A new method for modelling electroglottographic data

Stefano Coretta

1   Introduction

The position of the vocal folds within the oral tract makes their investigation difficult. Direct observation of the activity of the larynx via invasive methods, like laryngeoscopy, brings with it a series of practical drawbacks. Among other methods for obtaining information on glottal activity is electroglottography. Electroglottography, or EGG (Fabre 1957), is a technique that measures the degree of contact between the vocal folds (the Vocal Folds Contact Area, VFCA). A high frequency low voltage electrical current is sent through two electrodes which are in contact with the surface of the neck, one on each side of the thyroid cartilage (Figure 1). Impedance of this current is modulated by the VFCA, and greater vocal folds contact creates less impedance. The amplitude is inversely correlated with VFCA and impedance, so that higher amplitude values indicate a greater contact area (Titze 1990). The EGG unit registers the current impedance and converts it to relative amplitude values. The time-developing amplitude signal thus provides us with information on the changes in VFCA, i.e. on properties of vocal folds vibration (voicing).

A glottal cycle can be described in terms of two phases (Childers & Krishnamurthy 1985; Hampala et al. 2016): (a) a contacting phase, in which the vocal folds are approaching each other, (b) a de-contacting phase, in which the vocal folds move apart from each other. Transversal to this two-phase description, the glottal cycle can be described in terms of whether the glottis is closed or not. According to this classification, the cycle can be divided into (1) a closed phase, in which the glottis is completely closed and glottal flow is 0 (in some contexts this phase could be absent, like in breathy voicing), and (2) an open phase, in which there is no complete contact between the vocal folds. The timing of these phases can be approximated from the EGG signal, as demonstrated by both experimental and modelling work (Hampala et al. 2016). An example EGG signal is provided in Figure 2.

Two important landmarks of glottal movement are the closing instant (the timepoint of glottal complete closure) and the opening instant (the moment in which the glottis first opens). These points delimit the open and closed phases of a glottal cycle. The ratio of the closed phase relative to the total cycle duration, the closed quotient, has been used as an index of phonation type. Modal voice has higher close quotient values than breathy voice, and lower values than creaky voice. One method for the detection of the closing and opening instants is based on the first derivative of the EGG signal (the dEGG, see Figure 2). Herbst et al. (2017), however, showed that this method returns values that are just a surrogate of the actual articulatory movements, due to the complex anatomy of the vocal folds, and that there are no clear contacting and decontact instants, but rather intervals. Herbst et al. (2017) call this EGG-based closed quotient the 'contact quotient' and recommend to keep it distinct from the closed quotient obtained from direct observation of the vocal folds.

As an alternative to the contact quotient, Herbst et al. (2010) propose the wavegram, a visualisation method which does not reduce the EGG signal to a single value and thus suffers less from the limitations of the contact quotient. The wavegram exploits information from the whole signal to obtain an image of vocal folds activity. A wavegram is a 3D representation of the EGG signal developing in time. Its structure is similar to that of a classical phonetic spectrogram. The x-axis indicates the temporal sequence of individual glottal cycles. The y-axis represents the
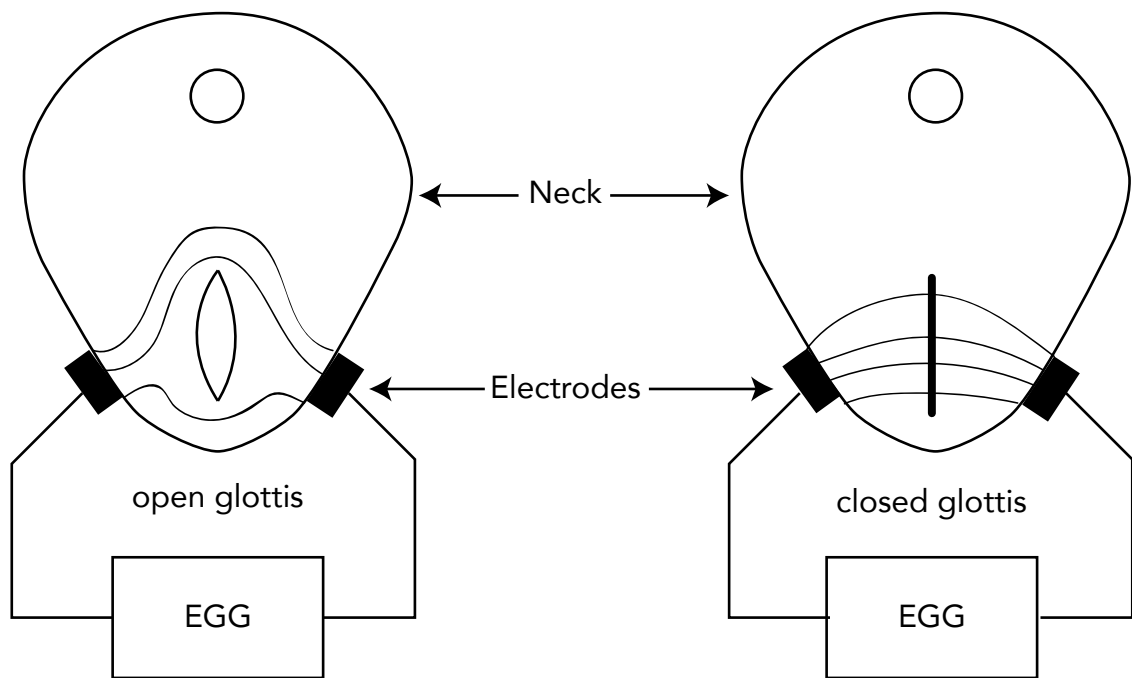
Figure 1: A schematics of the electroglottograph. A transverse section of the neck is shown with open glottis (on the left) and closed glottis (on the right). The electric field passing through the neck is represented by lines. When the vocal folds are apart, the opening distorts the electric field and impedance increases.
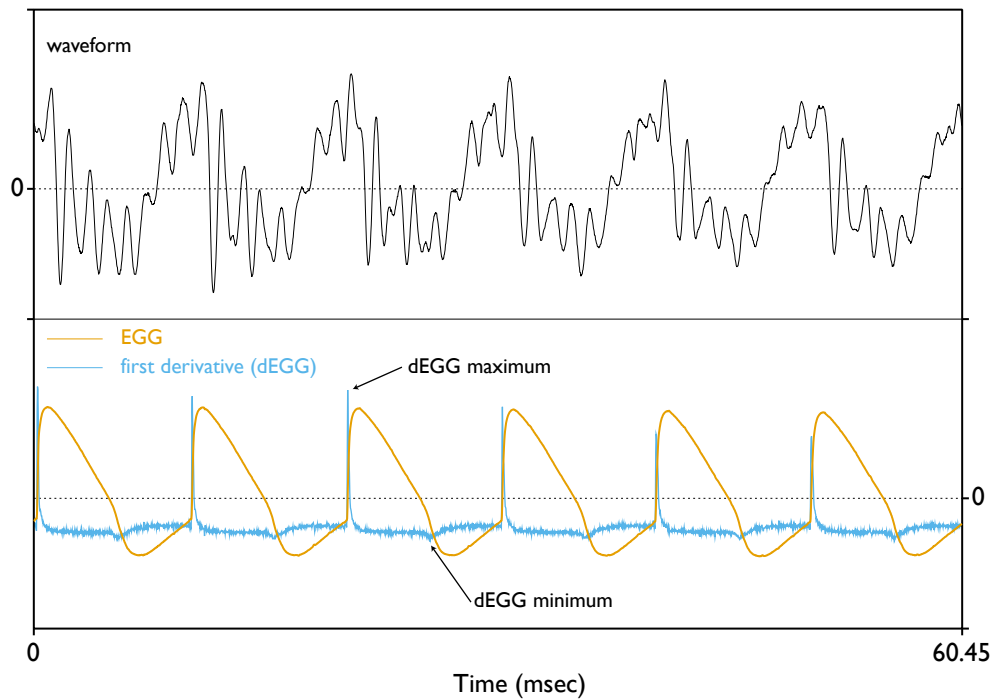

Figure 2: The electroglottographic signal (EGG) with corresponding first derivative (dEGG).

time within each glottal cycle, normalised between 0 and 1. Finally, the normalised amplitude of the signal corresponds to different colour intensities. Differences in intensity along the *x*-axis indicate changes in glottal activity. The procedure for constructing a wavegram is given in Figure 3. A wavegram can be produced for the EGG signal and for any of its transformations, like the dEGG.

A possible limitation of the wavegram method is that it is intended for a qualitative analysis based on visual inspection. However, wavegram data can be modelled using generalised additive models (GAMs, Hastie & Tibshirani 1986; Zuur 2012; Wood 2017). GAMs are a family of generalised models which can fit non-linear effects by additive combinations of smoothing splines. The flexibility of GAMs allows researchers to generate a fitted wavegram based on data from multiple repetitions of a single speaker and from multiple speakers. Random effects can also be included to account for idiosyncratic differences. Moreover, the potential for overfitting is reduced by a smoothing penalty parameter, which constraints the maximum number of basis functions used to construct the smoothing splines. This paper introduces wavegram GAMs as a way to quantitatively assess wavegram data. First, results from a pilot study which informally evaluates the proposed method are presented (Section 2). Section 3 illustrates how to conduct a wavegram GAM analysis of dEGG data through a practical example in which the wavegrams of vowels followed by voiceless and voiced stops are compared. Finally, **??** discusses limitations of the current implementation of the method and future directions.

2 Pilot study

Synchronised audio and EGG data were obtained from 5 trained phoneticians, who were asked to produce sustained tokens of /a/ with modal and breathy voice. The data was collected using a Glottal Enterprises EG2-PCX2 electroglottograph and a Movo LV4-O2 Lavalier microphone, at a sample rate of 44100 Hz (16-bit). The acquisition of the signals was controlled with Audacity running on a MacBook Pro (Retina, 13-inch, Mid 2014). The placement of the electrodes strap was checked with the height indicator on the EGG unit. Each participant uttered 10 consecutive tokens of a sustained /a/ in modal voice, followed by 10 tokens of a sustained breathy /a/. All subsequent data processing was performed in Praat (Boersma & Weenink 2018). The onset and offset of each token were detected with an automatic procedure which finds voiced and unvoiced intervals (`To TextGrid (vuv)`). The dEGG wavegram data was extracted from the first 8 glottal cycles of a 500 ms window, centred around the mid-point of each token. A glottal cycle was arbitrarily defined as the interval between two consecutive EGG minima (cf. with Herbst et al. 2010 for an alternative algorythm). From each glottal cycle, the relative amplitude of the dEGG signal was extracted every 10 samples.

A generalised additive mixed model (GAMM) was fitted to the data to statistically assess differences in vocal fold activity between modal and breathy voicing (see Sóskuthy 2017 and Wieling 2018 for a practical introduction to fitting GAMs in R). The following terms were included: the amplitude of the dEGG signal as the outcome variable, an interaction factor with language and phonation as a parametric term, a smooth over the normalised time of the beginning of the glottal cycle (as the proportion of the time relative to the entire token duration) and one over the normalised time of the sample within the glottal cycle (as the proportion of the time relative to the duration of the glottal cycle), two difference smooths over normalised time of the glottal cycle onset and normalised sample time using a by-variable with the language/phonation factor, a tensor product interaction between normalised cycle time and normalised sample time, and the same tensor product interaction with a language/phonation by-variable. Finally, inter-speaker differences were modelled with a by-speaker factor smooth over normalised cycle time. A first-order autoregressive (AR1) model was included to deal with the relatively high auto-correlation in the residuals.

Figure 4 shows the modelled wavegrams of modal and breathy tokens. Since the tokens
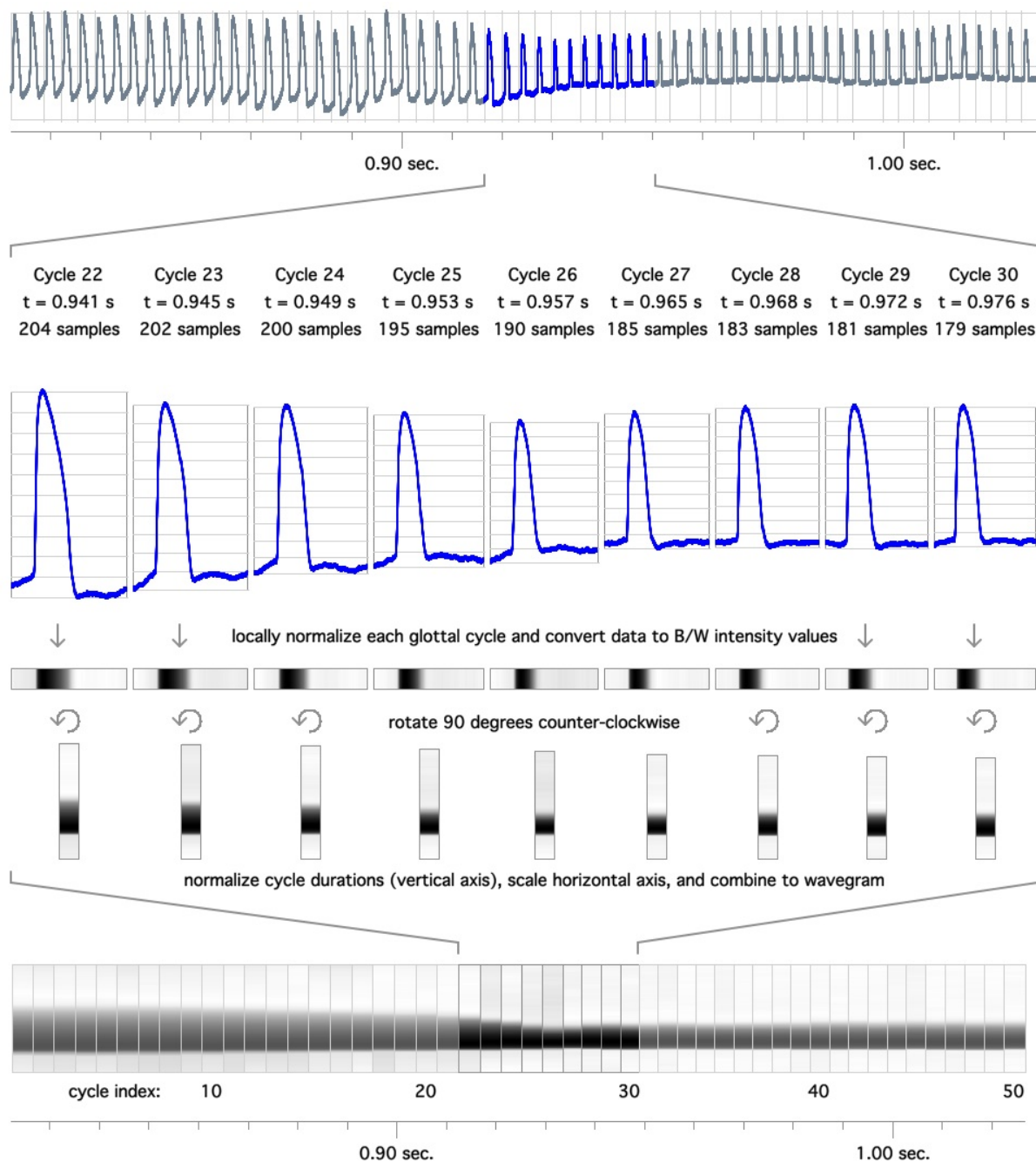
Figure 3: The wavegram. Created by Christian T. Herbst under a CC BY-SA 3.0 license.
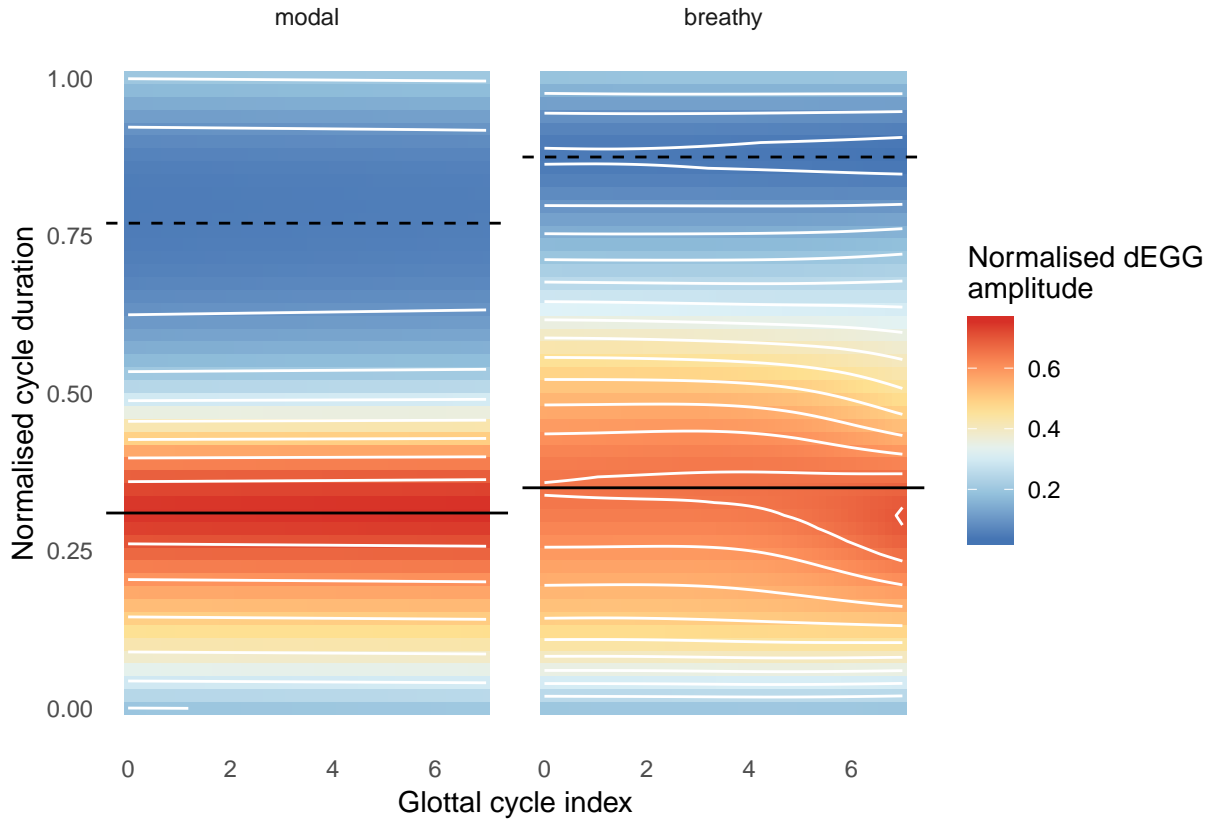
Figure 4: Fitted wavegram of modal and breathy phonation (Section 2). The horizontal lines represent the dEGG maximum (solid line) and minimum (dashed line).

were produced with sustained phonation, no appreciable change within each wavegram can be observed. However, the comparison of the wavegram of modal voice with that of breathy voice reveals differences between the two phonation types. As a general trend, the dEGG maximum and dEGG minimum are achieved later within the glottal cycle in breathy voicing relative to modal voicing. Moreover, differences in velocity of closing and opening movements of the vocal folds are signalled by the relative widths of the red-coloured bands (around the dEGG maximum) and the blue-coloured bands (around the dEGG minimum). While in modal voicing the blue band is wider, the red band is in breathy voicing, indicating that the velocity into and out of the beginning of the closed phase is slower in breathy voicing. According to the approximate significance of the smooth terms, phonation has an effect on the shape of the wavegram as expected (F(14.681) = 3.187, Ref.EDF = 19.027, $p < 0.001$).

3 Wavegram of vowels followed by voiceless vs. voiced stops

This section further illustrates the use of wavegram GAMs by discussing a dynamic analysis of changes in vocal folds activity during the production of vowels followed by voiceless vs. voiced stops in Italian and Polish. EGG data was obtained from 9 Italian speakers and 6 Polish speakers. The procedures for data processing and analysis were the same as with the pilot study, with the exception that data was extracted from every glottal cycle within the whole duration of the vowel. The vowel onset and offset were identified as the appearance and disappearance of higher formant structure respectively (Machač & Skarnitzl 2009).

Figure 5 shows the modelled wavegrams of vowels followed by voiceless (left) and voiced stops (right), in Italian (top) and Polish (bottom). The wavegrams indicate that vocal folds activity changes towards the end of the vowel if the following consonant is voiceless, both in Italian and
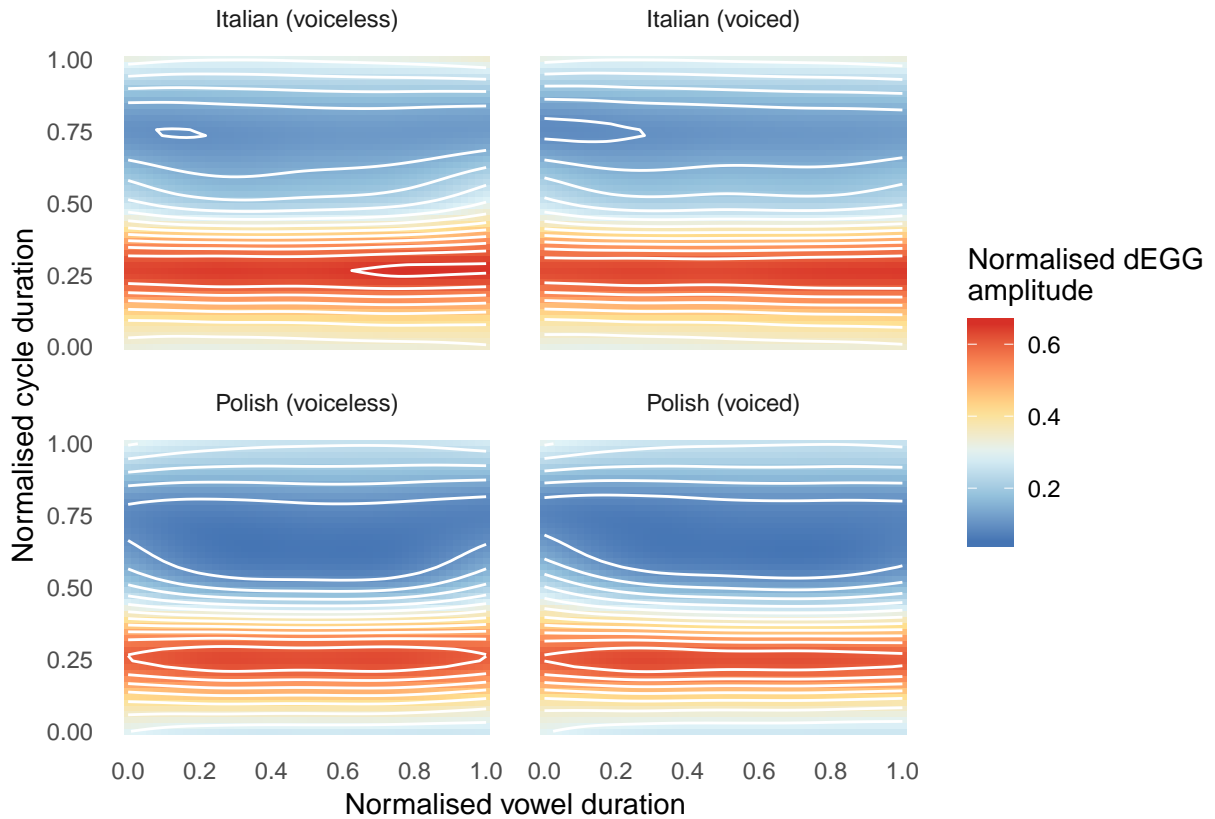
Figure 5: Fitted wavegram of vowels followed by voiceless and voiced stops in Italian and Polish (Section 3).

Polish. The glottal activity along the second half of vowels followed by voiced stops is, on the other hand, more stable. Moreover, the change in activity is initiated earlier in Italian (at around 60% into the vowel) than in Polish (at about 80%). What the wavegrams show is that glottal opening increases towards the end of the vowel, in anticipation of the open glottis typical of voiceless stops (and consonants more in general).

Another difference between the two languages concerns the initial voiceless stop (/p/). The change observed between 50 and 70% into the glottal cycle at the beginning of the vowel, which indicates a transition from a more spread glottis, is more extreme in Polish than in Italian. This means that phonation at vowel onset is slightly breathier in the former. Such difference is compatible with the finding that Polish voiceless stops have on average a longer VOT than Italian stops.

The observation of increased glottal opening in the production of vowels followed by voiceless stops in Italian is further compatible with the reported presence of pre-aspiration (breathy or voiceless) in Italian geminate stops (Stevens & Hajek 2004b,a, 2010; Stevens 2010; Stevens & Reubold 2014). Increased glottal spreading during vocal fold vibration is the logical precursor of voiceless pre-aspiration. The fact that voiceless pre-aspiration is not fully developed in Italian singleton voiceless stops offers interesting insights concerning the duration of voiceless stops and the emergence of pre-aspiration.

It is known that voiceless stops have longer closure durations than voiced stops. Lisker (1974) argues that, in English voiceless stops, closure occurs not long after the spreading gesture is initiated in order to avoid the emergence of full pre-aspiration. There are, however, varieties of English with pre-aspiration in stops and fricatives (Gordeeva & Scobbie 2007; Nance & Stuart-Smith 2013; Hejná 2015). The question arises as to how glottal spreading in vowels in the context of voiceless stops can lead to the emergence of pre-aspiration in some cases. One of the common

conceptions is that pre-aspiration arises when closure duration decreases while glottal spread remains. Ní Chasaide (1985), however, claims that the appearance of pre-aspiration (in the form of glottal spread before stop closure) diachronically precedes closure shortening. Stevens et al. (2014[1986]) further present experimental evidence from Italian that pre-aspiration and closure shorting are independent. Moreover, the presence of pre-aspiration increased the total duration of the VC sequence.

What can be learnt from this is that, once breathyness/pre-aspiration arises, there are two possible scenarios: according to one possible path, as noted by Lisker (1974), pre-aspiration is prevented by achieving stop closure earlier (thus increasing the stop closure duration); according to the other path, pre-aspiration can be enhanced, with subsequent closure shortening, as in the arguments by Ní Chasaide (1985) and Stevens et al. (2014[1986]).

References

Boersma, Paul & David Weenink. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.40. `http://www.praat.org/`.

Childers, Donald G. & Ashok K. Krishnamurthy. 1985. A critical review of electroglottography. *Critical reviews in biomedical engineering* 12(2). 131–161.

Fabre, P. 1957. Un procede electrique percutane d'inscrition de l'accolement glottique au cours de la phonation: glottographie de haute frequence. Premiers resultats. *Bulletin de l'Académie nationale de médecine* 141. 66.

Gordeeva, Olga B. & James M. Scobbie. 2007. Non-normative preaspirated voiceless fricatives in Scottish English: Phonetic and phonological characteristics. *QMU Speech Science Research Centre Working Papers* .

Hampala, Vít, Maxime Garcia, Jan G. Švec, Ronald C. Scherer & Christian T. Herbst. 2016. Relationship between the electroglottographic signal and vocal fold contact area. *Journal of Voice* 30(2). 161–171. doi:10.1016/j.jvoice.2015.03.018.

Hastie, Trevor & Robert Tibshirani. 1986. Generalized additive models. *Statistical Science* 1(3). 297–310.

Hejná, Michaela. 2015. *Pre-aspiration in Welsh English: A case study of Aberystwyth*: The University of Manchester dissertation.

Herbst, Christian T., W. Tecumseh S. Fitch & Jan G. Švec. 2010. Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *The Journal of the Acoustical Society of America* 128(5). 3070–3078. doi:10.1121/1.3493423.

Herbst, Christian T., Harm K. Schutte, Daniel L. Bowling & Jan G. Svec. 2017. Comparing chalk with cheese—the EGG contact quotient is only a limited surrogate of the closed quotient. *Journal of Voice* 31(4). 401–409.

Lisker, Leigh. 1974. On "explaining" vowel duration variation. In *Proceedings of the Linguistic Society of America*, 225–232.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*. Epocha.

Nance, Claire & Jane Stuart-Smith. 2013. Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. *Journal of the International Phonetic Association* 43(02). 129–152.

Ní Chasaide, Ailbhe. 1985. *Preaspiration in phonological stop contrasts: an instrumental phonetic study*: University of Wales dissertation.

Sóskuthy, Márton. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. arXiv.org preprint, arXiv:1703.05339.

Stevens, Kenneth N., Samuel Jay Keyser & Haruko Kawasaki. 2014[1986]. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and variability in speech processes*, 426–463. Psychology Press.

Stevens, Mary. 2010. How widespread is preaspiration in Italy? A preliminary acoustic phonetic overview. In *Proceedings of FONETIK 2010*, 97–102. Lund.

Stevens, Mary & John Hajek. 2004a. Comparing voiced and voiceless geminates in Sienese Italian: what role does preaspiration play? In *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 340–345.

Stevens, Mary & John Hajek. 2004b. Preaspiration in Sienese Italian and its interaction with stress in /VC:/ sequences. Paper presented at the International Conference Speech Prosody, Japan, March 23-26.

Stevens, Mary & John Hajek. 2010. Preaspirated /pp tt kk/ in Standard Italian: a sociophonetic v. phonetic analysis. In *Proceedings of the 2010 Speech Science and Technology Association Conference*, Melbourne, Australia.

Stevens, Mary & Ulrich Reubold. 2014. Pre-aspiration, quantity, and sound change. *Laboratory Phonology* 5(4). 455–488.

Titze, Ingo R. 1990. Interpretation of the electroglottographic signal. *Journal of Voice* 4(1). 1–9.

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70. 86–116. doi:10.1016/j.wocn.2018.03.002.

Wood, Simon. 2017. *Generalized additive models: An introduction with R*. Chapman and Hall/CRC 2nd edn.

Zuur, Alain F. 2012. *A beginner's guide to generalized additive models with R*. Highland Statistics Limited: Newburgh.