Developing a computational model of aging voice

Brad H. Story & Robin A. Samlan

Speech, Language, and Hearing Sciences, University of Arizona

QR code to access audio demonstrations of simulated vowels

AQL 2023

1. Background

- The purpose of this study was to test the feasibility of using glottal area waveforms (GAW) obtained from high-speed videoendoscopic (HSV) recordings directly as the vibratory source in a computational model of speech production. The GAW captures the variation of vibratory patterns that occur during natural voice production and provides a means for eventually enhancing voice source models so that they better represent the aging voice.
- The current project represents the first iteration of this approach and is focused on transforming raw glottal area waveforms to a form suitable for use in the model, and proof-of-concept that the output signals of the model are realistic simulations of natural vowel productions.

2. Participant selection

- Four participants were selected from a larger study that included audio recordings and HSV recordings of a group of individuals ages 65 and older. Some self-reported some degree of age-related dysphonia and others normal voice.
- One male and one female participant were selected whose cepstral peak prominence (CPP) calculated for connected speech samples were at the 25th and 75th percentiles of the larger dysphonic group. These participants will be referred to as M25, M75, F25, and F75, respectively. The age and measured CPP of the four participants was:

Participant	Age (yrs)	CPP (dB)
M25	78	2.9
M75	71	8.0
F25	78	4.4
F75	83	6.8

3. GAW: Measurement, processing, and analysis

- A one-second segment of a glottal area waveform (GAW) was extracted for each of the four participants from a high-speed video-endoscopic recording completed at 4000 frames per second while producing a sustained vowel (i.e., a nondescript vowel due to presence of the endoscope). The GAW was measured using "Glottal Analysis Tools" (GAT) developed by Michael Döllinger and colleagues at University Hospital Erlangen (Germany).
- Each GAW was resampled from 4000 Hz to 44100 Hz for the male participants, and to 50000 Hz for the female participants. The need for different sampling frequencies is due to the relation of vocal tract length and sampling frequency in the wave-reflection algorithm used in the computational model (described in a subsequent section).
- The resampled GAWs were then low-pass filtered with a 20th order FIR filter using a cutoff frequency of 2000 Hz to remove artifacts from the resampling process.
- In their raw form, as measured from HSV, the amplitude of the GAWs is in pixel units. Since there is no calibration of the images relative to actual cross-sectional area, they must be normalized to an assumed maximum area in order to use them in the computational model. For this study, the resampled and filtered GAWs were normalized such that their maximum value was 0.1 cm^2.
- To provide an indication of the degree of irregularity, for each GAW, the fundamental frequency, open quotient, and skewing quotient were determined for each glottal cycle over the time course of the waveform.

Acknowledgements

Research supported by NIDCD R21 DC016356 (Samlan), NIDCD R01 DC017998 (subaward, Story), and a Galileo Circle Fellows grant (Story).

3. GAW: Segments for modeling

- The one-second segments that were selected from HSV recordings of each of the four participants are shown in the left column of Figure 1, whereas shorter segments showing three glottal cycles are plotted in the right column. Note that the durations of the segments in the right column are dependent on the fundamental frequency.

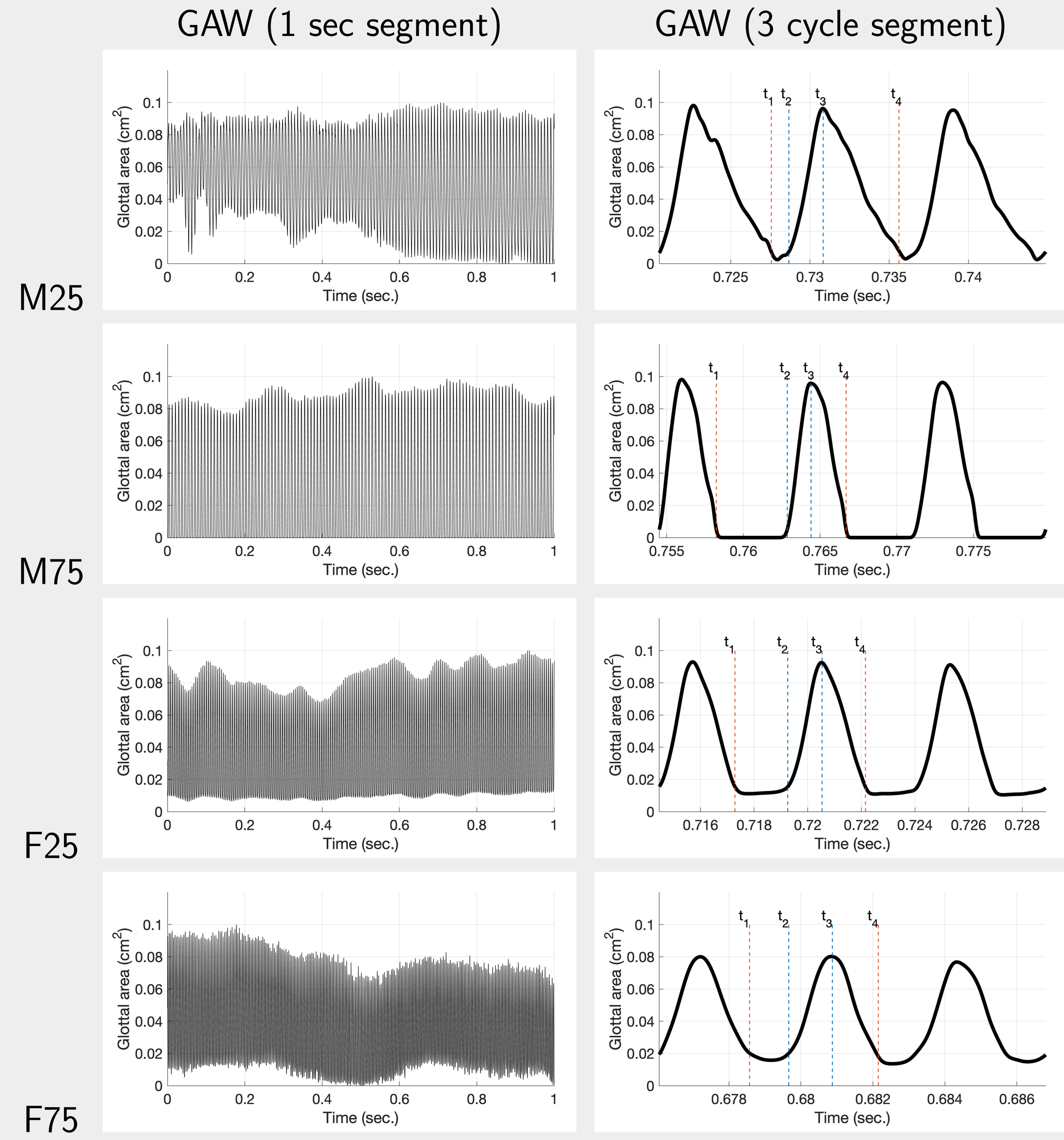

Figure 1: One-second (left column) and 3-cycle (right column) GAW segments for M25, M75, F25, and F75.

- The four time markers indicated in each short segment were determined for every cycle of the one-second glottal waveforms and then used to calculate the fundamental frequency, open quotient, and skewing quotient with the following equations:

$$f_o = \frac{1}{t_4 - t_1} \qquad Q_o = \frac{t_4 - t_2}{t_4 - t_1} \qquad Q_s = \frac{t_3 - t_2}{t_4 - t_3}$$

4. GAW Analysis: f_o, Q_o, and Q_s

- The analyses shown in Figure 2 for M25 and M75 indicate nearly the same f_o, but widely different Q_o values that are well aligned with their respective CPP measurements shown previously. For F25 and F75, the f_o differs by about 60 Hz, but the Q_o valuese are nearly the same. The Q_s values across all cases is about 1.0 or less. All quantities indicate cycle-to-cycle variability.

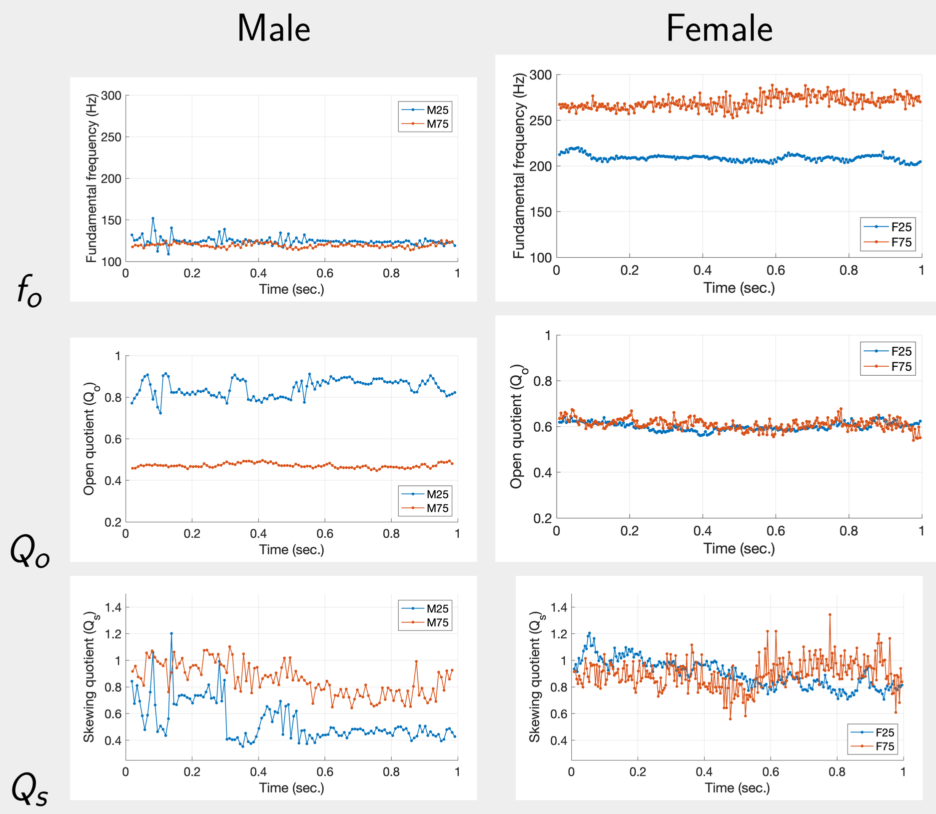

Figure 2: Time-varying f_o, Q_o, and Q_s for the four GAW segments from Figure 1.

5. Modeling

- Each of the four glottal area waveforms shown in Fig. 1 were used as the vibratory source in the TubeTalker model (cf., Story, 2013) as illustrated in Figure 3. Specifically, a GAW (shown in green) replaced the vocal fold model, and the simulation produced the other quantities shown (subglottal pressure, glottal flow, and radiated pressure).
- For each simulation, the vocal tract was configured as an /ɑ/ vowel based on the male and female age-dependent vocal tract model in Story et al., (2018). The vocal tract length was 17.5 cm for simulations of the male voices and 15.5 cm for simulations of the female voices.
- In each case, the particular tract shape was tuned (Story, 2006) so that the formant frequencies were closely aligned with the formant frequencies measured from a recording of the participant producing an /ɑ/ vowel.

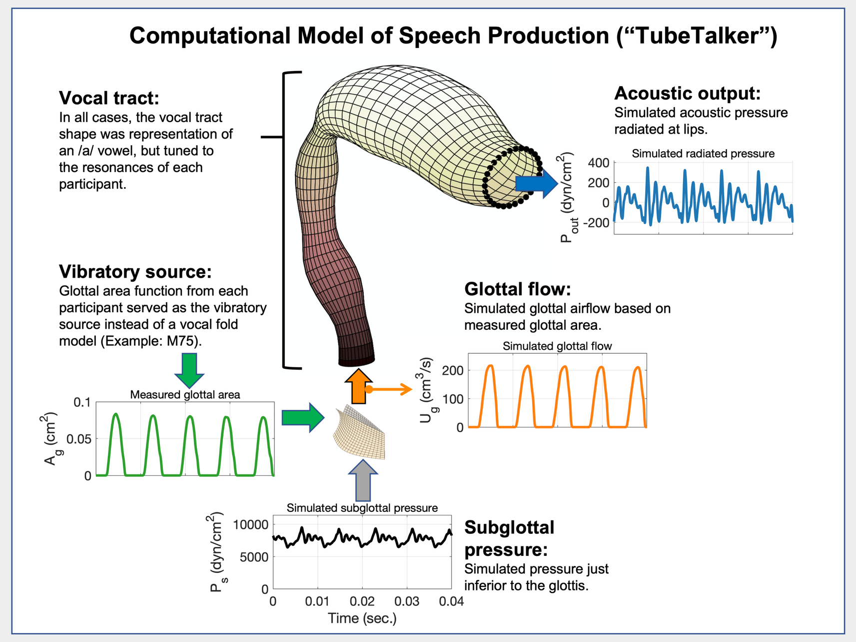

Figure 3: TubeTalker model using measured GAWs as the vibratory source.

6. Simulated cases: M25 and M75

- The simulated /ɑ/ vowels for the two male participants, M25 and M75, are shown in Figure 4. 0.04 second segments of the four waveforms are in the left column where the glottal area (green) is the measured GAW, and the other signals are simulated.
- In the right column of Figure 4 are the radiated pressure (audio) waveforms and corresponding narrow-band spectrograms of each simulation. Audio files can be heard by using the QR code at the top of the poster to access the sal.arizona.edu website.

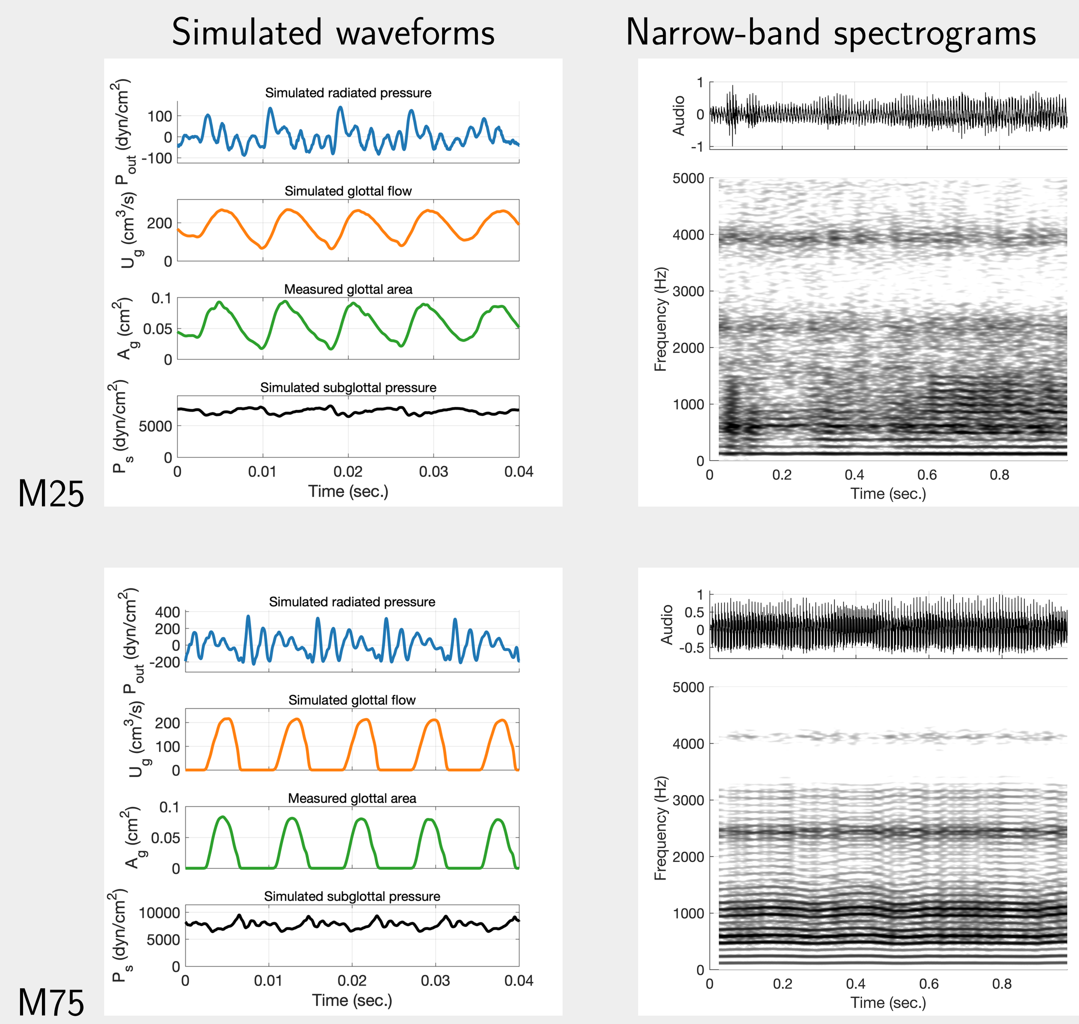

Figure 4: Waveforms and narrow-band spectrograms generated by TubeTalker based on GAWs produced by M25 and M75.

6. Simulated cases: F25 and F75

- The simulated /ɑ/ vowels for the two female participants, F25 and F75, are shown in Figure 5. 0.02 second segments of the four waveforms are in the left column where the glottal area (green) is the measured GAW, and the other signals are simulated.
- It is particularly apparent in these two cases that the glottal flow pulses are skewed rightward in time whereas the glottal area pulses (measured) are skewed to the left.
- In the right column of Figure 5 are the radiated pressure (audio) waveforms and corresponding narrow-band spectrograms of each simulation. Audio files can be heard by using the QR code at the top of the poster.

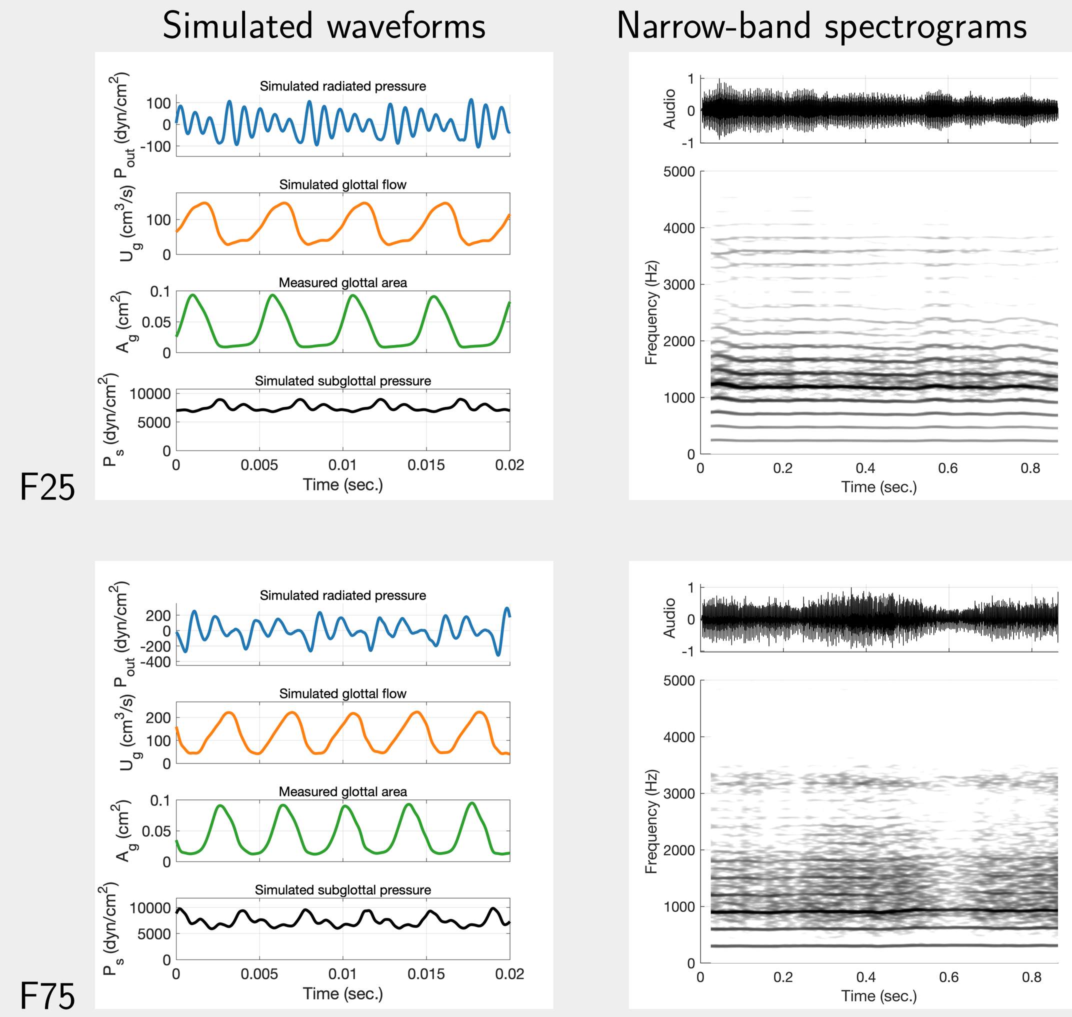

Figure 5: Waveforms and narrow-band spectrograms generated by TubeTalker based on GAWs produced by F25 and F75.

6. Conclusions

- The purpose of this study was a proof-of-concept project to determine the feasibility of using measured glottal area waveforms as the vibratory source in a computational speech production model.
- The results show that the simulations are realistic depictions of vowel production with varying degrees of irregularity characteristic of the aging voice. Informal listening to audio of the simulations and recordings of the participants indicates that the glottal area waveforms capture some idiosyncratic characteristics of the original talker.
- Next steps can include a series of voice therapy manipulations for aging voice with a goal of improving our overall understanding of voice production, disorders, and treatment for older adults.

References

Story, B.H., (2013). Phrase-level speech simulation with an airway modulation model of speech production, Computer Speech and Language. 27(4), 989-1010.

Story, B. H., (2006). A technique for "tuning" vocal tract area functions based on acoustic sensitivity functions, J. Acoust. Soc. Am., 119(2), 715-718.

Story, B. H., Vorperian, H., Bunton, K., and Durtschi, R., (2018). An age-dependent vocal tract model for males and females based on anatomic measurements, J. Acoust. Soc. Am., 143(5), 3079-3102. https://doi.org/10.1121/1.5038264