Acoustic Correlates of Glottal Gaps

[1]Gang Chen, [2]Jody Kreiman, [3]Yen-Liang Shue, and [1]Abeer Alwan

[1]Department of Electrical Engineering, University of California, Los Angeles
[2]Division of Head and Neck Surgery, UCLA School of Medicine, Los Angeles
[3]Dolby Australia

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

Abstract

During speech production, the vocal folds may not close completely. The resulting glottal gap (GG) or incomplete glottal closure has not been systematically studied in terms of GG acoustic and/or perceptual consequences. This paper uses high-speed imaging to investigate the relationship between GG area, source parameters, acoustic measures, and voice quality for 6 subjects. Results showed that the cepstral peak prominence (*CPP*) and the harmonics-to-noise ratio (*HNR*) are affected by GG area, indicating the presence of more spectral noise with increasing GG area. Analysis of a glide phonation from breathy to pressed for one female speaker showed that measures $H_1^* - H_2^*$ and $H_1^* - A_3^*$ were positively correlated with GG area under a steady fundamental frequency ($F0$). In some phonatory modes, increasing $F0$ may reduce the amplitude of vocal folds vibration, increase GG area, and produce a lower spectral tilt due to significant aspiration noise, leading to a negative correlation between GG area and the spectral tilt measure $H_1^* - A_3^*$.

Index Terms: glottal gap, acoustic measure, voice source

1. Introduction

During speech production, the voice source provides excitation to the vocal tract and contains important lexical information. In voiced sound production, the vocal folds oscillate to generate quasi-periodic excitations. With some phonations, the vocal folds may not close completely during the closing phase, producing a glottal gap (GG) or incomplete glottal closure.

The GG phenomenon has been found in normal speakers as well as in dysphonic speakers. GGs commonly exist in pathologic states associated with impaired vocal fold mobility and altered glottic closure, such as vocal fold paralysis, vocal fold atrophy, and sulcus vocalis [1]. Among normal speakers, direct examination of the vocal folds during normal phonation showed that the presence of a posterior glottal opening is common among female speakers, but happens less frequently for male speakers [2].

Although previous fiberscopic studies showed the prevalence of GGs, GG acoustic and perceptual consequences have not been systematically investigated. It has been shown [2, 3, 4] that larger GGs result in more aspiration noise, which is usually perceived as increased breathiness. It was hypothesized in [5] that speakers with larger posterior glottal openings would have larger spectral tilt (measured by $H_1^* - A_1$, the difference between the first harmonic magnitude, corrected for the effects of the formants, and the spectral magnitude at the first formant frequency and by $H_1^* - A_3^*$, the difference between the first harmonic magnitude and the spectral magnitude at the third formant frequency, corrected for the effects of the formants). The fiberscopic observation in that study confirmed that subjects with larger spectral tilt measures did exhibit larger posterior openings. Note that in [5], preliminary fiberscopic recordings were made for 4 subjects after the audio recordings. Thus, simultaneous audio/video recordings are needed to verify the hypothesis, as stated by the author.

Many studies investigated the relationship between harmonic amplitudes and glottal configuration. Varying levels of correlation between $H_1^* - H_2^*$ (difference between the first two spectral harmonic magnitudes, corrected for the effects of the formants), open quotient (OQ, defined as the proportion of time the vocal folds are open during a phonation cycle), and asymmetry quotient (AQ, the proportion of the opening phase relative to the open phase) were reported. In [6], it was shown that $H_1^* - H_2^*$ is minimally affected by AQ when OQ is small, but the influence of AQ increases as OQ increases. In this paper, we will show that $H_1^* - H_2^*$ could be predicted well by the relative GG area, when GG exists, and asymmetry quotient, suggesting that the relative GG area could be a new degree of freedom in predicting acoustic measures.

In [1], GG area was measured at the most closed point of vibration from digitized videostroboscopic images of speakers with different speech pathologies. Results showed that GG area affected pitch perturbation, harmonics-to-noise ratio, high-frequency power ratio, mean flow rate, and maximum phonation time. It was also reported that acoustic and aerodynamic measures were similar when GG sizes were similar, regardless of the dysphonia groups, suggesting that vocal function was predominantly influenced by GG size, not by the cause of glottic incompetence. In [2], glottal closure and perceived breathiness were evaluated in 9 female and 9 male subjects with no known speech pathology. Video-fiberstroboscopic recordings and audio recordings were judged by speech clinicians to evaluate the degree of glottal closure and the degree of perceived breathiness, respectively. Results showed that the degree of incomplete closure and the degree of perceived breathiness were significantly higher for females than for males; the degree of incomplete closure was not significantly affected by pitch levels.

In [7], two types of GGs were studied by computer simulation. One type of GG extended to the membranous glottis during a “linked leak”. The second type occurred when the GG was an orifice in the cartilaginous portion which was separated from the vibrating part of the glottis during a “parallel chink”. Results showed that the spectral slope is the major difference between the two types of GGs. The spectral slope of a parallel chink case is relatively close to or even slightly flatter than that of a no-gap case. During a linked leak, the spectral slope becomes much steeper. In [8], glottal area waveforms were extracted from high-speed images of the larynx. The effects of

GGs on voice source model parameters and acoustic measures were examined. Results showed that OQ, AQ, *CPP*, and spectral tilt measures ($H_1^* - A_2^*$, $H_1^* - A_3^*$, and $H_1^* - H_2^*$) were significantly affected by the presence/absence of a GG. Phonations with GGs had significantly higher $H_1^* - A_2^*$ and $H_1^* - A_3^*$ values than those without GGs, supporting the hypothesis in [5]. Note that in [8], only the effect of the presence/absence of a GG was analyzed, without quantitative measures of GG size. In this paper, GG size was quantitatively measured by the proposed relative GG area (*RGGA*).

The difficulty of relating acoustic measures to GGs can be attributed mainly to the difficulty of obtaining quantitative measures of GG area. In this paper, glottal area waveforms were extracted from high-speed images of the vocal folds. GG area, OQ, and AQ were measured from each glottal cycle. These parameters were studied together with acoustic measures using statistical analyses to determine the effects of GGs.

2. Data and Methods

2.1. Data

The data used are the same as those described in [8]. Synchronous audio and high-speed video recordings of the vocal folds were collected from six subjects, three females (denoted by F1-3) and three males (denoted by M1-3). Speakers varied their $F0$ (low, normal, and high) and voice quality (pressed, normal, and breathy) quasi-orthogonally. Speakers were asked to sustain the vowel /i/, and the most stable 1 second of phonation was extracted. In this paper, only phonations which exhibited GGs were selected for analyses. These phonations include: 16 out of 17 breathy phonations, 7 out of 33 non-breathy phonations, 8 out of 18 high-pitched phonations, 15 out of 32 normal and low-pitched phonations, 14 out of 26 phonations from female speakers, and 9 out of 24 phonations from male speakers.

2.2. Glottal gap area, OQ, and AQ calculations

As in [8], the first 150 images of each high-speed recording were manually segmented to obtain measurements of the glottal area waveforms. Due to the different positioning of the laryngoscope, GG area needs to be normalized. The relative GG area (*RGGA*) is proposed in this paper, and is defined as GG_{min}/GG_{max}, where GG_{min} is the minimum glottal area and GG_{max} is the maximum glottal area of each glottal cycle. *RGGA* is a relative measure in the sense that it measures GG size relative to the AC component. A similar vocal efficiency index measure in the form of AC/DC ratio of volume velocity is used in clinical research to reflect the efficiency of phonation [9]; note that it is suggested in [7] that the DC airflow alone is not a very good measure. *RGGA* was measured for each individual glottal cycle to represent GG size. OQ was calculated as the time from the first opening instant to the onset of maximum closure (or minimum area), divided by the cycle duration. AQ was calculated from the same glottal area waveform data by locating the first instants of glottal opening, the instants of maximum opening, and the onsets of maximum closure. AQ is defined as $t_o/(t_o + t_c)$, where t_o is the duration of opening phase and t_c is the duration of closing phase [6].

Note that in [8], the glottal area waveform of each individual phonation was averaged to obtain a single pulse which was representative of that particular phonation. In this study, each individual cycle of glottal area waveform generated one set of *RGGA* and AQ measures without averaging, allowing sufficient data for analyses on individual speakers.

2.3. Acoustic measurements

Acoustic measures were calculated for each phonation and include $H_1^* - H_2^*$ and $H_1^* - A_3^*$ (related to the spectral tilt), *CPP* (related to the periodic structure of the source [10]), and *HNR* between the frequencies 0–3.5 kHz (measuring the spectral noise level [11]).

These measures were calculated using VoiceSauce software [12] at a resolution of 1 ms. In each glottal cycle, acoustic measures were averaged to match *RGGA* and AQ of that cycle. Statistical analyses were performed using SPSS (v16.0). Tests where the null hypothesis had a probability of $p < 0.001$ were considered to be statistically significant.

3. Results

3.1. Experiment 1

In Experiment 1, *RGGA* was measured for all phonations with GGs.

3.1.1. Voice quality and pitch effects

Table 1 lists means and standard deviations of *RGGA* for the three voice quality types and the three pitch levels. Statistical analysis showed that breathy phonations had significantly larger *RGGA* than normal and pressed ones. There was no significant difference in *RGGA* between normal and pressed phonations. High-pitched cases had significantly larger *RGGA* than normal and low-pitched cases. No significant difference was found in *RGGA* between normal and low-pitched cases.

Table 1: *Means and standard deviations (in parentheses) of RGGA for the three voice quality phonations and the three pitch levels*

Voice quality	breathy	normal	pressed
RGGA	0.244 (0.135)	0.108 (0.077)	0.061 (0.005)

Pitch level	low	normal	high
RGGA	0.121 (0.082)	0.139 (0.077)	0.348 (0.189)

3.1.2. Correlation analysis

Table 2 lists the correlations between *RGGA* and various acoustic measures for the whole dataset. Correlation results for individual speakers are shown in Table 3.

In the presence of GGs, variation of OQ is small ($mean = 0.955$, $sd = 0.071$), suggesting that OQ may not be a good predictor of acoustic measures in the presence of GGs; hence, it is not surprising to see that OQ did not show a significant correlation with *RGGA*. AQ did not show a significant correlation with *RGGA* for the whole dataset, but showed a significant negative correlation in the individual analyses for all speakers except M3, suggesting that the way and/or degree of varying AQ could be speaker dependent. The earlier study [8] on the same data showed a significant negative correlation between AQ and OQ ($r = -0.5546, p < 0.001$). Similar to OQ, the increase in GG size characterizes a more "open" glottic configuration and associates with the trend of decreasing AQ.

The noise measures *CPP* and *HNR* were negatively correlated with *RGGA* for the whole dataset, which was confirmed in individual analyses. Increased GG size allows more airflow, producing more aspiration noise. This result is consistent with earlier studies [3, 4], which suggested that an increase in GG size results in an increase in the noise floor of the speech spectrum.

Table 2: *Correlation coefficients between RGGA and various acoustic measures. All values are significant.*

Measures	$F0$	*CPP*	*HNR*	$H_1 - A_3$
Correlation with *RGGA*	.491	-.646	-.330	-.522

Interestingly, $F0$ showed a significant positive correlation with *RGGA* on the whole dataset; moderate correlations were also observed for all 3 female speakers in the individual analyses. One explanation could be the increased stiffness of the vocal folds when increasing $F0$. A common approach to increase $F0$ is to increase activity in the cricothyroid muscle to stiffen the vocal folds. The increased tension could prevent the vocal folds from complete closure at high pitch [13]. It has been reported [14] that some individual elderly speakers tended to vary glottic configuration consistently with pitch level changes. In [2], the tendency of increasing degree of GG from habitual to high pitch for several female subjects was reported, but the effect of pitch was not significant for female subjects; no effect of pitch on degree of closure was found for male subjects. In that study, the lack of a significant effect of pitch may be attributed to the fact that the highest $F0$ was 262 Hz, while in our data $F0$ was as high as 330 Hz. In our study, a correlation analysis on female subjects showed that *RGGA* is modestly correlated with $F0$ ($r = 0.356, p < 0.001$). Male subjects did not show a significant effect of $F0$ on *RGGA*. For the largest GGs, the GG extended to the membranous portion, resulting in an anterior-posterior gap configuration. For each speaker, the largest GG size is from the breathy phonation at a high pitch. The increasing longitudinal tension upon the vocal ligaments would reduce the vibration amplitude, thus producing a larger GG. This statement is supported by similar findings among elderly [14] and female speakers [2], since those speakers are generally assumed to be more breathy.

Surprisingly, moderate negative correlations between the spectral tilt measure $H_1^* - A_3^*$ and *RGGA* were observed for the whole dataset; strong negative correlations were also observed for speakers F1, F2, F3, and M1 in the individual analyses. Thus, regression analyses were used to relate $F0$, *CPP*, and $H_1^* - A_3^*$ to *RGGA*. Results are shown in Table 4. The multiple linear regression models capture the acoustic consequences of *RGGA* well, in terms of R^2 values. Speaker F1, F2, F3, and M1 showed strong negative effects of *RGGA* on $H_1^* - A_3^*$. This is somewhat inconsistent with the hypothesis in [5] that larger GG would result in larger spectral tilt measures. This could be explained by the contribution of the breathy phonation at a high pitch. During this falsetto-like phonation, the small-amplitude vibration of the vocal folds is associated with strong aspiration noise caused by the ultra-large GG which extends to the membranous portion of the glottis. The relatively weak harmonic structure is overwhelmed by the high noise floor in the spectrum, leading to a small spectral tilt. Thus a large GG produces a small $H_1^* - A_3^*$ value under this mechanism.

3.2. Experiment 2: A glide phonation

Speaker F1 from Experiment 1 (a female phonetician) participated in this experiment. She gradually changed the phonation type from “breathy” to “pressed” while holding pitch and vowel quality steady. The duration of this utterance is 8 seconds. High-speed image and audio signals were recorded synchronously and analyzed. Glottal area waveform for the whole utterance was calculated from the high-speed images. *RGGA*, OQ, and AQ were calculated from the glottal area waveform using the same method described in Experiment 1. Since our focus is the role of GG size, measures from the portion of utterance with a GG were used for following correlation and regression analyses. Direct observation showed that the GG situated in the membranous portion of the glottis and kept decreasing in size during the phonation.

Table 3: *Correlation coefficients relating RGGA to source parameters and acoustic measures for each speaker. “ns” denotes not significant. All other values are significant. Measures which correlate with RGGA significantly for 4 or more speakers are shown.*

Speaker	F1	F2	F3	M1	M2	M3
AQ	-.551	-.681	-.933	-.903	-.701	ns
$F0$	.725	.807	.566	ns	ns	.697
CPP	-.432	-.770	-.648	-.947	-.974	-.631
HNR	ns	-.608	-.947	-.827	-.744	ns
$H_1^* - A_3^*$	-.858	-.969	-.969	-.958	ns	ns

Table 4: *Standardized regression coefficients and* R^2 *values for multiple linear regression analyses relating RGGA to* $F0$, *CPP, and* $H_1^* - A_3^*$ *for each speaker. “ns” denotes not significant. All other values are significant.*

Speaker	F1	F2	F3	M1	M2	M3
$F0$	ns	.296	.164	ns	-.262	.818
CPP	-.276	-.135	ns	ns	-.929	-.670
$H_1^* - A_3^*$	-.944	-.665	-.869	-.886	ns	ns
R^2	.828	.962	.956	.972	.975	. 949

3.2.1. Correlation analysis

Table 5 lists correlation coefficients relating *RGGA* to source parameters and acoustic measures. As expected, *RGGA* correlated with AQ and the noise measures (*CPP* and *HNR*) negatively, confirming the results in Experiment 1. $H_1^* - H_2^*$ showed a strong correlation with GG size. It was reported in [4] that listeners were more likely to rate a phonation as breathy if an increase in $H_1^* - H_2^*$ was accompanied by noise in perceptual experiments. In this glide phonation, the transition from breathy to modal is characterized by a gradual decrease of GG size, associated with decreasing noise measures and $H_1^* - H_2^*$. The close correlation with both noise measures and $H_1^* - H_2^*$ suggests that GG size could be a physiological indicator of breathiness. The measures $H_1^* - H_2^*$ and *RGGA* for the glide phonation are shown in Figure 1. Both measures displayed similar falling trend, while the ripples matched quite closely.

The spectral tilt measure $H_1^* - A_3^*$ showed a strong positive correlation with GG size. This is consistent with the hypothesis in [5]. Note that $F0$ is within habitual range and kept steady, allowing the result to be different from Experiment 1, where varying $F0$ level was affecting the physiological mechanism predominantly.

Table 5: *Correlation coefficients relating RGGA to source parameters and acoustic measures. All values are significant.*

Parameters/ Measures	AQ	$H_1^* - H_2^*$	*CPP*	*HNR*	$H_1^* - A_3^*$
Correlation with *RGGA*	-.821	.904	-.788	-.592	.801

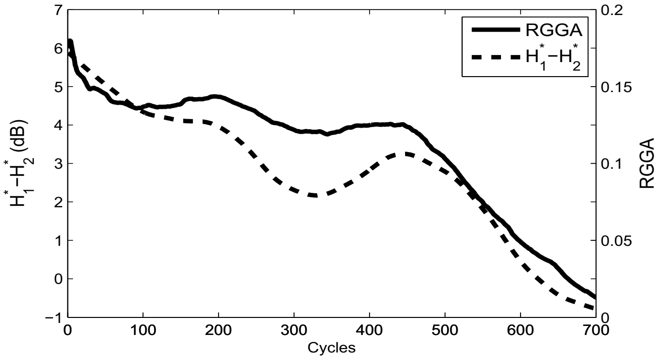

Figure 1: $H_1^* - H_2^*$ *and RGGA for a glide phonation (breathy to pressed)*

3.2.2. Relationship between $H_1^* - H_2^*$ and source parameters

A high correlation between *RGGA* and $H_1^* - H_2^*$ ($r = 0.904$) was observed. Regression analyses relating *RGGA*, *AQ*, and OQ to $H_1^* - H_2^*$ were conducted. Standardized regression coefficients and the proportion of explained variance (R^2) are listed in Table 6. $H_1^* - H_2^*$ could be well predicted using only *RGGA*, with an R^2 of 0.817. $H_1^* - H_2^*$ could also be well predicted by AQ with an R^2 of 0.765, while the model is further improved by combining *RGGA* and *AQ*. The regression model using only OQ to predict $H_1^* - H_2^*$ had a low R^2 of 0.107. In the presence of GG, *RGGA* and AQ are better predictors of $H_1^* - H_2^*$ than OQ, suggesting the emergence of a new degree of freedom when GG persists.

Table 6: *Standardized regression coefficients and* R^2 *relating RGGA, AQ, and OQ to* $H_1^* - H_2^*$.

RGGA	AQ	OQ	R^2
–	–	.329	.107
.904	–	–	.817
–	-.875	–	.765
.569	-.408	–	.871

4. Discussion and Conclusion

From the individual analyses in Experiment 1 and the glide phonation in Experiment 2, AQ showed a significant negative correlation with *RGGA*. An early study on the same data [8] showed a similar trend between AQ and OQ: AQ is negatively correlated with OQ ($r = -0.5546$, $p < 0.001$). Similar to OQ, the increasing GG size characterizes a more "open" glottic configuration, associating with the trend of decreasing AQ. The regression model in Experiment 2 showed that *RGGA* and AQ are better predictors of $H_1^* - H_2^*$ than OQ in the presence of GG. When the correlation between OQ and $H_1^* - H_2^*$ decreases, $RGGA$ emerges as a new degree of freedom in predicting acoustic measures. Note that we also compared the relationship between AQ, OQ, and $H_1^* - H_2^*$ in voices with high OQ but no GGs and did not find, unlike [6], similar results.

The existence of GGs among female speakers acts as an additional degree of freedom, resulting in more variation in voicing acoustics. As stated in [15]:"the size of a posterior glottal opening can be considered to provide an additional degree of variability in the acoustic parameters considered."

The noise measures (*CPP* and *HNR*) are negatively correlated with *RGGA*, indicating the presence of more spectral noise with increasing GG area. $F0$ showed a positive correlation with GG size, especially among female speakers.

The positive correlation between GG size and spectral tilt measure $H_1^* - A_3^*$ in Experiment 2 supported the hypothesis in [5]; while a negative correlation was found in Experiment 1 under varying $F0$ level. This suggests that Hanson's hypothesis that larger GG would result in larger spectral tilt measures is only valid under a steady $F0$ constraint. In some phonatory modes, increasing $F0$ may reduce the amplitude of vocal folds vibration, increase GG size, and produce a lower spectral tilt due to significant aspiration noise, leading to a negative correlation between *RGGA* and the spectral tilt.

The acoustic consequences of GGs have been shown to be quite complicated and more data are needed to generalize these results.

5. Acknowledgements

This work was supported in part by NSF grant IIS-1018863 and NIH/NIDCD DC01797.

6. References

[1] K. Omori, D.H.Slavit, A.Kacker, and S. Blaugrund, "Influence os size and etiology of glottal gap in glottic incompetence dysphonia," *The Laryngoscope*, vol. 108, pp. 514–518, 1998.

[2] M. Södersten and P.-A. Lindestad, "Glottal closure and perceived breathiness during phonation in normally speaking subjects," *Journal of Speech and Hearing Research*, vol. 33, pp. 601–611, 1990.

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

[4] D. Klatt and L. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *Journal of the Acoustical Society of America*, vol. 87, pp. 820–857, 1990.

[5] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *Journal of the Acoustical Society of America*, vol. 101, pp. 466–481, 1997.

[6] N. Henrich, C. D'Alessandro, and B.Doval, "Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data," in *Eurospeech*, 2001, pp. 47–50.

[7] B. Cranen and J. Schroeter, "Modeling a leaky glottis," *Journal of Phonetics*, vol. 23, pp. 165–177, 1995.

[8] Y.-L. Shue, G. Chen, and A. Alwan, "On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures," in *Interspeech*, 2010, pp. 34–37.

[9] N. Isshiki, *Assessment of Vocal Function. In: Phonosurgery: Theory and Practice.* Tokyo: Springer-Verlag, 1989.

[10] J. Hillenbrand, R. Cleveland, and R. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech and Hearing Research*, vol. 37, pp. 769–778, 1994.

[11] G. de Krom, "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," *Journal of Speech and Hearing Research*, vol. 36, pp. 254–266, 1993.

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

[13] M. Hirano, K. Kiyokawa, and S. Kurita, *Laryngeal muscles and glottic shaping. In O. Fujimura (Ed.), Vocal fold physiology: Vol. 2. Voice production. Mechanisms and functions*, 2nd ed. New York: Raven Press, 1988.

[14] S. Linville, "Glottal gap configurations in two age groups of women," *Journal of Speech and Hearing Research*, vol. 35, pp. 1209–1215, 1992.

[15] H. M. Hanson and E. S. Chuang, "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," *Journal of the Acoustical Society of America*, vol. 106, pp. 1064–1077, 1999.