GLOTTAL WIDTH PATTERNS IN NUMERICALLY SIMULATED DIPLOPHONIA

Jean Schoentgen([1]) and Philipp Aichinger([2])

([1])L.I.S.A, Faculty of Applied Sciences, Université Libre de Bruxelles, Brussels, Belgium

([2])Division of Phoniatrics and Logopedics, Department of Otorhinolaryngology, Medical University of Vienna, Vienna, Austria

jschoent@ulb.ac.be, philipp.aichinger@meduniwien.ac.at

ABSTRACT

The presentation explores diplophonia via numerical simulations of glottal vibrations. The aim of the study is to improve the understanding of glottal wall vibration and area waveform patterns of non-modal phonation as observed via laryngeal high-speed videos. Diplophonia has been described as the simultaneous perception of two pitches during voicing. A broader definition is the vibration of different glottal structures at two different frequencies. Simulations and direct observations suggest that glottal entrance-exit, left-right as well as anterior-posterior frequency asymmetries cause qualitatively distinct glottal pulse patterns. Additional results show that glottal entrance-exit and anterior-posterior phase shifts of single-frequency vibrations may cause double pulsing as well as spectral subharmonics. This suggests that definitions based on pulse counting within glottal metacycles or detecting spectral subharmonics are too broad to enable distinguishing glottal frequency from glottal phase asymmetries.

Keywords: Diplophonia, biphonation, voice quality, glottal area models.

1. INTRODUCTION

Conventionally, voiced speech sounds have been grouped into three categories according to whether their spectra are typified by a single series of harmonics (type I), more than one series of harmonics or partials (type II), or whether the spectra are continuous (type III). Diplophonic, biphonic and, occasionally, creaky voices are examples of type II voices. Type III voices are produced by vocal fold vibrations that are expected to be chaotic or "turbulent" [3].

The distinction between diplophonic and biphonic voices may be formal only. The former are observed when vibrations at distinct frequencies are synchronised and the latter when they are not. However, the difference between n:m synchronised vibrations with n and m large and unsynchronised vibrations is fuzzy in practice.

Diplophonia has been defined in several ways that may not agree perfectly. A definition that is popular in a clinical framework is the simultaneous auditory perception of two pitches during voicing [5][6][12][17]. But diplophonia has also been defined as follows.

- Vibration of both the ventricular folds and the vocal folds, producing two simultaneous voice tones [7]
- Subharmonics that "create an effect that has been perceptually described as diplophonia" [18]
- Quasiperiodic changes in the pitch, amplitude and/or waveform of the speech signal [8][9][10].
- A difference in the vibratory frequency of the left and right vocal folds or anterior-posterior parts of the vocal folds [9]
- Double pulsing [11]
- Desynchronization of the left and right vocal folds [13]
- A period-two up-down pattern of an arbitrary cyclic parameter [15][14]
- Existence of two distinct voice sources [4]
- Observation of spatially distinct glottal structures vibrating at different frequencies [1][6]

The first item in the list together with other extra-glottal sources, as well as turbulent airflows and glottal whistle that may cause two pitches to be heard, are not discussed. The remaining items are assumed to refer to diplophonia described at the glottal, acoustic or perceptual levels. The observation of glottal structures that vibrate at distinct frequencies is clinically relevant, because they are the signatures of severe imbalances in the masses or tensions of the vocal folds.

The observation of distinct vibratory frequencies at the glottis are not a sufficient condition for the perception of two pitches during voicing. When the frequencies are in a simple integer ratio (e.g. 1 : 2, 2 : 3), the overlap of the harmonics that associate with each fundamental increases and the pitches are likely to merge perceptually [2]. When the two frequencies are almost identical, a slow beating or modulation is expected, which may be perceived as roughness or tremulousness. The conditions that would predict the perception of two pitches are not known.

The numerical simulations focus on the glottal wall vibration → glottal width link because complex area patterns (and therefore complex glottal pulse patterns) may arise from simply described frequency or phase asymmetries of the vibrations of the glottal walls.

Motivations for the simulations are the following. First, case studies of artificially generated type II vocal patterns may contribute to the interpretation of observed glottal diplophonia. One of the aims of the study is indeed to improve the understanding of vocal fold vibration and glottal area waveform patterns of non-modal phonation observed via laryngeal high-speed video [1]. Second, the successful numerical synthesis of diplophonic voices relies on an intuitive understanding and control of type II glottal area patterns. Third, numerical simulations enable examining glottal entrance-glottal exit asymmetries that are difficult to observe in vivo.

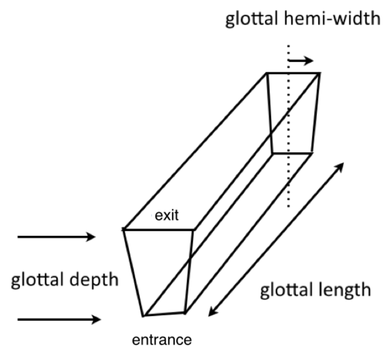
Hereafter, we argue, based on numerical simulations, that definitions relying on criteria such as double pulsing or subharmonics (that are the spectral signature of glottal metacycles) may be too broad because they include glottal phenomena other than genuine diplophonia. Other issues that are addressed are the lengths and complexities of the chains of causality that link glottal area waveforms to anterior-posterior, left-right and entrance-exit glottal frequency asymmetries, suggesting that disagreements between existing definitions may also be a consequence of diplophonia being phenomenologically inhomogeneous.

2. METHODS

2.1. Glottis model

The model of the glottis is a generalisation of the Phase-Delayed-Overlapping-Sinusoidal (PDOS) model [16]. It is a morphological model of the $3D$ glottis that generates glottal area waveform patterns via simple wall collision and $3D \rightarrow 2D$ projection operators. The waveforms so obtained are the effective glottal areas or widths observed from the top. The subsequent computation of the flow rate via a lumped-element model of the airflow in the glottis is not discussed here [16]. Fig. 1 shows the PDOS model geometrically and formulas (1) to (5) show it mathematically.

Figure 1: Rectangular glottal PDOS model.



A_0 designates the abduction hemi-amplitude, A the vibration hemi-amplitude, f the frequency of vibration, Φ the phase delay between glottal entrance and exit and w_h the evolving glottal hemi-width. Subscript h designates the *left* or *right* hemi-glottis.

(1)
$$w_{h,entrance} = A_{0,h,entrance} + A_{h,entrance} sin(2\pi f_{h,entrance})$$

(2) $$w_{h,exit} = A_{0,h,exit} + A_{h,exit} sin(2\pi f_{h,exit} - \Phi)$$

Expressions (3) and (4) are the entrance and exit widths, *max* is the collision operator that zeroes the widths as soon as the glottal walls touch and (5) is the effective width of the glottis.

(3) $$w_{entrance} = max(0, w_{left,entrance} + w_{right,entrance})$$

(4) $$w_{exit} = max(0, w_{left,exit} + w_{right,exit})$$

(5) $$w = min(w_{entrance}, w_{exit})$$

Assuming that the glottis is rectangular, the effective glottal area is equal to $w \times l$, with l equal to the glottal length. In the case of anterior-posterior asymmetries, the glottis is subdivided longitudinally into two separate anterior and posterior glottises, each with its length and width as well as frequency of vibration.

2.2. Simulation output

The simulation output is graphical. Fig. 2 (for instance) summarises the time evolution of a coronal slice of the glottis. The distances with regard to the longitudinal axis (to the left) and glottal widths (to the right) are given in *mm* on the horizontal axis and the time in *sec* on the vertical axis. Each vertical axis is shifted by $5mm$ to ease visibility. From left to right, the curves report the following: the evolution of the glottal left wall (in green) and right wall (in red) at the glottal entrance and exit, i.e. (1) and (2); the glottal entrance and exit widths, i.e. (3) and (4); and the effective glottal width, i.e. (5). The effective glottal width differs from the entrance and exit widths because of the masking of the entrance by the exit, which is simulated by the *min* operator in (5).

In the case of anterior-posterior asymmetries the glottal walls are dropped from the graph. In Fig. 3 (for instance) the three curves in blue and magenta report the areas associated with the anterior and posterior glottises respectively. The rightmost curve is the effective glottal area that determines the airflow rate. The area in mm^2 is given on the horizontal axis relative to the vertical axes that are shifted for better visibility and the time is given on the vertical axis in *sec*.

3. RESULTS

3.1. Double pulsing owing to phase shifts

Double pulsing, i.e. observing the glottis open and close twice within one glottal metacycle is not exceptional. Fig. 2 and Fig. 3 suggest that double pulsing is not necessarily caused by spatially distinct glottal structures that vibrate at different frequencies. In Fig. 2, double pulsing is caused by a large phase shift Φ between glottal entrance and exit as well as a large open quotient. In Fig. 3 it is caused by an out of phase motion of the anterior and posterior glottises.

Figure 2: Double pulsing owing to a large glottal entrance versus exit phase shift (by $3\pi/4$) and large open quotient. Axes and curve labels are explained in section 2.2.

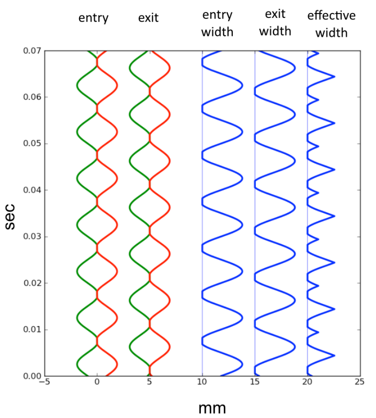
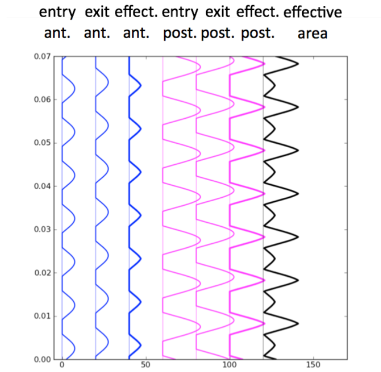


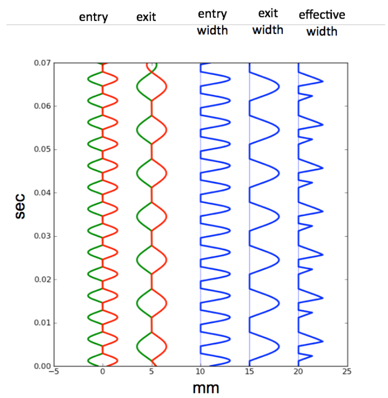
Figure 3: Double pulsing owing to phase shifted (by π) vibrations of anterior and posterior part-glottises. Axes and curve labels are explained in section 2.2.



3.2. Spatially distinct glottal structures vibrating at different frequencies

Figs 4, 5 and 6 report glottal width or area patterns that are caused by entrance and exit, right and left as well as anterior and posterior vibrations occurring at different frequencies f_1 and f_2. Wall vibration patterns in Figs 4 and 5 differ qualitatively even though the frequencies are in the simplest ratio possible, i.e. $f_1/f_2 = 2:1$, and the glottal model is very plain. Possible causes are discussed hereafter.

Figure 4: Glottal entrance-exit frequency asymmetry (2 : 1). Axes and curve labels are explained in section 2.2.



4. DISCUSSION AND CONCLUSION

Comparing Figs 2 and 3 to Figs 4 to 6 suggests that the observation of meta-cycles does not always predict the existence of genuine glottal diplophonia. Double pulsing in the glottal area or speech waveforms may indeed be caused by (substantial) phase

Figure 5: Glottal left-right frequency asymmetry (1 : 2). Axes and curve labels are explained in section 2.2.

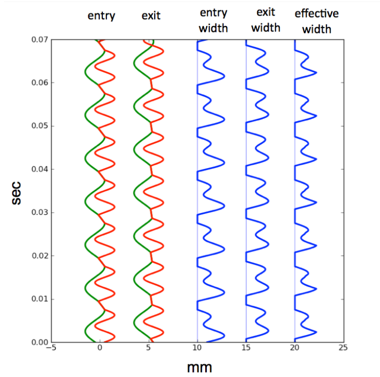
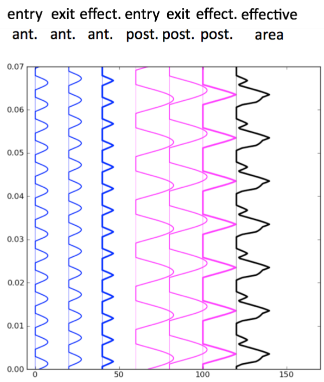


Figure 6: Glottal anterior-posterior frequency asymmetry (2 : 1). Axes and curve labels are explained in section 2.2.



shifts between spatially distinct glottal structures vibrating at the same frequency, a conjecture for which empirical evidence is pending. For instance, entrance-exit phase shifts are typical of modal voice. Double pulsing may occur when the folds are thick and slack because then the tissue wave takes longer to propagate from the entrance to the exit favouring thus large phase shifts. Double pulsing has been associated by some authors with creaky voice or glottal fry [13].

The existence of multiple definitions that overlap, but which do not concur perfectly, may be explained by the fact that diplophonia is under-researched and that it cannot always be auditorily distinguished from hoarseness or roughness or creakiness. A third possibility that is briefly discussed hereafter is that diplophonia might not be homogeneous as a phenomenon. A reason to discuss entrance-exit, left-right and anterior-posterior asymmetries separately is that they may have different aetiologies. A major contrast is the one between entrance-exit and left-right asymmetries on the one hand and anterior-posterior asymmetries on the other.

Reasons for distinguishing the former from the latter are several. First, the longitudinal split of the glottis may be the consequence of a growth that immobilises or dampens parts of the vocal folds, which continue vibrating anteriorly and posteriorly of the growth at frequencies and amplitudes that are different [9]. The split of the glottis into quasi-separate anterior and posterior glottises increases the number of degrees of freedom compared to the left-right and entrance-exit asymmetry cases. Second, the vibrating anterior and posterior glottises may function as additive pseudo-independent sources.

Examining expressions (1) to (5) suggests that simple additivity is a property of the anterior-posterior split only. This would suggest that the anterior-posterior case has properties (i.e. additivity, additional degrees of freedom) which it does not share in general with the left-right and entrance-exit cases. Indeed, to achieve quasi-additivity in the left-right case, the vocal folds must not touch and the glottal entrance and exit must vibrate in phase, thus dropping the nonlinear *max* and *min* operators from (3) to (5).

Examining Figs 4 and 5 shows that the glottal patterns are qualitatively distinct. The explanation is in expressions (1) and (2) that involve sums of sinusoids. Sums of sinusoids are themselves sinusoids only when the summands have the same frequency. If not, the sum is a complex signal that is aperiodic in general. The *max* and *min* operators in (3) to (5) therefore operate on a complex signal in the left-right and on a sinusoidal signal in the entrance-exit asymmetry case, thus explaining the qualitatively more complex glottal vibratory patterns in the left-right frequency asymmetry case.

To conclude, a condition for diplophonia to exist appears to be that different parts of the glottis are observed to vibrate at different frequencies. Different types of frequency asymmetry cause glottal area waveform patterns (and therefore speech cycle patterns) that are qualitatively different. The visual, acoustic or auditory detection of distinct frequencies is therefore not equally easy in entrance-exit, left-right and anterior-posterior asymmetry cases, which may have different aetiologies.

5. REFERENCES

[1] Aichinger, P. 2015. *Diplophonic voice, definitions, models and detection*. Graz: Graz University of Technology. PhD thesis.

[2] Aichinger, P., Sontacchi, A., Schneider-Stickler, B. 2011. Describing the transparency of mixdowns: The masked-to-unmasked-ratio. *130th Audio Engineering Society Convention* London.

[3] Behrman, A., Agresti, C., Blumstein, E., Lee, N. 1998. Microphone and EGG data from dysphonic patients : type 1, 2 and 3 signals. *J. Voice* 12, 249–260.

[4] Boone, D., McFarlane, S. 1994. *The voice and voice therapy*. Englewood Cliffs: Prentice Hall. p. 64.

[5] Colton, R., Casper, J., Leonard, R. 2006. *Understanding voice problems*. Philadelphia: Lippincott Williams and Wilkins. p. 20.

[6] Dejonckere, P., Lebacq, 1983. An analysis of the diplophonia phenomenon. *Speech Communication* 2, 47–56.

[7] Diplophonia. http://medical-dictionary.thefreedictionary.com. [Accessed: January 30, 2015].

[8] Kiritani, S., Hirose, H., Imagawa, H. 1991. Vocal fold vibration and the speech waveform in diplophonia. *ANN. BULL. RILP* 25, 55–62.

[9] Kiritani, S., Imagawa, H., Myiazi, M., Kumada, M., Niimi, S. 1995. Quasi-periodic perturbations in pathological vocal fold vibration. *ANN. BULL. RILP* 29, 27–33.

[10] Kiritani, S., Niimi, S., Imagawa, H., Hirose, H. 1995. Vocal fold vibrations with involuntary voice changes in certain pathological cases. In: Fujimura, O., Hirano, M., (eds), *Vocal fold physiology: voice quality control*. San Diego: Singular Publishing Group, Inc. 269–281.

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

[12] Rossbach, M. J. 1872. Doppeltönigkeit der Stimme (Diphthongie) bei ungleicher Spannung der Stimmbänder. *Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* 54, 571–574.

[13] Svec, J. 2000. *On Vibration Properties of Human Vocal Folds*. Groningen: Rijksuniversiteit Groningen. PhD thesis, p. 67, p. 93.

[14] Titze, I. R. 1994. Toward standards in acoustic analysis of voice. *J. Voice* 1, 1–7.

[15] Titze, I. R. 1995. Summary statement. *Workshop on acoustic voice analysis* Iowa. National Center for Voice and Speech.

[16] Titze, I. R. 2006. *The myoelastic aerodynamic theory of phonation*. Iowa City: The National Center for Voice and Speech. pp 259-262.

[17] Türck, L. 1866. *Klinik der Krankheiten des Kehlkopfes*. Vienna: Braumüller.

[18] Wong, D., Ito, M. R., Cox, N. B., Titze, I. R. 1991. Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. *J. Acoust. Soc. Am.* 89, 383–394.