Synchrony of Glottal Area Waveform Parameters During the Production of Obstruents in Vowel Context

João Vítor Possamai de Menezes[1], *Christian Kleiner*[1], *Marie-Anne Kainz*[2], *Matthias Echternach*[2], *Peter Birkholz*[1]

[1]*Institute of Acoustics and Speech Communication, Technische Universität Dresden,*
[2]*Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Munich University Hospital (LMU)*
joao_vitor.possamai_de_menezes@tu-dresden.de

Abstract: Obstruents are phonemes which require partial or total obstruction of airflow through the vocal tract. Their articulation also requires adjustments of the laryngeal settings, e. g., an abduction gesture to stop vocal fold vibration for voiceless obstruents. This study investigated the laryngeal settings during the production of voiced and voiceless obstruents in vowel context to analyze the degree of synchrony of the involved glottal gestures. High-speed laryngoscopy images were used to determine the glottal area waveform, from which the time functions of the parameters open quotient (OQ), fundamental frequency (f_0), and AC and DC amplitude (ACA and DCA) were calculated and analyzed. Significant correlations were found between all pairs of parameters, with *strong* correlations between some of them, e.g. Open Quotient and AC Amplitude. Correlations were also either consistently positive or negative for specific pairs of parameters across all investigated phonemes. These results could point to consistent patterns in laryngeal gestures that could enhance articulatory speech synthesis.

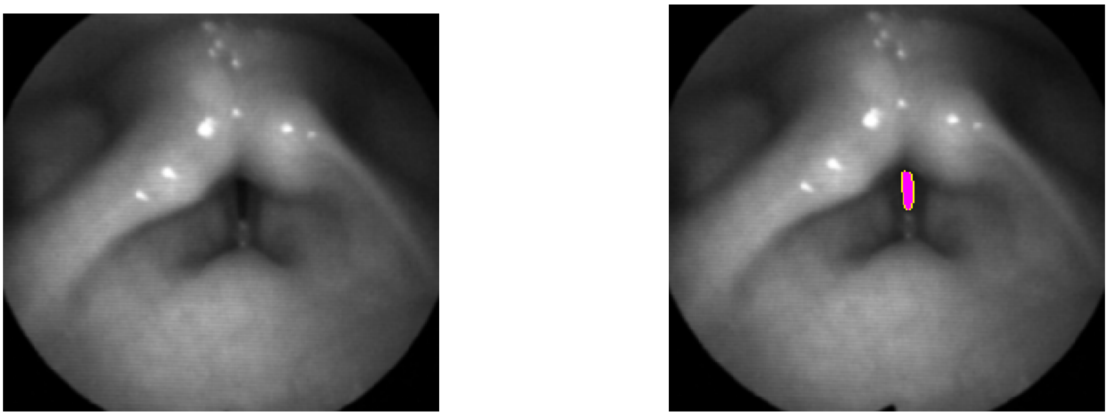
1 Introduction

The production of obstruents involves the adjustment of the laryngeal settings to the specific needs of the obstruents. For example, to generate a voiceless fricative in vocalic context, the vocal folds need to be abducted to stop vocal fold oscillation and reduce the glottal resistance to enable sufficient airflow for the generation of frication noise. Towards the end of the fricative, a vocal fold adduction gesture re-initiates phonation for the following vowel. These laryngeal gestures cause the change of multiple parameters of the glottal area waveform. For example, during the abduction phase there is an increase of the mean glottal area, a decrease of the glottal oscillation amplitude, and potential changes of the open quotient (OQ) and f_0 of the oscillations. The goal of the present study was to explore the specific temporal changes of these parameters in /eCe/ utterances with voiced and voiceless obstruents as the consonant C, and to investigate the potential synchrony or asynchrony between their time functions. For this purpose, glottal area waveforms were obtained from the analysis of high-speed laryngoscopy videos of a single German speaker. Similar investigations were previously performed by, e.g. [1, 2], but using less direct measurement techniques for the glottal parameters, for speakers of different languages then German, and without examining the synchrony of the parameters.

2 Method

This section describes our methodology, namely, how data was recorded, how data was preprocessed and how the glottal area waveform parameters were calculated.

DOI: 10.35096/othr/pub-7080, URN: urn:nbn:de:bvb:898-opus4-70808



(a) Glottis image after bitshift and cropping.

(b) Segmented glottal area using GlottalImageExplorer.

Figure 1 – An example of recorded and segmented glottis image. The glottal area is within the yellow boundaries and colored in pink in Figure 1b and was segmented based on three seed points (positioned manually inside the glottis) and their respective thresholds (also set manually). [color online]

2.1 Data recording

High-speed laryngoscopy with a frame rate of 5000 Hz was performed in a female native speaker of German while she uttered a corpus composed of repeated /Ce/ syllables, with the voiced obstruents /b, d, g, v, z, ʒ/ and the unvoiced obstruents /p, t, k, f, s, ʃ/ as the consonant C. The participant uttered each /Ce/ syllable 5 times in a row (without a pause between the syllables), resulting in /eCe/ clusters. A metronome was playing during the recordings and the speaker was asked to synchronize each CV syllable to the metronome's beat (period of 375 ms).

2.2 Data pre-processing

The raw recorded images of the glottis were subject to bit shift (from 2 to 6 bits), to increase contrast, and were cropped, to limit the image size to 256 × 256 pixels. After these procedures, the images were analyzed with the software GlottalImageExplorer [3], which allowed the segmentation of the glottal area based on the *seeded region growing* method. This yielded waveforms of the glottal area (in pixels) for all recorded images. Figure 1 shows one example of a laryngoscopy image of the glottis before and after segmentation of the glottal area.

The following pre-processing step was the upsampling of the glottal area waveforms from 5000 Hz to 20 000 Hz, increasing the resolution of the signal and, therefore, enabling a more precise recognition of the peaks of each glottal cycle. Pitch marks were then placed at the peaks of each glottal cycle (and at regular intervals in fully voiceless regions), and each interval between two consecutive pitch marks was considered as a glottal period. The peaks were determined using a minimal interval between two peaks of 80 samples. This means that the maximal frequency for the peak detection was 250 Hz ($\frac{f_s}{80} = \frac{20\,000\,\mathrm{Hz}}{80} = 250\,\mathrm{Hz}$). Figure 2 illustrates the upsampling of the original glottal area time series and the resulting pitch marks.

The individual utterances of the recorded obstruents were segmented as /eCe/ clusters manually centered at the obstruents and with the duration of 375 ms, matching the period of the metronome used for the recording. This resulted in 5 /eCe/ clusters for each of the 12 recorded obstruents.

Each /eCe/ cluster had its glottal area waveform ($f_s = 20\,000\,\mathrm{Hz}$) then split into two components: a DC component, containing the low frequencies of the signal, and an AC component, containing the high frequencies of the signal. These separate components allow the analysis of the slower changes related to voice quality (DC component) and of faster oscillations related to phonation (AC component) individually. The DC component is the result of a low-pass filtering of the glottal area waveform, whereas the AC component is the result of a high-pass filtering of

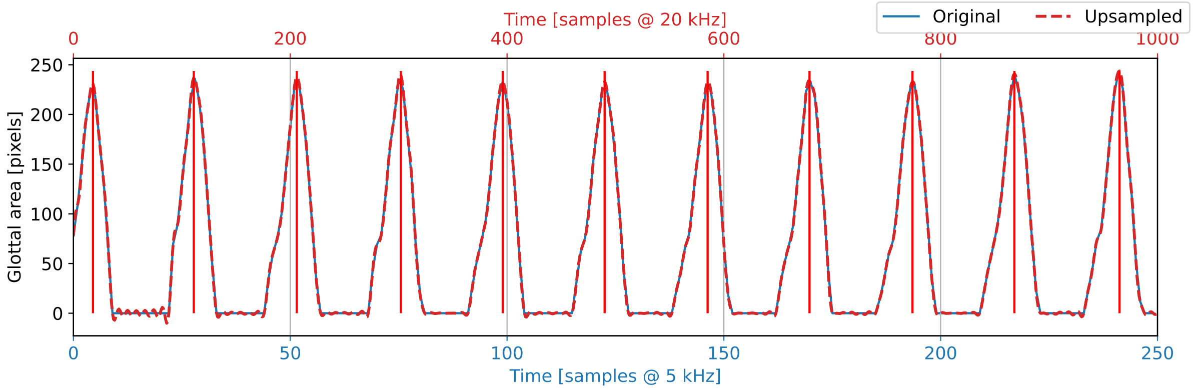


Figure 2 – The original glottal area waveforms ($f_s = 5000\,\text{Hz}$), the upsampled glottal area waveforms ($f_s = 20\,000\,\text{Hz}$) and the pitch marks (vertical red lines). [color online]

the same signal. The used low-pass and high-pass filter were zero-phase filters and had a cut-off frequency of $f_c = 50\,\text{Hz}$, guaranteeing that phonation was only present in the AC component, as the fundamental frequency of the recorded speaker lingered around $f_0 = 200\,\text{Hz}$. The used filters had a finite impulse response (FIR) and were designed using the Kaiser window method with following parameters: 60 dB rejection on the stop band and a 50 Hz-long transition between pass and rejection bands. Figures 3 and 4 show how the DC and AC components relate to the original glottal area waveform for a voiced and for a voiceless obstruent, respectively.

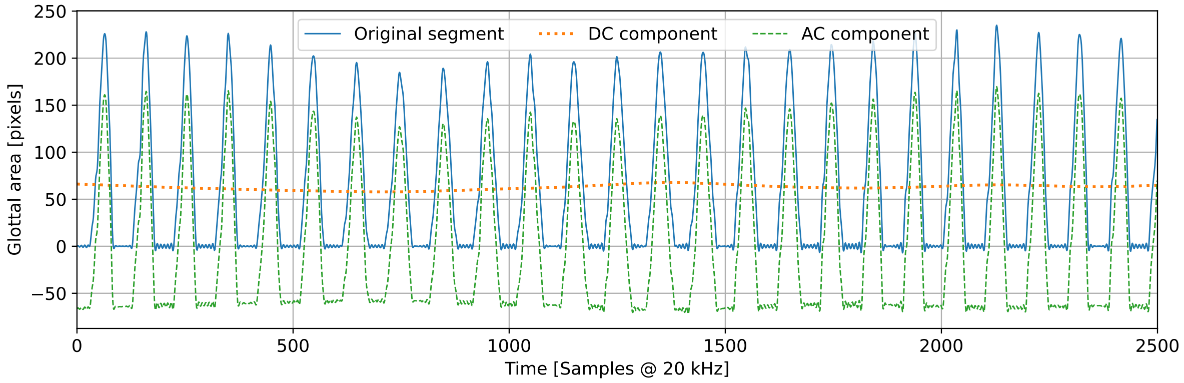


Figure 3 – Original glottal area waveform and its DC and AC components for a /ebe/ segment. [color online]

2.3 Calculation of glottal area waveform parameters

The calculated glottal area waveform parameters were the following: Open Quotient (OQ), Fundamental Frequency (f_0), AC Amplitude (ACA) and DC Amplitude (DCA). The parameters OQ, f_0 and ACA were calculated once for each glottal period (the interval between two pitch marks) as follows: OQ is the ratio of the glottal period in which the glottal area is equal or higher than 50 % of the maximal glottal area of that glottal period; f_0 is the inverse of the duration of the glottal period in seconds; and ACA is the amplitude of the AC component of the glottal period, that is, the difference between the maximal and the minimal glottal area. On the other hand, the DCA parameter is simply the DC component of the glottal area waveform.

The resulting values for OQ, f_0 and ACA, assigned to the middle of each period, were linearly interpolated to the same sampling frequency as the glottal area waveform, 20 000 Hz, since only one value per glottal period was calculated, and the glottal periods were not equally

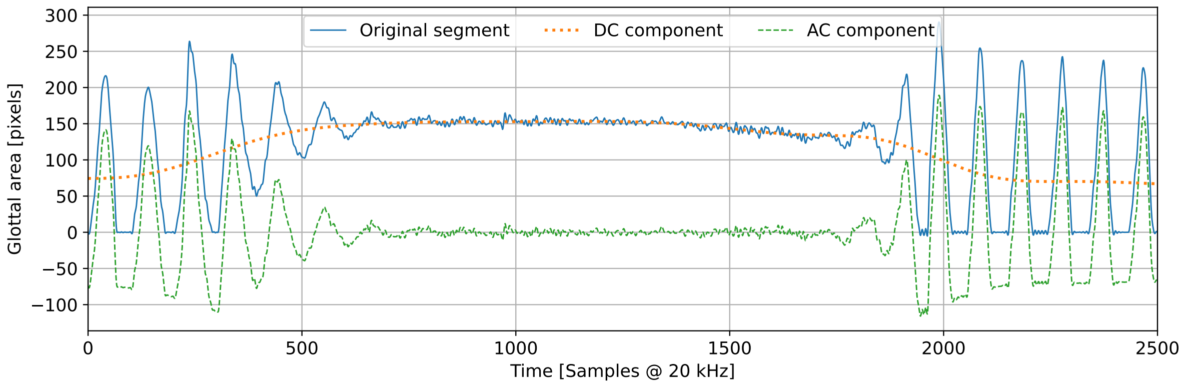


Figure 4 – Original glottal area waveform and its DC and AC components for a /epe/ segment. [color online]

distributed in time. The upsampled OQ, f_0 and ACA signals were then smoothed with a Gaussian filter ($\sigma = 150$ samples) to remove noise caused by suboptimal glottal area segmentation by the software. This was a problem because the lighting and the contrast in the high-speed laryngoscopy images were low and the detection of the actual glottis area was subject to noise.

2.4 Signal processing algorithms

The signal processing algorithms used in the methodology are all open source and available in the scipy library [4]. The upsampling of the glottal area waveform from 5000 Hz to 20000 Hz was done with *scipy.signal.resample*; the peaks of the glottal area waveform were determined with *scipy.signal.findpeaks*; the low-pass and high-pass filters used to split the glottal area waveform into its DC and AC components were designed using *scipy.signal.kaiserord*, to determine the necessary order of the filter, and *scipy.signal.firwin*, to generate the filter coefficients, and applied using *scipy.signal.lfilter*; and the glottal area waveform parameters were smoothed using *scipy.ndimage.gaussian_filter*.

3 Results and discussion

This sections presents the glottal area waveform parameters resulting from the data processing and describes how their synchrony was investigated. All glottal area waveforms and the extracted parameter time functions are available as CSV files in the supplemental material at https://www.vocaltractlab.de/index.php?page=birkholz-supplements.

3.1 Synchrony in voiced obstruents

Figure 5 presents the investigated glottal area waveform parameters for each voiced obstruent. The parameters OQ, f_0, ACA and DCA varied consistently during the production of voiced obstruents: OQ and DCA increased in synchrony, whereas f_0 and ACA decreased in synchrony. Signs of synchrony in these variations are corroborated by Pearson's correlation coefficients between these 4 parameters in Table 1, as significant correlation was found for all pairs of parameters for all phonemes ($\alpha = 0.01$).

The parameters OQ and DCA generally increased during the obstruent, whereas parameters f_0 and ACA generally decreased. The strength of their correlation varies depending on the phonemes: plosives showed stronger correlation between f_0 and ACA, and fricatives showed stronger correlation between OQ and DCA. Beyond that, OQ showed strong negative correlation to f_0 and ACA for the majority of the voiced obstruents, and the correlation between ACA and DCA was stronger for fricatives than for plosives.

A physiological background for these relations might be that higher values of OQ result from a greater abduction of the glottis, i.e., higher DCA and lower ACA. The f_0 decrease also happens during the production of voiced obstruents due to consonant-related f_0 perturbations ($CF0$) [5], which makes f_0 positively correlated to ACA. However, the f_0 decreases seemed to be less intense for fricatives than for plosives, as illustrated by the strong correlation values between f_0 and ACA only for plosives. Additionally, f_0 was consistently lower in fricatives than in plosives.

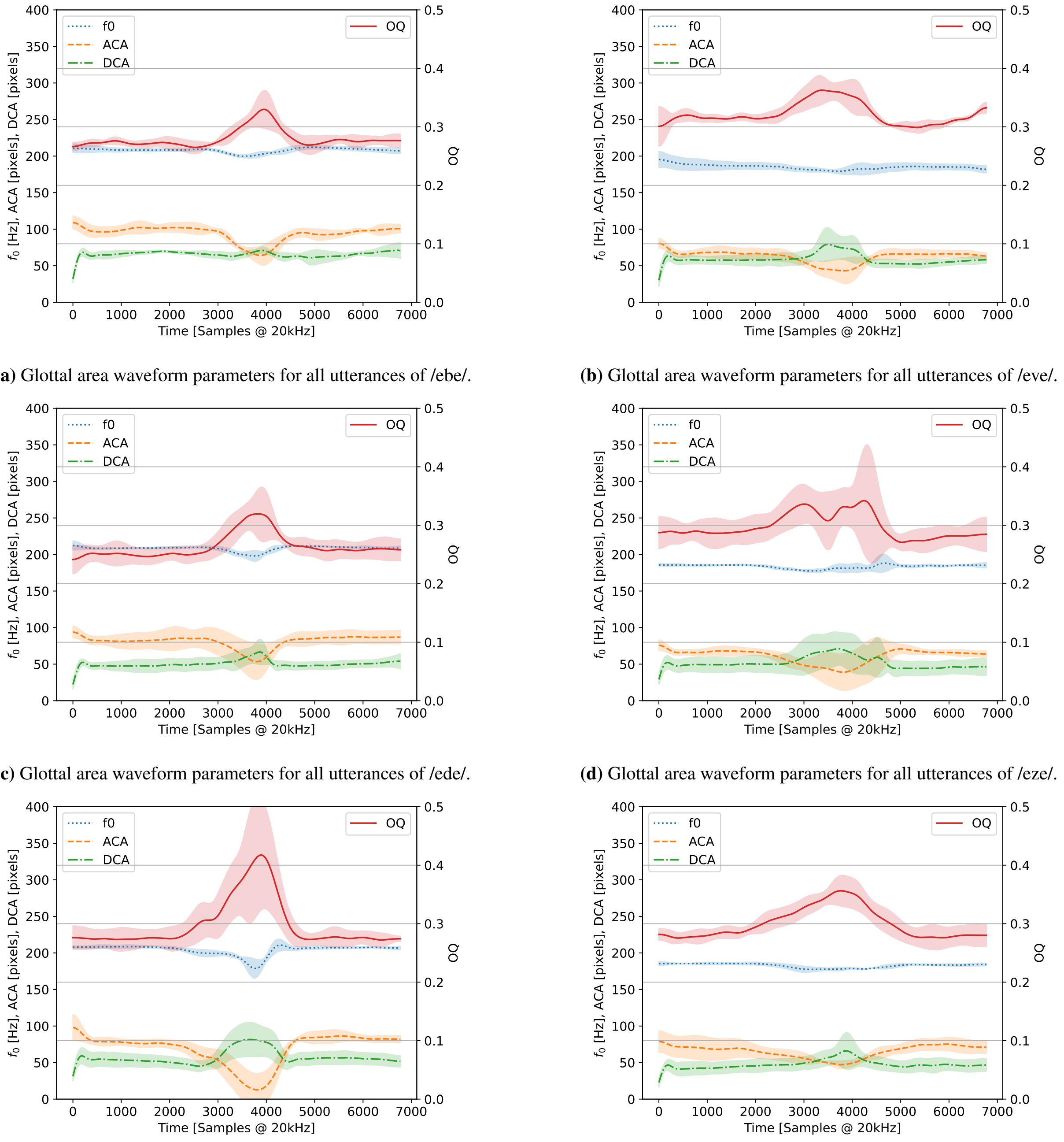
Table 1 – Mean and standard deviation of Pearson's correlation coefficient between the glottal area parameters. The correlation coefficients were calculated based on 150 ms long segments centered at the obstruent, to consider variations at vowel offset, obstruent and vowel onset (voiceless periods were not considered). All coefficients are significant with $p < 0.01$ and 2998 degrees of freedom. Strong correlations, where at least 3 of the 5 utterances showed absolute values higher or equal to 0.7, are written in bold.

Phoneme	OQ × f_0	OQ × ACA	OQ × DCA	f_0 × ACA	f_0 × DCA	ACA × DCA
/b/	**−0.59 ± 0.42**	**−0.95 ± 0.02**	0.34 ± 0.39	**0.73 ± 0.36**	0.12 ± 0.39	−0.21 ± 0.45
/d/	**−0.73 ± 0.25**	**−0.87 ± 0.16**	0.41 ± 0.59	**0.79 ± 0.24**	−0.10 ± 0.33	−0.17 ± 0.59
/g/	−0.61 ± 0.22	**−0.90 ± 0.12**	**0.57 ± 0.61**	**0.74 ± 0.12**	−0.33 ± 0.59	**−0.61 ± 0.74**
/v/	**−0.51 ± 0.39**	**−0.92 ± 0.06**	**0.77 ± 0.09**	0.42 ± 0.44	−0.49 ± 0.29	**−0.78 ± 0.20**
/z/	−0.07 ± 0.37	**−0.62 ± 0.42**	**0.60 ± 0.43**	0.03 ± 0.52	0.05 ± 0.22	**−0.52 ± 0.48**
/ʒ/	**−0.59 ± 0.41**	**−0.93 ± 0.06**	**0.84 ± 0.11**	0.54 ± 0.23	−0.50 ± 0.43	**−0.74 ± 0.16**
/p/	−0.31 ± 0.53	**−0.79 ± 0.13**	**0.67 ± 0.28**	0.47 ± 0.51	−0.41 ± 0.48	**−0.75 ± 0.23**
/t/	−0.57 ± 0.13	**−0.85 ± 0.05**	**0.71 ± 0.13**	0.63 ± 0.13	**−0.61 ± 0.17**	**−0.76 ± 0.14**
/k/	**−0.67 ± 0.09**	**−0.84 ± 0.12**	**0.80 ± 0.19**	**0.74 ± 0.07**	−0.50 ± 0.27	−0.53 ± 0.33
/f/	−0.42 ± 0.26	−0.64 ± 0.25	**0.81 ± 0.07**	0.29 ± 0.37	−0.58 ± 0.23	**−0.67 ± 0.16**
/s/	−0.28 ± 0.24	**−0.82 ± 0.10**	**0.69 ± 0.07**	0.29 ± 0.32	−0.44 ± 0.23	**−0.78 ± 0.06**
/ʃ/	−0.60 ± 0.11	**−0.86 ± 0.07**	**0.74 ± 0.14**	**0.73 ± 0.12**	**−0.71 ± 0.12**	**−0.74 ± 0.09**

3.2 Synchrony in voiceless obstruents

Figure 6 presents the investigated glottal area waveform parameters for all utterances of each voiceless obstruent. Similarly to the voiced obstruents, significant correlation was found between all glottal area waveform parameters, as shown in Table 1. OQ, f_0, and ACA behaved consistently for all voiceless obstruents.

However, the parameter DCA reached a higher amplitude at the beginning of fricatives compared to plosives. This means that the degree of glottal abduction was overall stronger for voiceless fricatives than for voiceless plosives. For the fricatives, we also observe an asymmetry of DCA around the voiceless portion: DCA is high at the fricative onset and low at the fricative offset. This is a hysteresis effect of vocal fold oscillation: The oscillation continues for a while during glottal abduction at the onset of the fricative, but is re-initiated after the fricative only as soon as the vocal folds are sufficiently adducted.



(a) Glottal area waveform parameters for all utterances of /ebe/.

(b) Glottal area waveform parameters for all utterances of /eve/.

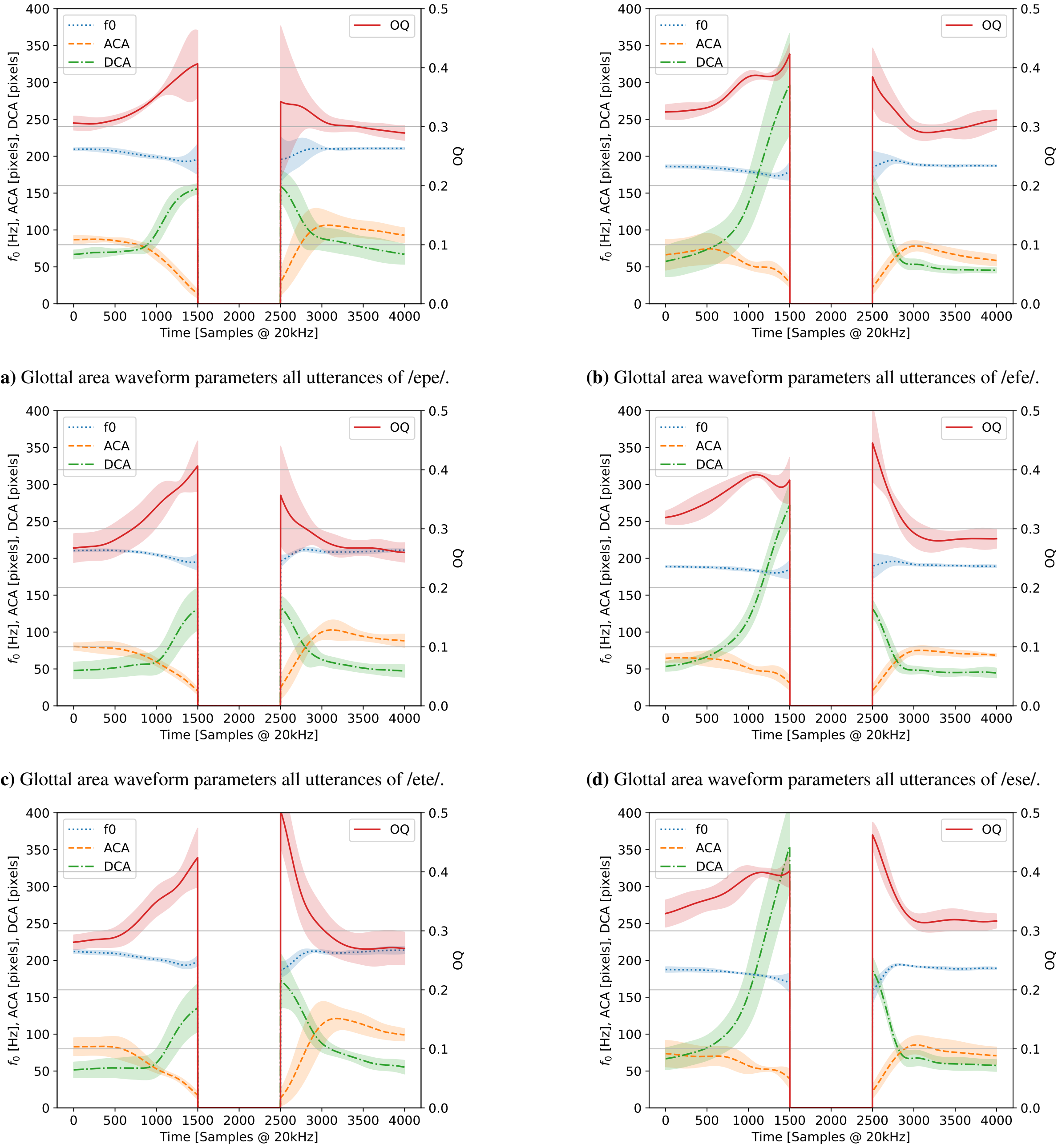
(c) Glottal area waveform parameters for all utterances of /ede/.

(d) Glottal area waveform parameters for all utterances of /eze/.

(e) Glottal area waveform parameters for all utterances of /ege/.

(f) Glottal area waveform parameters for all utterances of /eʒe/.

Figure 5 – Mean and standard deviation of glottal area waveform parameters OQ, f_0, ACA and DCA for all utterances of each recorded voiced obstruent. The values of parameters f_0, ACA and DCA are shown on the y-axis on the left of each plot, whereas the values of parameter OQ are shown on the y-axis on the right of each plot. [color online]



(a) Glottal area waveform parameters all utterances of /epe/.

(b) Glottal area waveform parameters all utterances of /efe/.

(c) Glottal area waveform parameters all utterances of /ete/.

(d) Glottal area waveform parameters all utterances of /ese/.

(e) Glottal area waveform parameters all utterances of /eke/.

(f) Glottal area waveform parameters all utterances of /eʃe/.

Figure 6 – Glottal area waveform parameters OQ, f_0, ACA and DCA for all utterances of each recorded voiceless obstruent. The values of parameters f_0, ACA and DCA are shown on the y-axis on the left of each plot, whereas the values of parameter OQ are shown on the y-axis on the right of each plot. All utterances were aligned by vowel offset and vowel onset. The central portion of each graph, where the value of all parameters are set to zero, corresponds to the voicelss obstruent. [color online]

4 Conclusions

It is important to highlight that there are not, to our knowledge, any studies of this nature using the same type of data used here, high-speed laryngoscopy images directly from the glottis. Results from this study, however, appear to be consistent with [1, 2] when suggesting higher values of OQ, i.e. breathier voice, at vowel offset and onset around voiceless obstruents. The well known phenomenom of consonant-related f_0 perturbations ($CF0$) [5] were also present in this study's results.

The most important results of this study are the correlations, which were *significant* for all pairs of glottal area waveform parameters and *strong* in many cases, e.g. for OQ × ACA, OQ × DCA and ACA × DCA. Correlations between pairs of glottal area waveform parameters were also either consistently positive or negative for all investigated phonemes (except for $f_0 \times$ DCA in voiced obstruents), suggesting similar physiological coupling of these parameters.

In future studies, sonorant consonants like /m/ should also be recorded to be used as a baseline, since the laryngeal behavior for their articulation is much more similar to vowels. The main application of the patterns found in this study is the enhancement of articulatory speech synthesis models [6, 7]. The observed effects could be used to improve glottal models and, possibly, improve synthesis quality. Finally, future high-speed laryngoscopy recordings with more speakers and crisper images should be carried out in the future to corroborate the results of this pilot study.

References

[1] LÖFQVIST, A. and R. S. MCGOWAN: *Influence of consonantal environment on voice source aerodynamics*. *Journal of Phonetics*, 20(1), pp. 93–110, 1992.

[2] LÖFQVIST, A., L. L. KOENIG, and R. S. MCGOWAN: *Vocal tract aerodynamics in /aca/ utterances: Measurements*. *Speech Communication*, 16(1), pp. 49–66, 1995.

[3] BIRKHOLZ, P.: *GlottalImageExplorer - an open source tool for glottis segmentation in endoscopic high-speed videos of the vocal folds*. In *Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*, Studientexte zur Sprachkommunikation, pp. 39 – 44. TUDPress, Dresden, Germany, 2016.

[4] VIRTANEN, P. ET AL.: *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. *Nature Methods*, 17, pp. 261–272, 2020.

[5] KIRBY, J. P. and D. R. LADD: *Effects of obstruent voicing on vowel f0: Evidence from "true voicing" languages*. *The Journal of the Acoustical Society of America*, 140 4, pp. 2400–2411, 2016. URL `https://api.semanticscholar.org/CorpusID:217177924`.

[6] BIRKHOLZ, P. and S. DRECHSEL: *Effects of the piriform fossae, transvelar acoustic coupling, and laryngeal wall vibration on the naturalness of articulatory speech synthesis*. *Speech Communication*, 132, pp. 96–105, 2021.

[7] KRUG, P. K., S. STONE, and P. BIRKHOLZ: *Intelligibility and naturalness of articulatory synthesis with VocalTractLab compared to established speech synthesis technologies*. In *Proc. 11th ISCA Speech Synthesis Workshop (SSW 11)*, pp. 102–107. 2021.