, University of California, Los Angeles

yshue@ee.ucla.edu, gangchen@ee.ucla.edu, alwan@ee.ucla.edu

Abstract

In human speech production, the voice source contains important non-lexical information, especially relating to a speaker's voice quality. In this study, direct measurements of the glottal area waveforms were used to examine the effects of voice quality and glottal gaps on voice source model parameters and various acoustic measures. Results showed that the open quotient parameter, cepstral peak prominence (*CPP*) and most spectral tilt measures were affected by both voice quality and glottal gaps, while the asymmetry parameter was predominantly affected by voice quality, especially of the breathy type. This was also the case with the harmonic-to-noise ratio measures, indicating the presence of more spectral noise for breathy phonations. Analysis showed that the acoustic measure $H_1 - H_2$ was correlated with both the open quotient and asymmetry source parameters, which agrees with existing theoretical studies.

Index Terms: voice source, voice quality, acoustic measures

1. Introduction

An essential component of the human speech production system is the voice source which provides excitation to the vocal tract. The voice source contains important non-lexical information such as cues relating to a speaker's voice quality, prosody and emotional status. In medical applications, analysis of the voice source can aid in the diagnosis of vocal cord diseases.

There are two main methods for voice source analysis: 1) estimating the time-domain source signal from the speech signal, and 2) using acoustic measures which are related to the voice source. Accurate extraction of the time-domain source signal is non-trivial as it requires the decoupling of a non-linear system between the voice source and the vocal tract. Instead of directly estimating the voice source signal, one can study acoustic measures which are related to the voice source or voice quality. These measures are usually estimated from speech spectra; some common examples include $H_1^* - H_2^*$ (difference between the first two spectral harmonic magnitudes, corrected for the effects of the vocal tract), $H_1^* - A_3^*$ (difference of the first harmonic magnitude and the spectrum magnitude at the third formant frequency, corrected for the effects of the vocal tract), *CPP* (cepstral peak prominence), and spectral noise measures. While there is evidence that relate some acoustic measures to certain voice qualities (for example $H_1^* - H_2^*$ is related to breathiness), there is a general lack of empirical data linking these acoustic measures to the physiological movements of the vocal folds. Knowing how acoustic measures relate to vocal fold configurations could lead to a better understanding of the human speech production system and more accurate voice source analyses.

$H_1^* - H_2^*$ and its uncorrected version, $H_1 - H_2$, has often been taken as a correlate of the open quotient (*OQ*) [1], which is broadly defined as the proportion of time the vocal folds are open during a phonation cycle. This relationship can be shown theoretically using a simple sinusoid, but the relationship is more complex for speech sounds. Since *OQ* is often thought to be correlated with breathiness [2, 3, 4], by association, $H_1^* - H_2^*$ has also been used as a measure of breathiness. In perceptual studies, [5] and [6] found $H_1 - H_2$ to be moderately correlated with perceived breathiness. Other studies involving languages such as Mazatec [7], Gujarati [8, 9], and Hmong [10] have also found that $H_1 - H_2$ can be used to distinguish breathy phonations from non-breathy phonations. However, in [11], it was shown that the theoretical relationship between $H_1^* - H_2^*$ and *OQ* is not as strong as previously thought, and depends on other voice source parameters such as the asymmetry coefficient (α).

The measure $H_1^* - A_3^*$ was shown in [12] to be related to the source spectral tilt. In that study, it was suggested that source spectral tilt, as measured by $H_1^* - A_3^*$ and $H_1^* - A_1$ (difference between the first harmonic magnitude, corrected for the effects of the formants, and the spectral magnitude at the first formant frequency), may be correlated with the speed of closure of the vocal folds. It was also speculated that lower spectral tilt values should correspond to more abrupt glottal closures and higher values may be an indication of non-simultaneous closure. Fiberscopy of a small subset of speakers confirmed this to be the case, although in that study, the fiberscopic images were not collected simultaneously with the acoustic data. Other correlates of source spectral tilt include the measures $H_1^* - A_1^*$ and $H_1^* - A_2^*$ [13], and their respective uncorrected variants, $H_1 - A_1$ and $H_1 - A_2$. *CPP* is defined in [5] as "a measure of cepstral peak amplitude normalized for overall amplitude". In theory, the peaks in the cepstral domain (conventionally known as "rahmonics") reflects the properties of the source, and a well defined periodic source should have larger peaks than a less periodic one. Hence, the *CPP* measure has been used to differentiate between modal phonations (larger *CPP* value) and breathy phonations (smaller *CPP* value). Noise in the speech spectrum is usually thought to be correlated with breathiness. In [6], perceptual experiments were used to show that when random noise was added to a synthesized source signal with a large $H_1 - H_2$, English listeners were more likely to rate the signal as being breathy than if only $H_1 - H_2$ was used by itself. Estimation of the spectral noise level can be done through a harmonic-to-noise ratio (*HNR*) measure, as in [14].

The challenge in relating acoustic measures and voice quality to physiological vocal fold movements can be attributed mainly to the difficulty of obtaining direct observations of the vocal folds. In this paper, high-speed images of the vocal folds were used to extract glottal area waveforms. There waveforms were then fitted to a four-parameter source model [15], consisting of the open quotient (OQ), asymmetry coefficient (α), and

the speed of opening and closing (S_{op} and S_{cp} respectively), which represent the vocal fold closure speed. These parameters were studied together with acoustic measures using statistical analyses to determine the effects of voice quality and incomplete glottal closures. Correlations between model parameters and acoustic measures were also examined.

2. Data and Methods

2.1. Data

The data used are the same as those described in [15]. Six subjects (3 males/3 females) were asked to vary their F_0 (low, normal and high) and voice quality (pressed, normal and breathy) quasi-orthogonally while sustaining the vowel /i/. During these phonations, synchronous audio and high-speed imaging of the larynx were performed. The most stable 1 second of phonation was extracted for analysis.

2.2. Voice source model parameters and acoustic measures

As in [15], the first 150 images of each high-speed recording were manually segmented to obtain measurements of the glottal area waveforms. These 150 measurements were then averaged to obtain a single pulse which was representative of that particular phonation. Figure 1 shows this process for low F_0, breathy phonation of a female subject. The averaged glottal area waveforms were also resampled to be of length 100 samples to allow comparisons across phonations.

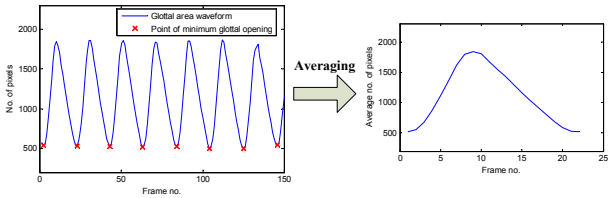

Figure 1: Example of glottal area waveform averaging. Data are from a low F_0 and breathy phonation of a female subject.

For each phonation, the normalized glottal area waveform was fitted to the source model described in [15]. This model was chosen because it was derived from the same high-speed imaging data and consisted of the four parameters: OQ, α, S_{op}, and S_{cp}. A minimum squared error (MSE) criterion was used for the waveform fitting.

Acoustic measures were calculated for each phonation and include the (supposed) open quotient correlate $H_1 - H_2$, spectral tilt measures $H_2 - H_4$, $H_1 - A_1$, $H_1 - A_2$ and $H_1 - A_3$, *CPP*, and noise measurements as calculated from the harmonic-to-noise ratio (*HNR*) measures [14] between the frequencies 0–500 Hz (*HNR05*), 0–1.5 kHz (*HNR15*) and 0–2.5 kHz (*HNR25*). These measures were calculated using the VoiceSauce software application [16] at a resolution of 1 ms. The values for each measure were then averaged across each phonation type.

Statistical analyses were performed using SPSS (v16.0). For two-way analysis of variance (ANOVA) tests, fixed factors included the subject plus one other factor from either voice quality effects or glottal gap effects. Tests where the null hypothesis had a probability of $p < 0.001$ were considered to be statistically significant. Due to the limited number of subjects, the results were not separated in terms of gender.

3. Results

3.1. Voice quality effects

Voice source model parameters and acoustic measures which were affected by the voice quality type in a statistically significant way ($p < 0.001$) are listed in Table 1. Not surprising, OQ was shown to be lowest for the pressed phonations and highest for the breathy phonations. This is in agreement with other studies ([2, 3, 4]). Analysis of individual subjects showed that, with the exception of one male subject, all subjects had the same trend for OQ: pressed $<$ normal $<$ breathy.

Table 1: Voice source model parameters and acoustic measures which were affected by voice quality in a statistically significant way ($p < 0.001$). Values shown are means and standard deviations (in parentheses) for the three voice qualities. Values in bold are the highest among the three voice qualities

	Mean (s.d.) of parameter/measure		
Parameter	Pressed	Normal	Breathy
OQ	.65(.13)	.80(.13)	**.94**(.06)
α	**.51**(.06)	.49(.09)	.39(.04)
Measure			
CPP	**25.08**(3.29)	23.99(2.40)	18.04(2.74)
HNR05	**15.41**(10.56)	13.50(7.82)	3.44(6.52)
HNR15	**24.59**(10.54)	23.44(5.76)	13.33(6.87)
HNR25	**27.41**(10.10)	26.31(5.84)	16.40(6.38)
$H_1 - A_2$	13.50(7.05)	17.27(8.59)	**23.03**(6.50)
$H_1 - A_3$	20.50(6.16)	24.35(6.72)	**29.80**(6.32)
$H_1 - H_2$	-0.22(6.79)	1.67(6.21)	**11.19**(4.58)

A somewhat unexpected result was seen for the model parameter α, although a post-hoc analysis showed that the main effect was due to the pressed/normal vs. breathy voice qualities. The results showed that, on average, the pressed and normal phonations were more symmetrical (equal durations for opening and closing phases) than the breathy phonations which were skewed towards a shorter opening phase. This is demonstrated in Figure 2 using the mean OQ and α values in Table 1 with the other model parameters (S_{op} and S_{cp}) set to 0.5. The smaller mean value for the breathy phonations was surprising because the duration of the opening phase has conventionally been thought to be always longer than the duration of the closing phase, due to the effort required to separate the vocal folds, and also because this is what has been seen in EGG and airflow signals. Individual subject analysis showed that all subjects had the lowest α values for breathy phonations.

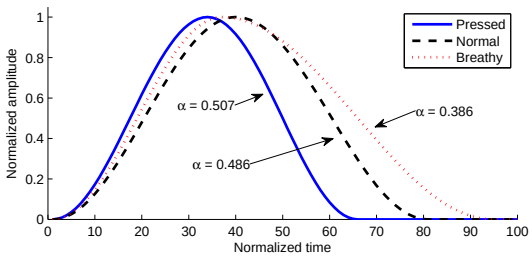

Figure 2: Examples of voice source shapes for the mean OQ and α values listed in Table 1.

Statistical analysis on the acoustic measures showed that most of the measures (*CPP*, *HNR*, and $H_1 - H_2$) thought to be related to breathiness were statistically significant. However, post-hoc analysis on these measures revealed that most of

the statistical significance was mainly from the pressed/normal vs. breathy phonations. For the *CPP* measure, the mean values were higher for the pressed and normal phonations than for the breathy phonations. This was as predicted in [5], and can be attributed to the rising of the noise floor in the speech spectrum for breathy phonations. Similarly, the *HNR* (*HNR05*, *HNR15* and *HNR25*) measures were much lower for the breathy phonations due to increased noise in the spectrum. Interestingly, the $H_1 - H_2$ measure had similar means for the pressed and normal phonations, but a significantly larger value for the breathy phonation. This is slightly different from the trends observed for the OQ parameter which had progressively increasing values from the pressed to normal to breathy voice qualities.

On average, the spectral tilt measures $H_1 - A_2$ and $H_1 - A_3$ were smallest for the pressed phonation and largest for the breathy phonation. These results confirm the hypothesis in [12] that voice sources with more abrupt glottal closures may exhibit more high frequencies in speech spectrum.

3.2. Glottal gap effects

Results in Section 3.1 showed that for the parameter α and the voice source related measures, there were few differences separating the pressed and normal phonations. However, breathy phonations had significantly different values from either the pressed or normal phonations. A possible cause for this effect could be due to the existence of incomplete glottal closures for breathy phonations. Indeed from the images, it was observed that 16 out of 17 breathy phonations had glottal gaps, while 7 out of 33 non-breathy phonations exhibited glottal gaps.

Table 2 lists the model parameters and acoustic measures which were statistically significant in ANOVA analysis, with the presence/absence of the glottal gap as the other fixed factor. Given that glottal gaps usually occurred with the breathy phonations, it was not surprising to see the OQ parameter been associated with the glottal gap effect. Similarly, it was shown previously that the α parameter had the lowest mean value for the breathy phonations, hence the statistical significance with the glottal gap factor. While it can be seen from these results that OQ is dependent on both the type of voice quality (pressed, normal or breathy) and the existence/absence of the glottal gap, it is not clear as to how or which factor is predominantly affecting α. Analysis of phonations which contained a glottal gap and were not of a breathy voice quality showed that the α values for these phonations were not necessarily the lowest for their corresponding voice quality group. However, for all subjects, the breathy phonation had the lowest α values when averaged across each subjects' F_0 type. From these results, it would be reasonable to hypothesize that it is the breathy phonations which affect the α values, but more data would be needed to confirm this.

Interestingly, the parameter S_{op}, which did not show a statistically significant effect of voice quality, showed a statistically significant effect of the glottal gap. The larger mean value for the presence of the glottal gap translates to a slower initial rise during the opening phase. A possible explanation for the slower initial rise during the opening phase could be due to the smaller distance required to reach the maximum open position of the vocal folds. Without the glottal gap, the distance from the closed position to the maximum open position is much greater, hence requiring a faster initial rise during the opening phase. Another way of interpreting these results could be that the opening phase is dictated by a constant "curve", and the glottal gap simply moves the starting point up this curve. This interpretation would also result in a lower S_{op} value when using the full range of the curve and a high value when starting near the middle of this curve.

Table 2: Voice source model parameters and acoustic measures which were statistically significant to the effects of incomplete glottal closures. Values shown are means and standard deviations (in parentheses).

	Mean (s.d.) of parameter/measure	
Parameter	Glottal gap	No glottal gap
OQ	.922(.068)	.694(.139)
α	.413(.066)	.498(.077)
S_{op}	.550(.066)	.481(.074)
Measure		
CPP	19.631(3.515)	24.605(3.258)
$H_1 - A_2$	22.070(6.812)	14.569(7.958)
$H_1 - A_3$	29.451(6.127)	21.150(6.116)
$H_1 - H_2$	9.371(5.937)	-0.014(6.283)

With the exception of the three *HNR* measures, the acoustic measures which were statistically significant to the voice quality factor were also statistically significant to the presence/absence of the glottal gap. This is not surprising given that the same measures appear to be predominantly affected by the breathy voice quality which contains most of the phonations with glottal gaps. The mean values for the measures *HNR05*, *HNR15* and *HNR25* were also lower, inferring more noise, for the presence of the glottal gap, but these were not statistically significant. This suggests that noise may be more prevalent in breathy phonations as opposed to phonations with incomplete glottal closures, which may or may not be breathy. A related study ([6]) found that, during perceptual experiments, listeners were more likely to rate a phonation as breathy if an increase in $H_1 - H_2$ was accompanied by noise; increases in $H_1 - H_2$ alone were sometimes rated as having a nasalized voice quality. Similarly, in [9], it was found that $H_1^* - H_2^*$ separated breathy vs. modal vowels for only some speakers of Gujariti.

In [12], it was suggested that speakers with high $H_1^* - A_1$ and $H_1^* - A_3^*$ values may have a posterior opening in the vocal folds. This hypothesis is supported here by the related measure, $H_1 - A_3$, which has a high mean value for the glottal gap case. Although the mean values for $H_1 - A_1$ also showed the same trend, the effect was not statistically significant.

3.3. Correlations between model parameters and acoustic measures

Table 3 lists the correlations (r) between voice source model parameters (OQ, α and S_{op}) and the acoustic measures (*CPP*, *HNR*, $H_1 - A_1$, $H_1 - A_2$, $H_1 - A_3$ and $H_1 - H_2$). Parameter S_{cp} and measure $H_2 - H_4$ did not show any strong correlations.

It can be seen that the parameter OQ is moderately correlated with the parameters α and S_{op}, and also with the measures *CPP*, $H_1 - A_1$, $H_1 - A_2$, $H_1 - A_3$ and $H_1 - H_2$. The correlations with α and S_{op} were not surprising given that α appeared to be affected by voice quality and S_{op} by the presence/absence of the glottal gap, both effects which were correlated with OQ. The negative correlation with *CPP* is most likely attributable to the breathy voice quality; since breathy phonations were seen to induce larger OQ values and also more spectral noise, hence resulting in a smaller *CPP* value. Correlations with the spectral tilt measures $H_1 - A_1$, $H_1 - A_2$ and $H_1 - A_3$ could be explained using the reasoning from [12]. That is, when the glot-

Table 3: Correlations between voice source model parameters and acoustic measures. Correlations with $r > 0.4$ are in bold and were all statistically significant.

Parameters/Measures	Voice source model parameters		
	OQ	α	S_{op}
α	**-0.5546**	–	–
S_{op}	**0.5034**	-0.3306	–
CPP	**-0.5445**	**0.5256**	-0.1617
HNR05	-0.3187	**0.4112**	-0.0985
HNR15	-0.3536	**0.4151**	-0.1814
HNR25	-0.3370	**0.4606**	-0.1521
$H_1 - A_1$	**0.4998**	-0.3053	0.2452
$H_1 - A_2$	**0.4454**	-0.3170	0.0808
$H_1 - A_3$	**0.5520**	-0.2250	0.0957
$H_1 - H_2$	**0.6563**	**-0.4730**	0.2641

tal closures become less abrupt, as in the case when OQ increases, the high frequency components are generally reduced. The moderate correlation with $H_1 - H_2$ was as predicted by [1], although the correlation here was not quite as strong as that study ($r = 0.656$ vs. $r = 0.693$). However, as shown by the mean values in Table 1, the mean $H_1 - H_2$ values did not increase linearly for the three types of voice qualities as occurred with the parameter OQ. Furthermore, $H_1 - H_2$ also showed a slight correlation with the asymmetry coefficient, α. This is similar to the findings in [11], which used the LF model to theoretically show that $H_1^* - H_2^*$ was dependent on both OQ and the asymmetry coefficient.

Apart from the measure $H_1 - H_2$, α was also correlated with *CPP*, and the three *HNR* measures. *CPP* was also moderately correlated with the parameter OQ which was affected by the voice quality. Interestingly, the *HNR* measures were more strongly correlated with α than OQ, although the correlations are moderately weak for both parameters. This is not surprising since it was shown previously that both α and the *HNR* measures were thought to be predominantly affected by the breathy voice quality.

The lack of any meaningful correlations with the parameter S_{cp} is surprising given that S_{op} is moderately correlated with OQ; the correlation coefficient between OQ and S_{cp} is $r = 0.1825$ compared with $r = 0.5034$ for S_{op}. Since the tension of the laryngeal muscles is assumed to be constant during a cycle of phonation, this result requires further exploration.

3.4. The effects of normalization

It is conceivable that the normalization used in producing the glottal area waveforms may have removed some important information. This is particularly true for the glottal area peaks (representing the maximum opening during each cycle) and the DC offsets (representing the level of incomplete glottal closures). Preliminary analysis of data recorded from one female subject saying a glide from a breathy voice quality to a pressed one showed that model parameters were not significantly affected by the changing peak and DC offset values. Results are shown in Figure 3.

4. Summary and Conclusion

In this paper, direct measurements of the glottal area waveforms were used to examine the voice source model parameters and acoustic measures in relation to the effects of voice quality and glottal gaps. Using ANOVA tests, it was found that the model parameter OQ and the spectral tilt measures $H_1 - A_2$, $H_1 - H_2$ and $H_1 - A_3$ were affected by both voice quality and glottal gaps, while the parameter α was predominantly affected by voice quality, especially of the breathy type. This was also the case with many of the acoustic measures, such as *CPP* and the three *HNR* measures, indicating the presence of more spectral noise for breathy phonations. Interestingly, the source parameter S_{op} was seen to be affected by the presence of glottal gaps, while no such effect was observed for the parameter S_{cp}. Correlation analysis showed that the measure $H_1 - H_2$ was correlated with both the parameters OQ and α, which agrees with existing theoretical studies. However, the correlation between OQ and S_{op} and the lack of correlation between OQ and S_{cp} is puzzling and requires further research.

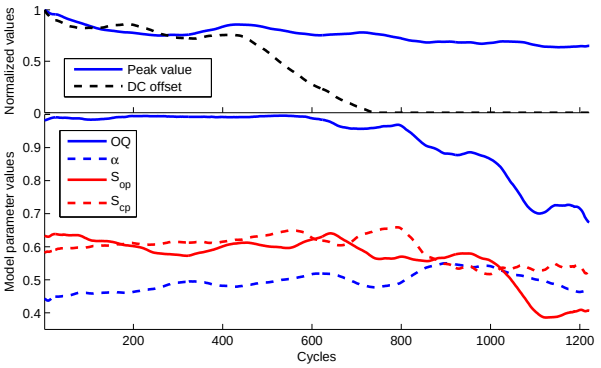

Figure 3: Model parameters and normalized peak and DC offset values for a glide phonation (breathy to pressed) for a female.

5. Acknowledgements

This work was supported in part by the NSF, and by grant DC01797 from the NIH/NIDCD. We thank Jody Kreiman for many productive discussions. High speed imaging was done by Drs. Jody Kreiman and Bruce Gerratt.

6. References

[1] E. Holmberg, R. Hillman, J. Perkell, P. Guiod, and S. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *JSHR*, vol. 38, pp. 1212–1223, 1995.

[2] M. Huffman, "Measures of phonation type in Hmong," *JASA*, vol. 81, no. 2, pp. 495–504, February 1987.

[3] E. Fischer-Jorgensen, "Phonetic analysis of breathy (murmured) vowels in Gujarati," *Indian Linguist*, vol. 28, pp. 71–139, 1967.

[4] M. Södersten and P.-A. Lindestad, "Glottal closure and perceived breathiness during phonation in normally speaking subjects," *JSHR*, vol. 33, pp. 601–611, 1990.

[5] J. Hillenbrand, R. Cleveland, and R. Erickson, "Acoustic correlates of breathy vocal quality," *JSHR*, vol. 37, pp. 769–778, 1994.

[6] D. Klatt and L. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *JASA*, vol. 87, no. 2, pp. 820–857, February 1990.

[7] B. Blankenship, "The time course of breathiness and laryngealization in vowels," Ph.D. dissertation, UCLA, 1997.

[8] C. Esposito, "The effects of linguistic experience on the perception of phonation," Ph.D. dissertation, UCLA, 2006.

[9] S. Khan, "An acoustic and electroglottographic study of breathy phonation in gujarati," in *JASA*, San Antonio, TX, 2009, p. 2222.

[10] C. Esposito, J. Ptacek, and S. Yang, "An acoustic and electroglottographic study of white hmong phonation," in *JASA*, San Antonio, TX, 2009, p. 2223.

[11] N. Henrich, C. d'Alessandro, and B. Doval, "Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data," in *Proc. EUROSPEECH*, Aalborg, Denmark, 2001, pp. 47–50.

[12] H. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *JASA*, vol. 101, pp. 466–481, 1997.

[13] J. Kreiman, B. R. Gerrat, M. Iseli, J. Neubauer, Y.-L. Shue, and A. Alwan, "The relationship between open quotient and $H_1^* - H_2^*$," in *Proc. 6th ICVPB*, Tampere, Finland, Aug. 2008.

[14] G. de Krom, "A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals," *JSHR*, vol. 36, pp. 254–266, 1993.

[15] Y.-L. Shue and A. Alwan, "A new voice source model based on high-speed imaging and its application to voice source estimation," in *ICASSP*, Dallas, TX, March 2010, pp. 5134–5137.

[16] Y.-L. Shue, "VoiceSauce: a program for voice analysis," 2010, http://www.ee.ucla.edu/~spapl/voicesauce/.