

Simulation of vocal fold oscillation behaviour by a self-oscillating glottis model

B.J. KRÖGER

Institut für Phonetik der Universität zu Köln, Greinstr. 2, 50939 Köln, Germany

Abstract: As a good compromise between simplicity and approximation of real human vocal fold oscillation behaviour the Ishizaka-Flanagan two-mass model [1] is widely used in articulatory speech synthesis. But its poor results in modelling the oscillatory behaviour during glottal abduction/adduction and in modelling leakage flow during normal phonation are drawbacks. The two-mass model can be improved without increasing its complexity, if a non-oscillating and aperture-dependent bypass is added.

1. INTRODUCTION

A self-oscillating glottis model leads to physiology-related control parameters. Only simple rules for generating time functions for these control parameters as well as for controlling supraglottal articulation are needed [2].

2. THE TWO-MASS MODEL

The acoustics and aerodynamics of the vocal tract are modelled by a reflection type line analog [3, 4] (fig. 1). The vibration behaviour of the vocal folds is modelled by a two-mass approximation (fig. 2), i.e. the mechanical part of the glottis model. In this model the upper and lower part of each vocal fold are represented by different harmonic oscillators (damped mass-spring-systems with mass m_i , spring constant s_i , and damping r_i ; $i=1,2$). The masses are stiffness-coupled (k_C). The mechanical part is driven by pressure-induced forces acting on the inner surfaces of the vocal folds. The forces acting on the masses can be approximated from the pressure values of the tube representing the glottis p_G during the open phase of the glottis, and from subglottal pressure p_{sub} during the closed phase of the glottis [1]. The control parameters acting directly on the mechanical part of the glottis model are glottal aperture GA and cord tension CT (fig. 1). GA characterizes the equilibrium position of the vocal folds, i.e. the degree of abduction/adduction which can be controlled actively. The output of the mechanical part of the glottis model is the glottal area a_g which describes the instantaneous area of glottal constriction resulting from the actively controllable positioning of the folds and from their vibration.

3. THE NON-OSCILLATING BYPASS

Even during normal phonation, vocal fold vibration is a complex three-dimensional process. The Ishizaka-Flanagan two-mass approximation models two of them: The first dimension is defined by the direction of the displacement for the oscillatory masses (lateral direction) and the second dimension is defined by the direction from trachea to pharynx (vertical or inferior-superior direction), i.e. the direction in which the two masses are ordered. These dimensions define the plane of figure 2. According to the phase lag between the two mass pairs during oscillation, the two-mass approximation is able to model the elliptical movement for the center of mass of the folds occurring during normal phonation: The opening of the folds is accompanied

by an upward shift of glottal closure (e.g. [5], p. 265f). The third dimension is defined by the anterior-posterior direction, i.e. the direction given by the length of the glottal slit. According to the physiological fact that the anterior attachment of the folds is on the surface of the arytenoid cartilages and according to the fact, that these cartilages still form a part of the glottal slit (e.g. [6], pp. 135-140), the vocal folds can be divided into a cartilaginous and a membranous part (fig. 3). In our simple modification of the two-mass model, we take a non-oscillating bypass in order to model the cartilaginous part while the oscillating masses model the membranous part of the glottis (fig. 4). The ordinary two-mass model calculates the cross-sectional areas a_1 and a_2 between each mass pair m_1 and m_2 . The glottal area a_g is calculated by taking the minimum of a_1 and a_2 and by adding the bypass area a_{by} . The bypass can be implemented in two different ways as a linked leak or as a parallel chink [7, 8]. Since we are interested in dynamical glottal abduction/adduction behaviour, we here used a linked leak.

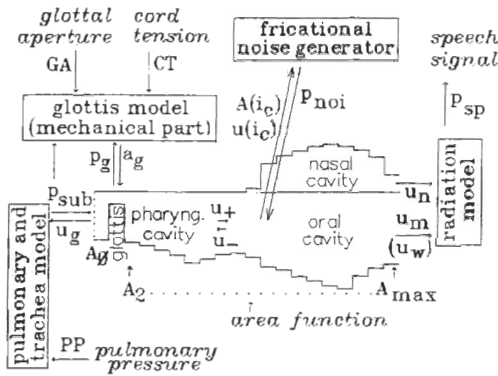


Figure 1 The aerodynamic-acoustic vocal tract model (in the center of figure), and its coupling to the pulmonary/trachea model, to the mechanical part of glottis model, to the frictional noise generator, and to the radiation model. Inputs and outputs in italics.

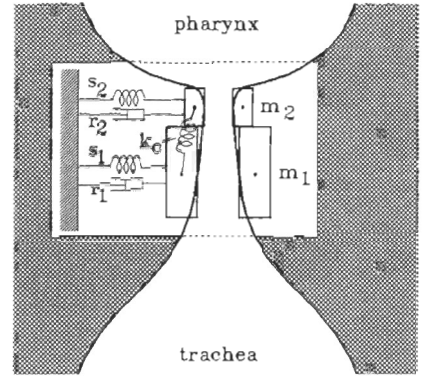


Figure 2 Cross-sectional frontal view of the glottis and its approximation by the two-mass model.

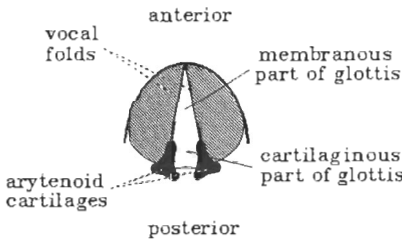


Figure 3 The glottis as seen from above. (superior view)

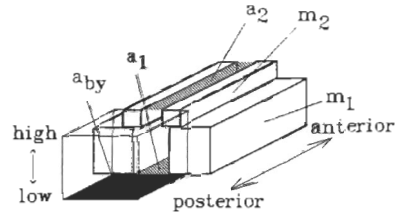


Figure 4 The modified two-mass model.

4. NORMAL PHONATION

Three cases can be differentiated for normal phonation ([5], pp. 272-274): a) the arytenoidal portion is involved in vocal fold vibration, leading to an unsimultaneous glottal closure; b) the arytenoid cartilages are spread apart, leading to a leaky glottis even in the so called closed phase of the glottal cycle; c) the arytenoids are tightly approximated. Case (c) is modelled well by the unmodified Flanagan-Ishizaka two-mass model and represents normal phonation. An abrupt glottal closure occurs. Case (b) represents a breathy voice and can be modelled using the linked leak case as implemented in our model. Fig. 5 illustrates the transition from case (c) to case (b) generated by our model. Most striking, the excitation strength (i.e. the negative peak of flow derivative u_g') decreases and the flow curve becomes more rounded, leading to less energy at high frequencies. Case (a) cannot be modelled by our simple modification. Here for example a constant tilt angle must be introduced for both mass pairs as done by [9]. It must be noted that in addition a fourth case is possible: The arytenoids are tightly approximated only at the joint with the membranous part and a triangular chink is formed between the arytenoids. This chink is separated from the time-varying glottal opening produced by the membranous part. In contrast to the linked leak the spectrum can be enhanced at high frequencies (relative to the no-leakage case) by the parallel chink [8].

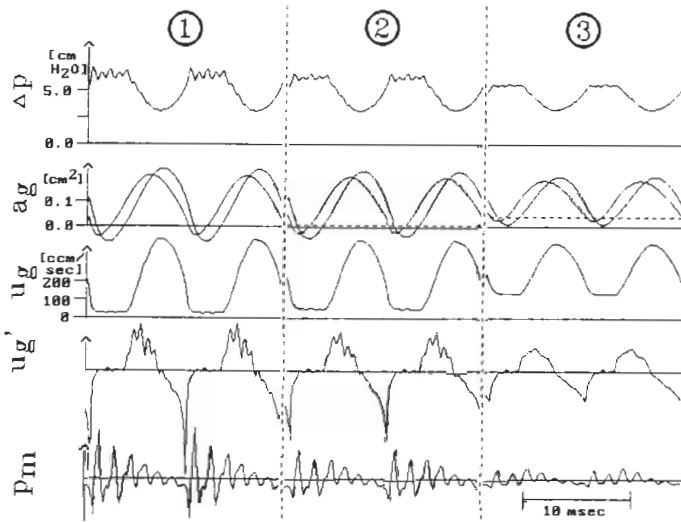


Figure 5 Transglottal pressure difference p , glottal area a_g , glottal flow u_g , its derivative u_g' , and radiated sound pressure p_m for phonation of /a:/ with increasing bypass areas.

- (1) $a_{by} = 0.005\text{cm}^2$,
- (2) $a_{by} = 0.01\text{cm}^2$, and
- (3) $a_{by} = 0.03\text{cm}^2$.
($a_{g0} = 0.04\text{cm}^2$).

5. ABDUCTION/ADDUCTION BEHAVIOUR

Glottal abduction/adduction gestures occur during the respiratory phase between phrases and during the production of voiceless sounds. The most important abductor is the posterior cricoarytenoid muscle, which mainly rotates the arytenoid cartilages but its lower fibers also produce a medial gliding movement ([6], p. 150). Its antagonists are the lateral cricoarytenoid muscle, which adducts the arytenoid cartilages mainly by rotation, and the interarytenoid muscle, which adducts them mainly by translation. In order to model abduction/adduction behaviour we want to use only glottal aperture GA as control parameter and we assume that the phonation neutral area a_{g0} as well as the bypass area a_{by} are functions of the glottal aperture GA . For positive values of GA (i.e. during glottal opening) we assume a constant ratio between a_{g0} and a_{by} (fig. 6). In the case of negative GA , a_{by} is zero but a_{g0} can become negative which is important for modelling medial compression. In order to estimate the ratio of bypass area to total glottal area ($a_{g0}+a_{by}$) it is important to know the length ratio of the membranous portion to the cartilaginous portion. Length ratios vary from 6:1 (after [10]) to 6:4 (after [11], p. 522 and [12]). Then from simple geometrical considerations (taking into account rotation or translation of the arytenoids) area ratios (bypass area to total area) in the vicinity of 0.14 to 0.64 can be calculated.

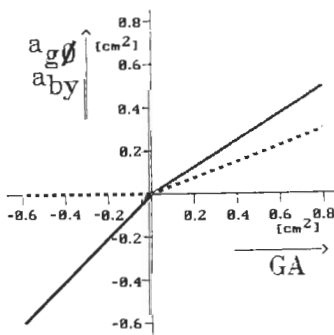


Figure 6 Phonation neutral area a_{g0} (solid line) and bypass area a_{by} (dashed line) as functions of glottal aperture GA . GA equals total area $a_{g0}+a_{by}$. (Ratio of bypass area to total area is 0.4 during glottal opening.)

Figure 7 shows glottal aperture GA , glottal area a_g and derivative of glottal flow u_g' for three abduction/adduction gestures with different values of maximal glottal aperture [13]. The vocal tract is fixed (configuration of /a:/). In column (1) no bypass is used and in column (2) the bypass is included and glottal aperture controls a_{g0} and a_{by} as shown in figure 6. Figure 7a shows increasing peak values of glottal area during abduction (See [1], p. 1253f: increasing phonation-neutral area increases glottal vibration amplitude but leads to undesirably high voicing build-up times). But the excitation strength remains roughly constant. With addition of the bypass the excitation strength decreases considerably during abduction. Figure 7b and 7c show an increase of the ac-amplitude for glottal area (a_g) in the beginning of the abduction gesture for the unmodified two-mass model. This behaviour is avoided by using the bypass.

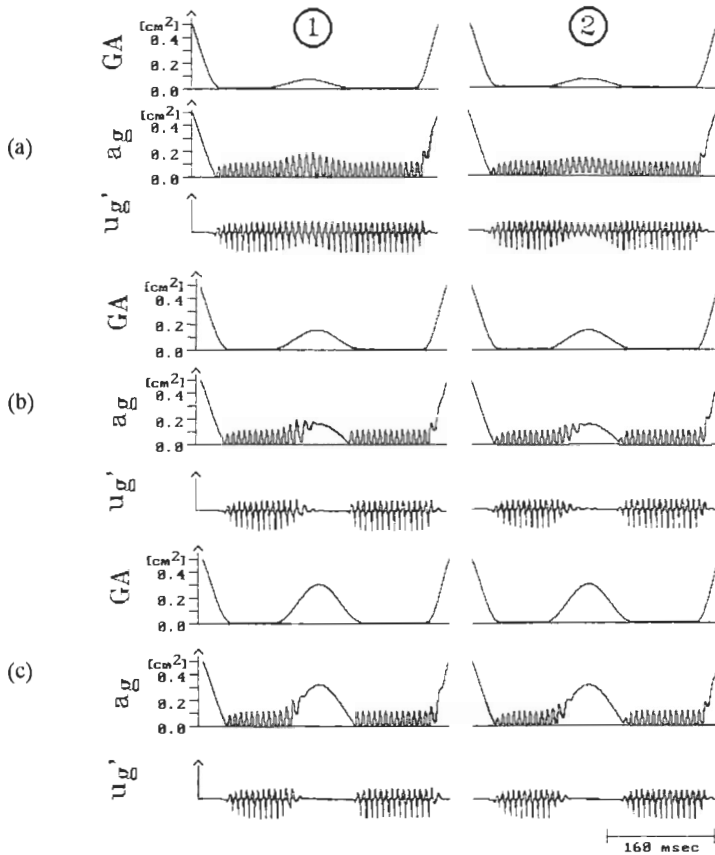


Figure 7 Glottal aperture, glottal area and derivative of glottal flow for three abduction/adduction gestures. (1) Without bypass, and (2) with bypass (area ratio is 0.4). Maximal glottal aperture:
 (a) $GA = 0.07\text{cm}^2$,
 (b) $GA = 0.15\text{cm}^2$,
 (c) $GA = 0.30\text{cm}^2$.

ACKNOWLEDGEMENTS

This work was supported in part by Deutsche Forschungsgemeinschaft DFG Grant He 434/21-1 and in part by ESPRIT-BR Project Nr. 6975 (SPEECH MAPS).

REFERENCES

- [1] Ishizaka K., Flanagan J.L., "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Tech. J.* **51** (1972) 1233-1268.
- [2] Kröger B.J., "Minimal rules for articulatory speech synthesis", in: Vandewalle J. et al. (eds.): *Signal processing VI: Theories and applications* (Amsterdam: Elsevier, 1992), pp. 331-334.
- [3] Kelly J.L., Lochbaum C.C., "Speech synthesis", *Proc. Fourth Intern. Congr. Acoust.* 1962, Paper G42. Reprinted in: Flanagan J.L., Rabiner L.R. (eds.), *Speech synthesis* (Stoudsburg: Dowden, Hutchinson & Ross, 1973) pp. 127-130.
- [4] Liljencrants J., *Speech synthesis with a reflection-type line analog*, Dissertation, Royal Inst. Techn., Stockholm (1985).
- [5] Stevens K.N., "Physics of laryngeal behavior and larynx modes", *Phonetica* **34** (1977) pp. 264-279.
- [6] Zemlin W.R., *Speech and hearing science. Anatomy and physiology* (Englewood Cliffs: Prentice-Hall, 1968).
- [7] Cranen B., Boves L., "On subglottal formant analysis", *J. Acoust. Soc. Am.* **81** (1987), pp. 734-746.
- [8] Schroeter J., Cranen B., "Physiologically-motivated modelling of the voice source in articulatory analysis/synthesis", *Proc. Eurospeech* (Berlin, 1993) pp. 95-98.
- [9] Childers D.G., Alsaka Y.A., Hicks D.M., Moore G.P., "Vocal fold vibrations: An EGG Model", in: Bear T. et al. (eds.), *Laryngeal function in phonation and respiration* (Boston: Little, Brown & Comp., 1987) pp. 181-202.
- [10] Hirano M., Kurita S., Nakashima T., "The structure of the vocal folds", in: Stevens K.N., Hirano M. (eds.) *Vocal fold physiology* (Tokyo: Univ. of Tokyo Press, 1981) pp. 33-44.
- [11] Holmberg E.B., Hillman R.E., Perkell J.S., "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", *J. Acoust. Soc. Am.* **84** (1988) pp. 511-529.
- [12] Hirano M., Yoshida T., Kurita S., Kiyokawa K., Sato K., Tateishi O., "Anatomy and behavior of the vocal process", in: Bear T. et al. (eds.) *Laryngeal function in phonation and respiration* (Boston: Little, Brown & Comp., 1987) pp. 3-13.
- [13] Lindqvist J., "Laryngeal articulation studied on swedish subjects", *Speech Transm. Lab. - Quart. Prog. Stat. Rep.* **2-3** (1972) pp. 10-27.