Investigating the relationship between glottal area waveform shape and harmonic magnitudes through computational modeling and laryngeal high-speed videoendoscopy

[1]Gang Chen, [2]Robin A. Samlan, [2]Jody Kreiman, and [1]Abeer Alwan

[1]Department of Electrical Engineering, University of California, Los Angeles
[2]Department of Head and Neck Surgery, School of Medicine, University of California, Los Angeles
Work supported by NSF and NIH

Objective

- Investigate the relationship between OQ (the glottal open quotient; the relative amount of time the glottis is open within a glottal vibratory cycle) and H1*-H2* (the relative amplitudes of the first two harmonics of the voice source).
- Compare results from a computational voice production simulation and high-speed laryngeal videoendoscopy data.

Background

- Increases in OQ are widely assumed to be the physical precursors of perceived breathiness, in part because of consequent increases in H1*-H2* [1].
- Empirical studies used electroglottographic (EGG) data or inverse-filtered acoustic signals, and varying levels of correlation between H1*-H2* and OQ have been reported [2, 3].
- In the LF model [4] the relationship is expressed as H1*-H2*=-6+0.27exp(0.055 OQ) [5].
- Modeling studies do not currently fully explain the observed variability in experimental studies.

Data and methods

Human subject data

- Synchronous audio and high-speed video recordings of the vocal folds
- Five subjects (4 male + 1 female)
- Vowel /i/
- Gradually change their phonations from breathy to pressed while holding F0 and vowel quality as steady as possible
- High-speed imaging: 10,000 frames/sec; a resolution of 208×352 pixels

Extract glottal area from high-speed recording of the vocal folds

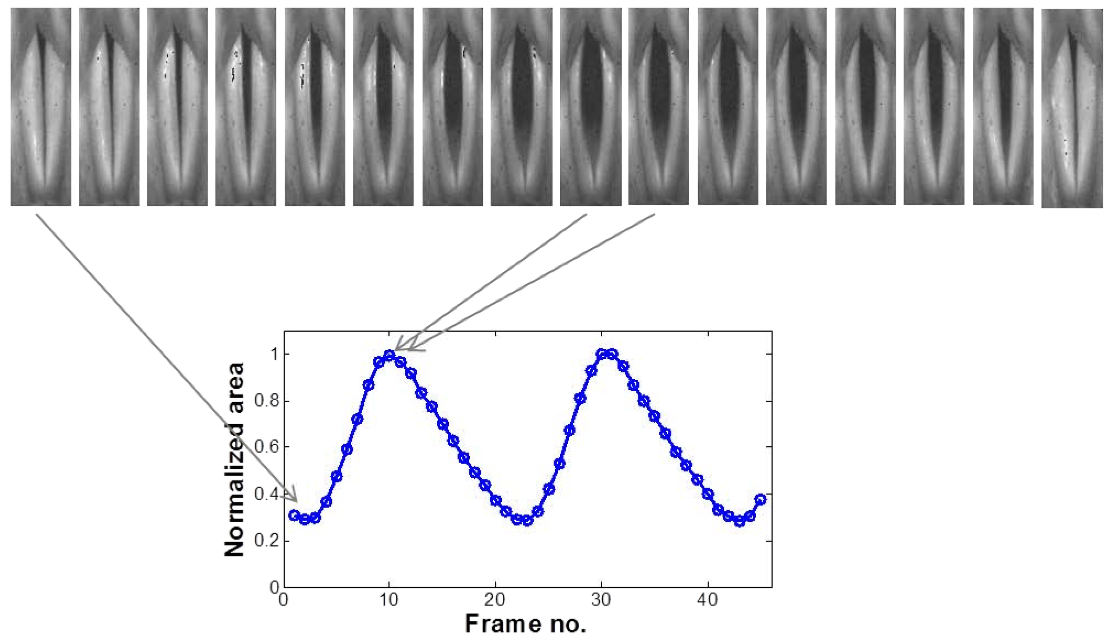


Figure 1: Glottal area extraction

Data and methods

Measures from high-speed imaging

- OQ: the time from the first opening instant to the onset of maximum closure (or minimum area), divided by the length of the current glottal cycle.
- DC (i.e., the glottal gap size): defined as the minimum glottal area normalized by the maximum glottal area in each glottal cycle.

Acoustic measure

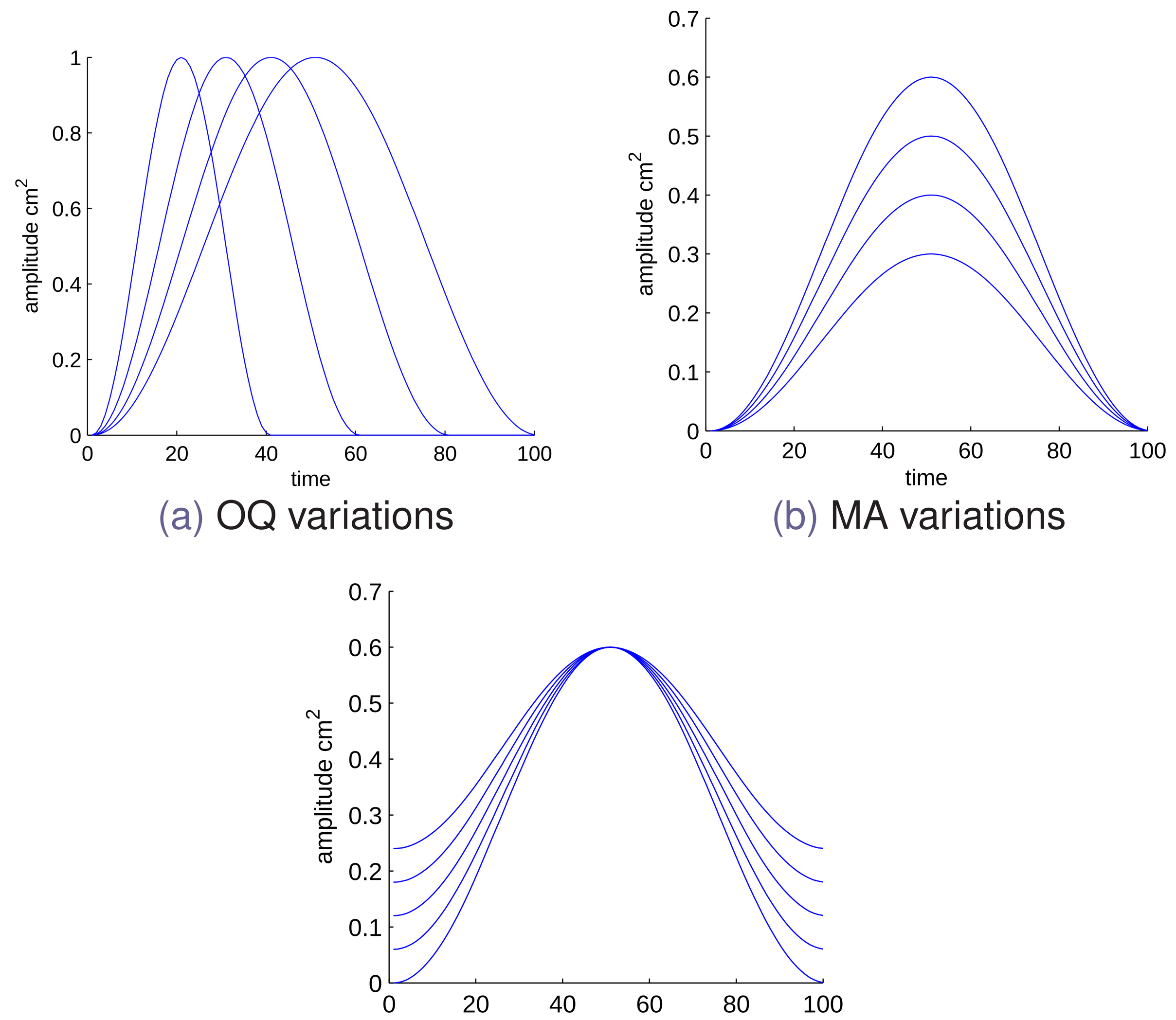
- H1*-H2*: measured from the audio signals with VoiceSauce software [6].

Computational model simulation

Generating glottal area waveforms

The parametric voice source model in [7] (denoted EE2) was chosen for this study.

- Provides greater glottal pulse shape flexibility than the LF model.
- Allows for direct control of the glottal area pulse shape compared to kinematic models.
- Parameters:
  (1): OQ
  (2): The maximum amplitude (MA) of the glottal area waveform
  (3): DC



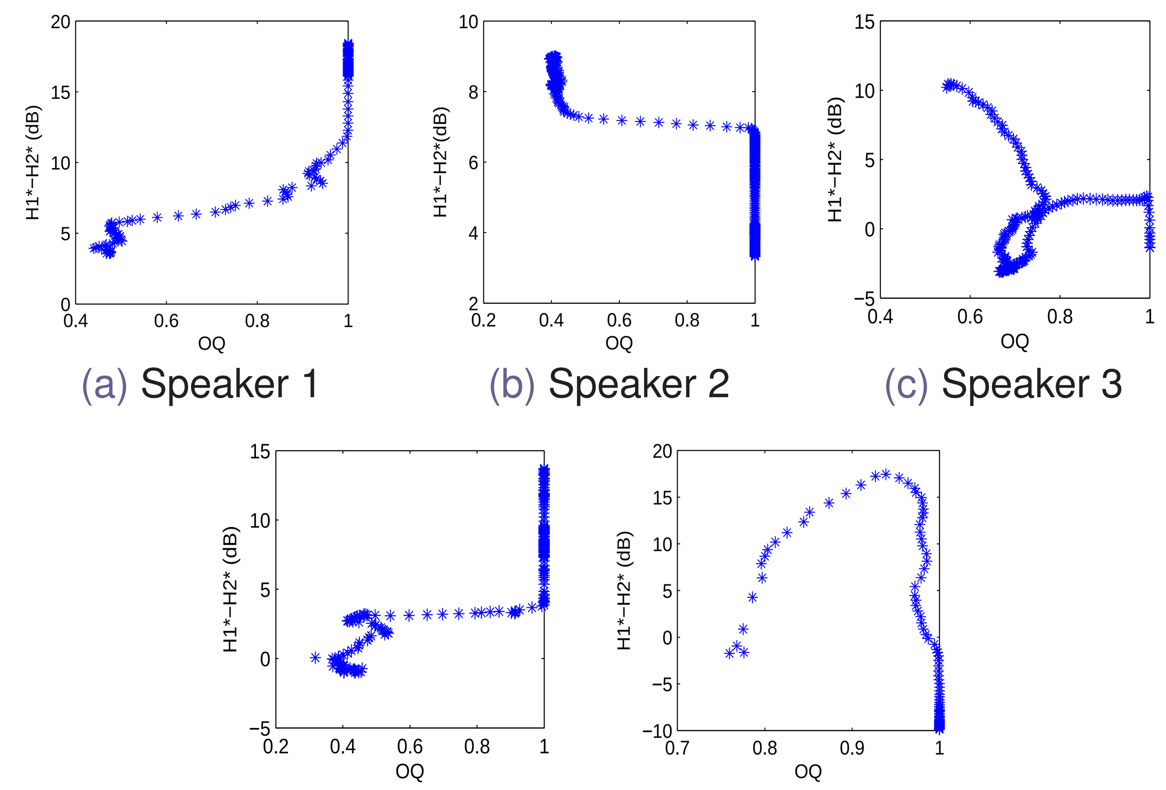
(a) OQ variations (b) MA variations (c) DC variations

Figure 2: *Generated glottal area waveforms using the EE2 source model.*

Simulating nonlinear source-filter interactions

- The glottal area was acoustically coupled to the trachea and vocal tract airway system.
- Nonlinear source-filter interactions were simulated using "LeTalker" [8] software.

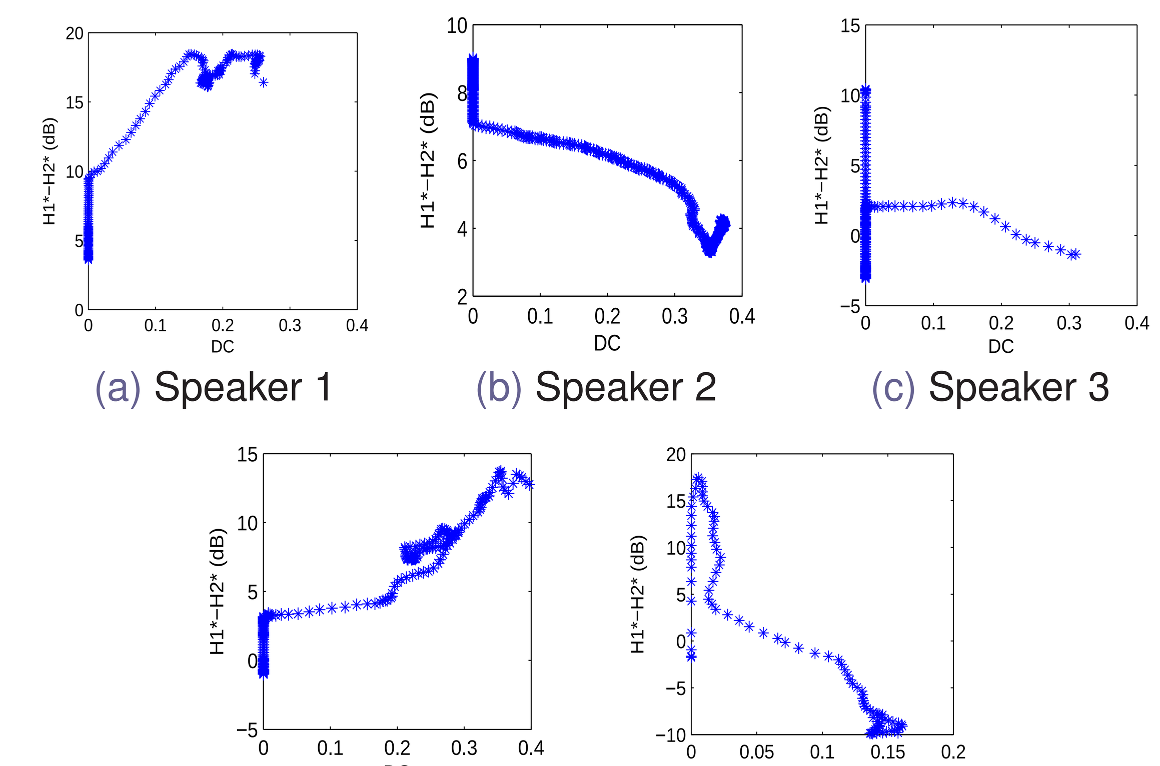
Results: Human subject data



(a) Speaker 1 (b) Speaker 2 (c) Speaker 3 (d) Speaker 4 (e) Speaker 5

Figure 3: *H1*-H2* vs. OQ for speakers 1-5.*

These varying patterns suggest that the relationship between H1*-H2* and phonatory characteristics may be speaker dependent.



(a) Speaker 1 (b) Speaker 2 (c) Speaker 3 (d) Speaker 4 (e) Speaker 5

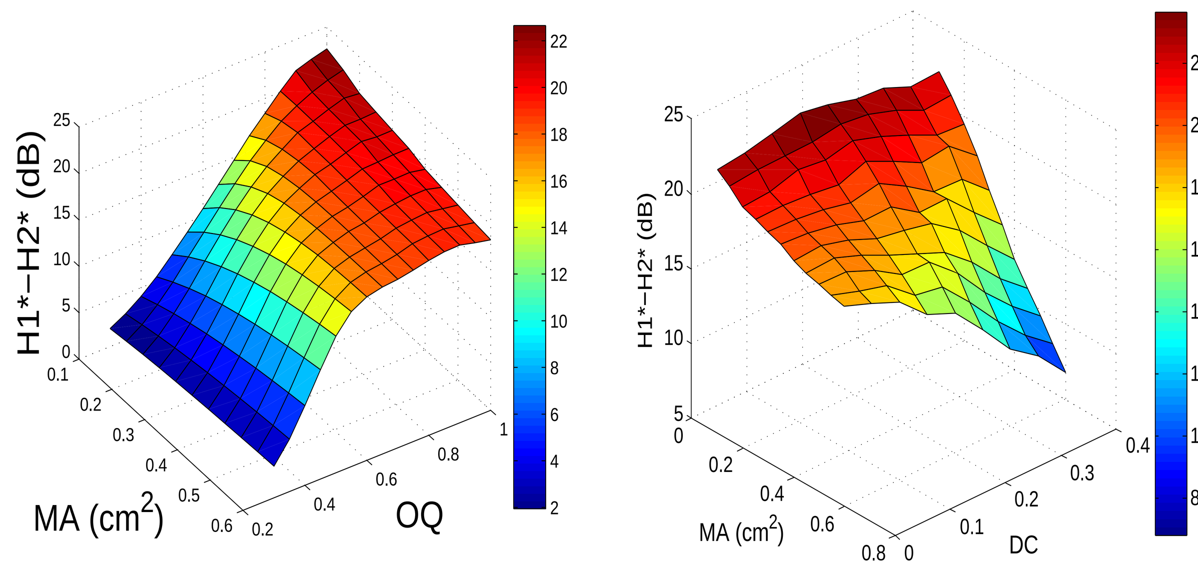
Figure 4: *H1*-H2* vs. DC for speakers 1-5.*

- Despite this variability, the cases in which glottal gaps were observed are typically assigned an OQ of 100% or close to 100%.
- Thus, the variability in H1*-H2 with varying DC partially contributes to the observed variability in the relationship between H1*-H2* and OQ in previous studies

References

[1] D. Klatt and L. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *JASA.*, vol. 87, pp. 820–857, 1990.
[2] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. Guiod, and S. L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *JSHR.*, vol. 38, pp. 1212–1223, 1995.
[3] M. Swerts and R. Veldhuis, "The effect of speech melody on voice quality," *Speech Commun.*, vol. 33, no. 4, pp. 297–303, 2001.
[4] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, vol. 4, pp. 1–13, 1985.
[5] G. Fant, "The LF-model revisited. Transformations and frequency domain analysis," *STL-QPSR*, vol. 36, pp. 119–156, 1995.
[6] Y.-L. Shue, "VoiceSauce: a program for voice analysis," 2010, http://www.ee.ucla.edu/~spapl/voicesauce/.
[7] G. Chen, Y.-L. Shue, J. Kreiman, and A. Alwan, "Estimating the voice source in noise," in *Interspeech*, 2012, pp. 1600–1603.
[8] B. Story, "LeTalker: Lumped-element model of phonation," 2012, http://sal.shs.arizona.edu/~bstory/LeTalkerMain.html.
[9] 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," *JASA.*, vol. 132, pp. 2625–2632, 2012.
[10] 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," *JASA.*, vol. 130, pp. 3999–4009, 2011.

Results: Model simulations



When MA is relatively small

- H1*-H2* increases monotonically with increasing OQ (similar to the observations in Figures 3a and 3e).
- H1*-H2* varies very little with increasing DC (about 2 dB).

When MA is relatively large

- H1*-H2* first increases and then slightly decreases with increasing OQ.
- H1*-H2 decreases monotonically with increasing DC (similar to the human data in Figures 4b and 4e).
- Attributes to an increased degree of source-filter interaction. OQ (DC) increases → the mean glottal area also increases → a higher degree of source-filter interaction → the effect of "skewing" the glottal flow waveform → results in decreased H1*-H2*.

Conclusion

- Human subject data: the effects of OQ and glottal gap size on H1*-H2* may be variable and speaker dependent. H1*-H2* may increase or decrease with increasing glottal gap size, allowing more variability of relationship between H1*-H2* and OQ to be observed.
- Model simulations: supported the observed variabilities and suggested that this relationship depends on mean glottal area (MA), a parameter associated with the degree of source-filter interaction but not directly measurable from high-speed images of the vocal folds.
- It is possible that the result derived from laryngeal high-speed recordings (based on time quotient measures) could be somewhat incomplete or inconclusive.
- The simulation results in this study may also provide a possible explanation for the large interspeaker variability and weak correlations between time domain measures and acoustic measures reported in previous high-speed laryngoscopy-based studies [9, 10].

Gang Chen - Electrical Engineering Department - University of California, Los Angeles

Mail: gangchen@ucla.edu

WWW: http://www.seas.ucla.edu/~gangchen