Investigating the effect of F0 and vocal intensity on harmonic magnitudes: Data from high-speed laryngeal videoendoscopy

Gang Chen[1], *Soo Jin Park*[1], *Jody Kreiman*[2], *Abeer Alwan*[1]

[1]Department of Electrical Engineering, University of California, Los Angeles, USA
[2]Department of Head and Neck Surgery, School of Medicine, University of California, Los Angeles, USA
{gangchen, sj.park, jkreiman}@ucla.edu, alwan@seas.ucla.edu

Abstract

The relative magnitude of the first two harmonics of the voice source (H1*-H2*) is an important measure and is assumed to be one exponent of changes in vocal quality along a breathy-to-pressed continuum. H1*-H2* is often associated with glottal open quotient (OQ) and glottal pulse skewness (as quantified by speed quotient, SQ), but may also covary with fundamental frequency (F0) and vocal intensity. We examined the relationship between H1*-H2*, F0, and vocal intensity using phonations in which vocal qualities varied continuously in F0 and intensity. Glottal area measures (OQ and SQ) and acoustic measures (F0, intensity, and H1*-H2*) were studied using simultaneously-collected laryngeal high-speed videoendoscopy and audio recordings from 9 subjects. Analyses of individual speakers showed that H1*-H2* may sometimes vary as a function of F0 alone, with OQ and SQ remaining rather constant, hypothetically when nonlinear source-filter interaction is strong. Although conventionally H1*-H2* is assumed to decrease with increasing vocal intensity due to a decrease in OQ, results showed examples where H1*-H2* increased with increasing vocal intensity, hypothetically when the effect of decreasing pulse skewness exceeds the effect of decreasing OQ. In some phonatory modes, the relationship between SQ and H1*-H2* may not be as monotonic as previously assumed.

Index Terms: voice source, voice quality, harmonic magnitudes, laryngeal high-speed videoendoscopy, glottal area waveform, open quotient, speed quotient

1. Introduction

The voice source represents the excitation signal to the speech production system. A better understanding of possible dependencies between voice quality and source properties such as the source spectrum, fundamental frequency (F0), and intensity may lead to more natural voice qualities in speech synthesis systems, as well as more valid and reliable linguistic analyses and modeling. However, uncovering the links between these factors and quality has been challenging. It is known that changing vocal intensity has pronounced effects on the speech spectrum, which are often quantified via the relative magnitudes of the first two harmonics of the source spectrum, denoted by H1*-H2* and measured from audio signals with correction for the effects of vocal tract resonances [1–3]. Intensity-related spectral changes can be traced (at least in part) to adjustments in vocal fold stiffness in response to changing subglottal pressure, which in turn lead to differences in the glottal waveform shape. Glottal flow and area waveforms have been used to study these factors, and the mathematical relationship between them can be modeled in the non-linear source-filter interaction theory [4]. This interaction has the effect of "delaying" the peak of the glottal flow pulse, making it more rightward-skewed than the glottal area pulse [5, 6]. Dynamically-changing area and flow are quantified mainly in terms of the open quotient (OQ), which equals the relative amount of time the glottis is open within a glottal vibratory cycle [7]. Assuming all other influences are constant, an increase in OQ is assumed to produce an increase in the intensity of the first harmonic (and thus in H1*-H2*). An increase in H1*-H2* is in turn often assumed to indicate a change in vocal quality from "pressed" to "breathy" [8].

However, recent evidence is inconsistent with the strong version of this account. Several studies suggest that OQ is not the only determinant of H1*-H2*, and that other factors such as glottal pulse skewness [9, 10], glottal gap [11, 12], and F0 [10] also contribute. Analysis of laryngeal high-speed videoendoscopy (HSV) data in [11] showed that F0 and OQ both contributed significantly to statistical prediction of H1*-H2* (implying independence), but the exact relationship seemed to be speaker-dependent. In contrast, [2] found that across subjects H1*-H2* was strongly and positively correlated with F0 for F0 below 175 Hz (r=0.77), while a negative correlation with F0 was found above that frequency (r=-0.47). Additional studies used HSV, electroglottographic (EGG) and/or inverse filtering analyses to investigate the relationship between F0 and OQ, again with variable results. For example, no relationship between OQ and F0 was found in [13–15], while increases in OQ with increasing F0 were reported in analyses of glottal flow [16], in studies using HSV and photoglottography or EGG within a modal speech register [17, 18], or during a glissando from modal to falsetto register [19].

In a study of inverse-filtered speech waveforms [10], F0 and H1*-H2* were negatively correlated for 6 cases, positively correlated for 18 cases, and not correlated for 4 cases. The authors argued that variables such as intensity may correlate even more strongly with H1*-H2* than does F0, leading to interspeaker variability. For example, [20] showed that H1-H2 increased by about 6 dB when the overall intensity was lowered by 10 dB below normal levels. In [7], 18 out of 20 subjects showed relatively strong (r>0.70) negative relationships between H1-H2 and sound pressure level. This relationship was further assumed to occur because OQ tended to decrease with increasing vocal intensity. Studies in [17, 21, 22] reported that a louder voice tended to result in a smaller OQ. In [23], EGG data showed that OQ tended to be negatively correlated with vocal intensity, at least for some glottal configurations.

In summary, various studies suggest that F0, vocal intensity, and perceptually-important voice source characteristics such as H1*-H2* might be inter-related, but systematic experimental

This work was supported in part by NSF Grant No. IIS-1018863 and by NIH/NIDCD Grant Nos. DC01797 and DC011300.

validation is rather limited and results are variable. This study used HSV to investigate variations in glottal waveform shape with changing F0 and vocal intensity, along with the effects of these changes on the source spectrum. H1*-H2*, F0, and intensity were measured from recorded acoustic signals. OQ and speed quotient (SQ, the length of the closing phase relative to the opening phase [9]) were measured synchronously from HSV. By gathering multiple tokens from male and female speakers who varied F0 or vocal intensity continuously within an utterance, we hoped to assess the effect of each variable (and their interactions) on H1*-H2*. We hypothesized that the relationship between H1*-H2*, F0, and intensity is speaker dependent, and that these dependencies are mediated at least partly by the degree of source-filter interaction [4].

2. DATA AND METHODS

2.1. High-speed videoendoscopy and audio recording

Synchronous audio recordings and HSV of the vocal folds were collected from 9 subjects (4 male, 5 female; range 18-43 years; mean age=26.8 years; SD=9.0 years). The data collection procedures are similar to those in [24–28]. Briefly, in the first experiment speakers produced the vowel /i/ while gradually increasing F0 and holding intensity and vowel quality as constant as possible. In the second experiment, speakers gradually increased intensity while holding F0 and vowel quality as constant as possible. The vowel /i/ was selected to optimize the view of the vocal folds [29]; across tokens vowel quality ranged from /I/ to approximately cardinal vowel /ɛ/. HSV of the vocal folds was recorded using a Phantom V210 camera at a sampling rate of 10,000 frames/second, with a resolution of 208×352 pixels. Audio signals were synchronously recorded with a Brüel & Kjær microphone (type 4193-L-004) and directly digitized at a sampling rate of 60 kHz (later downsampled to 16 kHz for analysis). Synchronized audio and HSV were recorded for 6 s. All speakers performed both tasks, except M3 and M4, who only participated in the first or second task, respectively.

2.1.1. *Measures from high-speed imaging*

Glottal area waveforms of the complete utterances were extracted using "GlotAnTools," a software toolkit that automatically segments the glottal area from HSV [30]. Segmented glottal areas were visually examined to ensure accuracy. Following [11, 24, 31], each cycle of glottal vibration was tracked from the extracted glottal area waveforms by marking the first instants of glottal opening when glottal closure was complete. When no complete glottal closure occurred, the moments of minimal glottal area were tracked. These cycle boundaries were detected using a customized automatic algorithm similar to the syllable detection method in [32]. For each individual glottal cycle, OQ was calculated as the time from the first opening instant to the onset of maximum closure (or minimum area), divided by the duration of the current glottal cycle. Glottal pulse skewness was measured using the speed quotient (SQ; [9]), calculated as the duration of the closing phase relative to the opening phase. OQ and SQ values were smoothed over 100 ms windows.

2.1.2. *Acoustic measures*

H1*-H2* was measured pitch-synchronously from the audio signals with VoiceSauce software [33] using an analysis window of six periods with a 1 ms shift. The harmonic magnitudes, H1* and H2*, were calculated from the speech spectrum and corrected for the effects of the first two formant frequencies using the formula in [2]. F0 values were obtained from the STRAIGHT algorithm [34]. Formant frequencies were estimated using Snack Sound Toolkit software [35]. Results were verified by visual inspection. Because intensity was only analyzed within each recording, absolute sound pressure levels were not measured. Instead, vocal intensity was measured from the audio signals by calculating the RMS energy with a 50 ms window. For subsequent analysis these acoustic measures were aligned with measures from HSV.

3. Results

3.1. Experiment 1: The effect of increasing F0

Table 1 shows the correlations among F0 and H1*-H2*, OQ, and SQ for each speaker, and Figure 1 shows H1*-H2*, OQ, and SQ as a function of F0 for each speaker.[1] Both positive and negative correlations of similar magnitudes between F0 and H1*-H2* were observed, consistent with results reported in [10]. Also consistent with previous studies, the relationship between F0 and OQ was variable as well, with positive correlations for 4 speakers, negative correlations for 2 speakers, and no significant correlation for 2 speakers. As Figure 1 shows, as F0 varied OQ remained fairly constant at a value close to 1 for speakers F1, F2, F4, F5, and M3, and varied by at least 0.2 with increasing F0 for speakers F3, M1, and M2, consistent with the variable correlations just described.

Multiple linear regression was applied to examine the extent to which F0, OQ, SQ, and intensity contributed jointly to predicting H1*-H2* values in phonations with changing F0. Results of these regressions are shown in Table 2, which lists the standardized regression coefficients and R^2 values for each speaker. Regression coefficients reflect the relative importance of the different factors in predicting H1*-H2* in that analysis. F0 and OQ were significant predictors of H1*-H2* for all speakers. SQ was also a significant (but less important) factor for 6 speakers. The effect of intensity was not significant for 5 speakers, and was relatively small for the remaining 3 speakers. This is expected because speakers were asked to maintain a constant vocal intensity during phonation, and increases in F0 are not necessarily accompanied by intensity increases.

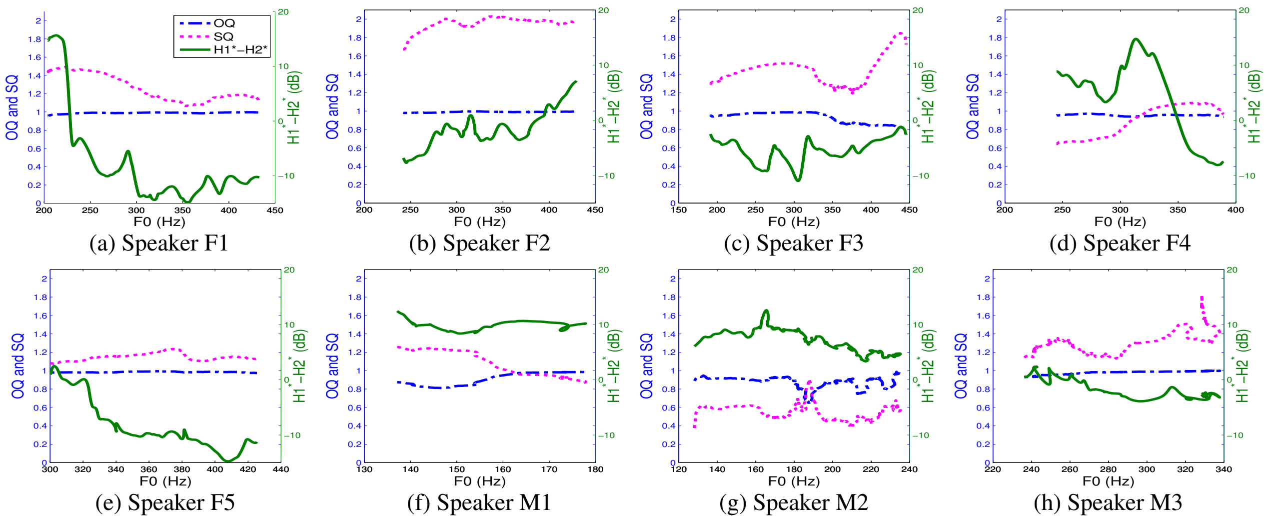
The present results suggest that H1*-H2* may sometimes vary as a function of F0 alone, with OQ and SQ remaining rather constant (e.g., speaker F5). This can be explained by source-filter interaction [4], which has the effect of skewing the glottal flow pulse compared to the glottal area pulse. This effect in turn may result in decreased H1*-H2*, especially when source-filter interaction is strong. According to [4], the degree of interaction depends on the mean glottal area, which appears to be rather high for speaker F5, whose OQ is close to 1. Visual inspection of the HSV also confirmed that speaker F5 exhibits a prominent posterior glottal gap throughout the utterance, suggesting the presence of strong source-filter interactions. The other speakers also exhibit glottal gaps, although the size varies during the phonation.

3.2. Experiment 2: The effect of increasing vocal intensity

In this subsection, we examine data from phonations with changing intensity. Table 3 shows the correlations between intensity and OQ, SQ, and H1*-H2*, and Figure 2 shows the same variables as a function of increasing intensity for speakers F1-F5 and M1, M2, and M4.[2] For 6 out of 8 speakers, intensity was negatively correlated with H1*-H2* (cf. [20]), and it was

[1] Speaker M4 did not participate in Experiment 1. Direct visualization of the vocal folds was not available for the entire utterance for speakers F5 and M3, due to the position and angle of the laryngoscope. Only the segment with a clear view of the glottis was selected for analysis. Therefore the F0 range is limited for these two speakers.

[2] Speaker M3 did not participate in Experiment 2.



(a) Speaker F1 (b) Speaker F2 (c) Speaker F3 (d) Speaker F4

(e) Speaker F5 (f) Speaker M1 (g) Speaker M2 (h) Speaker M3

Figure 1: *OQ, SQ, and H1*-H2* as a function of F0 for speakers F1-F5 and M1-M3.*

Table 1: Correlations between F0 and H1*-H2*, OQ, and SQ for phonations with changing F0 from 8 different speakers. Coefficients shown are significant at p<0.001 (p values have been corrected for multiple comparisons).

Speaker	F1	F2	F3	F4	F5	M1	M2	M3
F0&H1*-H2*	-0.66	0.78	0.39	-0.75	-0.86	-	-0.85	-0.79
F0&OQ	0.65	0.40	-0.82	-0.45	-	0.89	-	0.90
F0&SQ	-0.85	0.56	0.27	0.92	-	-0.96	-0.23	0.77

Table 2: Standardized regression coefficients and R^2 values for multiple linear regression analyses relating F0, OQ, SQ, and intensity to H1*-H2*. Coefficients shown are significant at p<0.001.

Speaker	F1	F2	F3	F4	F5	M1	M2	M3
F0	0.33	0.78	-0.25	-1.73	-0.84	-2.32	-0.69	-0.40
OQ	-0.78	0.19	-0.69	-0.50	-0.24	1.11	0.09	-0.70
SQ	0.56	-	0.33	0.62	-0.11	-1.28	-	0.20
Intensity	-	-	-	0.33	-	-	-0.23	0.12
R^2	0.87	0.64	0.42	0.91	0.83	0.72	0.76	0.89

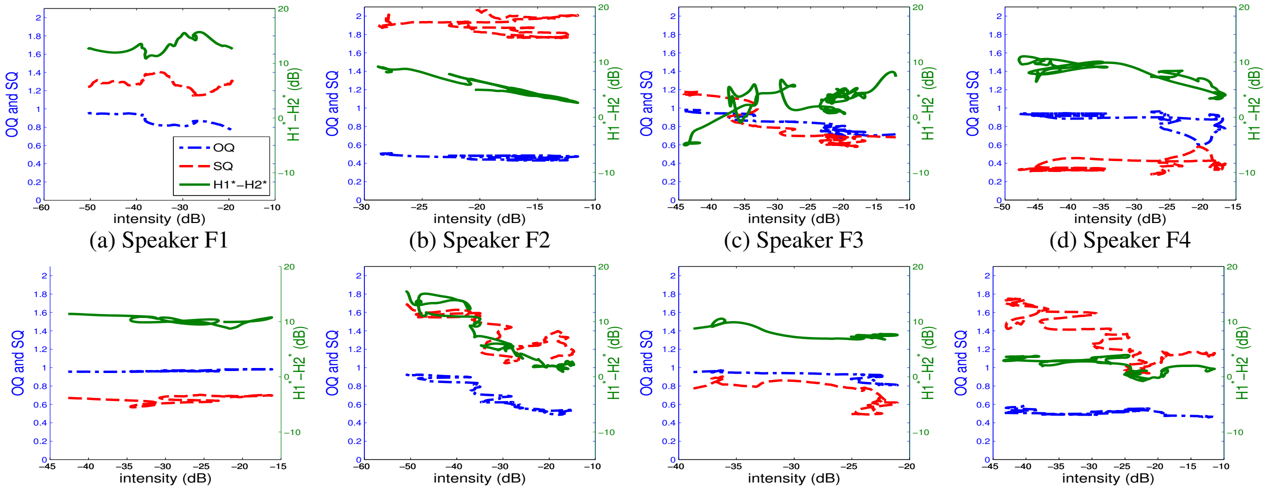
negatively correlated with OQ for 7 out of 8 speakers. Intensity and OQ were positively correlated for speaker F5, for whom OQ was nearly constant (range = 0.95 to 0.98).

Table 3: Correlations between intensity and H1*-H2*, OQ, and SQ, for phonations with changing intensity from 8 different speakers. Coefficients shown are significant at p<0.001. “Int” denotes intensity.

Speaker	F1	F2	F3	F4	F5	M1	M2	M4
Int&H1*-H2*	0.51	-0.98	0.60	-0.82	-0.51	-0.94	-0.81	-0.64
Int&OQ	-0.79	-0.55	-0.91	-0.53	0.87	-0.94	-0.62	-0.19
Int&SQ	-0.50	-	-0.91	0.34	0.63	-0.64	-0.66	-0.82

Contrary to predictions, the correlation between intensity and H1*-H2* was positive for speakers F1 and F3. One hypothesis is that the increase in vocal intensity was accomplished by an increase in subglottal pressure, which had the effect of decreasing pulse skewness [36], with the resultant more symmetric pulse shape leading to a higher H1*-H2*. Although a louder voice is assumed to have a lower OQ [17, 21, 22], which in turn is assumed to lead to a lower H1*-H2*, the effect of decreasing pulse skewness may have offset or even exceeded the effect of decreasing OQ. Regression analysis showed that SQ is indeed the best predictor of H1*-H2* for speaker F1 (see Table 4). This hypothesis could be verified by a computational voice production model.

For speaker F3, SQ was not a significant predictor of H1*-H2* (Table 4). Figure 2 shows that for this speaker SQ drops from approximately 1.2 to 0.6 as intensity increased across the utterance. Theoretically, a symmetric glottal pulse should have a higher H1*-H2* than an asymmetric pulse, assuming all other factors (such as OQ) are constant. This suggests that H1*-H2* should increase, then decrease when SQ drops from 1.2 to 0.6 (SQ=1 corresponds to a symmetric pulse, and therefore should have the highest H1*-H2*). Conventionally, the glottal opening phase is assumed to be longer than the glottal closing phase (rightward-skewed, or SQ<1), at least for glottal flow data. Therefore, glottal pulse skewness has generally been demonstrated to be monotonically correlated with H1*-H2* in theoretical modeling studies [9, 37]. However, HSV-based studies have recently reported that the opening phase can be much shorter than the closing phase (i.e., leftward-skewed glottal pulse) [38, 39]. For example, the duration of the opening phase was as short as only 1/2 of that of the closing phase (i.e., SQ=2), for breathy voices in [38] and some subjects in [39]. We hypothesize that, in such cases, the glottal flow pulse may also be leftward-skewed, especially when nonlinear source-filter coupling (the effect of delaying the pulse peak) is weak. An example could be a loud voice: as subglottal pressure increases and vocal folds become more adducted, the glottal source impedance is much higher than the impedance to the vocal tract (linear source-filter coupling), and therefore the flow pattern resembles the area function more closely [4, 36]. Under this hypothesis, the relationship between SQ and H1*-H2* is not monotonic and can not be explicitly modeled in a linear regression analysis. Specifically, H1*-H2* should first increase, then decrease with increasing SQ, with a maximum corresponding to SQ=1. One alternative is to use a new parameter to account for the “absolute skewness” (AS) of the area pulse, defined as AS=|SQ-1|. With this definition, AS is always positive regardless of leftward or rightward skewness of the pulse shape, and captures the “deviation” from perfect pulse symmetry. Replacing SQ with AS in the regression analysis (Table 4) for speaker F3 shows that AS is a significant predictor of H1*-H2* (β=-0.68), with R^2 increased from 0.45 to 0.55. For the other speakers, SQ is either above 1 (speakers F1, F2, M1, and M4) or below 1 (speakers F4, F5, and M2) throughout the entire utterance, so its theoretical relationship to H1*-H2* should be



(a) Speaker F1 (b) Speaker F2 (c) Speaker F3 (d) Speaker F4

(e) Speaker F5 (f) Speaker M1 (g) Speaker M2 (h) Speaker M4

Figure 2: *OQ, SQ, and H1*-H2* as a function of increasing intensity for speakers F1-F5 and M1, M2, and M4.*

monotonic. This hypothesis could also be verified using a computational voice production model with a nonlinear source-filter framework (e.g., [39]).

Multiple linear regression was applied to relate intensity, OQ, SQ, and F0 jointly to predicting H1*-H2* values. As Table 4 shows, intensity was a significant predictor of H1*-H2* for all the speakers. F0 was also a significant predictor of H1*-H2* for all the speakers except F2.

Table 4: Standardized regression coefficients and R^2 values for multiple linear regression analyses relating intensity, OQ, SQ, and F0 to H1*-H2*. Coefficients shown are significant at p<0.001.

Speaker	F1	F2	F3	F4	F5	M1	M2	M4
F0	-1.21	-	-0.38	0.55	-1.05	0.04	-0.11	-0.67
OQ	-0.69	-	-0.44	-	0.98	0.89	-0.46	-
SQ	-1.23	-0.07	-	0.23	-0.36	-0.04	-	0.73
Intensity	0.53	-0.94	0.32	-1.26	-0.37	-0.15	-1.05	0.55
R^2	0.87	0.97	0.45	0.81	0.67	0.98	0.79	0.78

4. Discussion

While previous theoretical studies typically measured pulse skewness from glottal flow pulses, in this study glottal pulse skewness is measured from glottal area waveforms. The glottal flow pulse is known to be more rightward-skewed than the glottal area pulse [5, 6], due to the interaction between pressure from the lungs and the glottal area function [40] as well as the interaction between glottal area and the vocal tract system [41]. Although glottal flow pulse skewness is assumed to be directly related to harmonic magnitudes, glottal area pulse skewness has been used instead in previous HSV-based studies [11, 39] because of its close relationship to glottal flow pulse shape. It is possible that the skewness of the glottal flow pulse, which was not measured in this study, is responsible for some variability observed in the results due to source-filter interactions.

We hypothesized that the relationship between H1*-H2*, F0, and intensity is speaker dependent, and that these dependencies are mediated at least partly by the degree of source-filter interaction. The use of HSV of the vocal folds allows for the direct observation of glottal vibratory pattern such as incomplete glottal closure, and provides insights into the degree of non-linear source-filter interaction. These observations lend experimental evidence to our hypothesis, which partially explains some variable trends observed in this study. The hypothesis can be further verified using a computational voice production model with a nonlinear source-filter framework (e.g., [39]).

Previous studies have suggested that increases in loudness are typically accompanied by F0 increases in speakers without formal singing training [22,42]. Although speakers in this study were asked to vary F0 and intensity separately, they inevitably increased F0 when trying to increase loudness. On average, F0 increased by 35 Hz during phonations in Experiment 2. This implies that the effect of increasing intensity may have been "disturbed" by the effect of increasing F0. However, on the other hand, this may also suggest that perhaps these two effects should be studied jointly, due to their frequent co-variation in natural speech.

Although similar studies have been done on EGG data (e.g., [23]), it should be kept in mind that the results of this study are not directly comparable to those results. The EGG signal measures the in-depth contact of vocal folds and reflects both vertical and horizontal contact, but neglects leakages caused by incomplete glottal closure. The glottal area from HSV reflects the area of separation between the vocal folds as projected by the image of the glottis, which captures glottal gap but does not reflect vertical closure. A weak interclass correlation between OQ values from HSV and EGG signals was reported in [19]. The glottal area is a quantitative component in nonlinear source-filter interaction theory [4], but the theoretical status of the waveform of the EGG signal remains somewhat unclear.

5. Conclusions

This study investigated the relationship between H1*-H2*, F0, vocal intensity, and measures of glottal pulse shape. Analyses of synchronous audio and laryngeal high-speed video recordings showed that H1*-H2* may sometimes vary as a function of F0 alone, with OQ and SQ remaining rather constant, hypothetically when nonlinear source-filter interaction is strong. Although conventionally H1*-H2* is assumed to decrease with increasing vocal intensity due to a corresponding decrease in OQ, results showed examples where H1*-H2* increased with increasing vocal intensity, hypothetically when the effect of decreasing pulse skewness exceeds the role of decreasing OQ. In some phonatory modes, the glottal area pulse can be significantly leftward-skewed, and the relationship between SQ and H1*-H2* may no longer be as monotonic as previously assumed. Future work will include using a computational voice production model to verify these hypotheses, as well as collecting more data to enable across-speaker statistical analyses.

6. References

[1] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *J. Acoust. Soc. Am.*, vol. 101, pp. 466–481, 1997.

[2] M. Iseli, Y.-L. Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *J. Acoust. Soc. Am.*, vol. 121, pp. 2283–2295, 2007.

[3] G. Chen, X. Feng, Y.-L. Shue, and A. Alwan, "On using voice source measures in automatic gender classification of children's speech." in *Interspeech*, 2010, pp. 673–676.

[4] I. Titze, "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.*, vol. 123, pp. 2733–2749, 2008.

[5] M. Rothenberg, *Acoustic interaction between the glottal source and the vocal tract, Vocal fold physiology*. Tokyo: University of Tokyo Press, 1981, pp. 305–323.

[6] K. N. Stevens, *Acoustic Phonetics*. Cambridge, MA: MIT Press, 1998, pp. 55–126.

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

[8] D. Klatt and L. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, vol. 87, pp. 820–857, 1990.

[9] 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.

[10] M. Swerts and R. Veldhuis, "The effect of speech melody on voice quality," *Speech Commun.*, vol. 33, pp. 297–303, 2001.

[11] J. Kreiman, Y.-L. Shue, G. Chen, M. Iseli, B. Gerratt, J. Neubauer, and A. Alwan, "Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation," *J. Acoust. Soc. Am.*, vol. 132, pp. 2625–2632, 2012.

[12] G. Chen, J. Kreiman, Y.-L. Shue, and A. Alwan, "Acoustic correlates of glottal gaps," in *Interspeech*, 2011, pp. 2673–2676.

[13] D. Childers, D. Hicks, G. Moore, L. Eskenazi, and A. Lalwani, "Electroglottography and vocal fold physiology," *J. Speech Lang. Hear. Res.*, vol. 33, p. 245, 1990.

[14] D. G. Hanson, B. R. Gerratt, and G. S. Berke, "Frequency, intensity, and target matching effects on photoglottographic measures of open quotient and speed quotient," *J. Speech Hear. Res.*, vol. 33, pp. 45–50, 1990.

[15] J. Sundberg, T. F. Cleveland, R. Stone Jr, and J. Iwarsson, "Voice source characteristics in six premier country singers," *J. Voice*, vol. 13, pp. 168–183, 1999.

[16] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch," *J. Voice*, vol. 3, pp. 294–305, 1989.

[17] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*. Cengage Learning, 2000.

[18] J. Koreman, "The effects of stress and F0 on the voice source," *Phonus*, vol. 1, pp. 105–120, 1995.

[19] M. Echternach, S. Dippold, J. Sundberg, S. Arndt, M. F. Zander, and B. Richter, "High-speed imaging and electroglottography measurements of the open quotient in untrained male voices' register transitions," *J. Voice*, vol. 24, pp. 644–650, 2010.

[20] G. Fant, J. Liljencrants, I. Karlsson, and M. Båvegård, "Time and frequency domain aspects of voice source modelling." *ESPRIT*, 1995, BR Speechmaps (6975), Deliverable 27 WP 1.3.

[21] J. Pierrehumbert, "A preliminary study of the consequences of the intonation for the voice source," *STL-QPSR*, vol. 4, pp. 23–36, 1989.

[22] E. Holmberg, R. Hillman, and J. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *J. Acoust. Soc. Am.*, vol. 84, pp. 511–529, 1988.

[23] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo, "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," *J. Acoust. Soc. Am.*, vol. 117, pp. 1417–1430, 2005.

[24] G. Chen, J. Kreiman, B. R. Gerratt, J. Neubauer, Y.-L. Shue, and A. Alwan, "Development of a glottal area index that integrates glottal gap size and open quotient," *J. Acoust. Soc. Am.*, vol. 133, pp. 1656–1666, 2013.

[25] G. Chen, R. A. Samlan, J. Kreiman, and A. Alwan, "Investigating the relationship between glottal area waveform shape and harmonic magnitudes through computational modeling and laryngeal high-speed videoendoscopy," in *Interspeech*, 2013, pp. 3216–3220.

[26] G. Chen, M. Garellek, J. Kreiman, B. R. Gerratt, and A. Alwan, "A perceptually and physiologically motivated voice source model," in *Interspeech*, 2013, pp. 2001–2005.

[27] G. Chen, J. Kreiman, and A. Alwan, "The glottaltopogram: A method of analyzing high-speed images of the vocal folds," *Computer Speech & Language*, vol. 28, pp. 1156–1169, 2014.

[28] ——, "The glottaltopograph: A method of analyzing high-speed images of the vocal folds," in *ICASSP*, 2012, pp. 3985–3988.

[29] M. R. Draper, B. Blagnys, and D. J. Premachandra, "To 'EE' or not to 'EE'," *J. Otolaryngology*, vol. 36, pp. 191–195, 2007.

[30] "GlotAnTools," [Computer software], Version 5, Department of Phoniatrics and Pediatric Audiology, University Hospital, Erlangen, Germany. Retrieved Nov. 11, 2011.

[31] G. Chen, Y.-L. Shue, J. Kreiman, and A. Alwan, "Estimating the voice source in noise," in *Interspeech*, 2012, pp. 1600–1603.

[32] P. Mermelstein, "Automatic segmentation of speech into syllabic units," *J. Acoust. Soc. Am.*, vol. 58, pp. 880–883, 1975.

[33] Y.-L. Shue, "VoiceSauce: a program for voice analysis," 2010, http://www.seas.ucla.edu/spapl/voicesauce/ (Last viewed Apr. 1, 2012).

[34] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech Commun.*, vol. 27, pp. 187–207, 1999.

[35] K. Sjölander, "The snack sound toolkit," KTH Stockholm, Sweden, 2004, http://www.speech.kth.se/snack/ (Last viewed Aug. 1, 2009).

[36] G. Fant, "Preliminaries to analysis of the human voice source," *STL-QPSR*, vol. 4, pp. 1–27, 1982.

[37] ——, "The LF-model revisited. Transformations and frequency domain analysis," *STL-QPSR*, vol. 36, pp. 119–156, 1995.

[38] 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.

[39] D. D. Mehta, M. Zañartu, T. F. Quatieri, D. D. Deliyski, and R. E. Hillman, "Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy," *J. Acoust. Soc. Am.*, vol. 130, pp. 3999–4009, 2011.

[40] T. V. Ananthapadmanabha and G. Fant, "Calculation of true glottal flow and its components," *Speech Commun.*, vol. 1, pp. 167–184, 1982.

[41] I. Titze and B. Story, "Acoustic interactions of the voice source with the lower vocal tract," *J. Acoust. Soc. Am.*, vol. 101, pp. 2234–2243, 1997.

[42] P. Gramming, J. Sundberg, S. Ternström, R. Leanderson, and W. H. Perkins, "Relationship between changes in voice pitch and loudness," *J. Voice*, vol. 2, pp. 118–126, 1988.