A SURVEY OF SELF-OSCILLATING LUMPED-ELEMENT MODELS OF THE VOCAL FOLDS

Peter Birkholz

Clinic of Phoniatrics, Pedaudiology, and Communication Disorders
University Hospital Aachen and RWTH Aachen University
pbirkholz@ukaachen.de

Abstract: Self-oscillating biomechanical models of the vocal folds are indispensable tools for articulatory speech synthesis, speech coding, and speech analysis. A successful compromise between simplicity and completeness are the low-dimensional vocal fold models with lumped elements. This paper provides a survey of this class of models and compares them in terms of their mechanical design, aerodynamic simulation, and with respect to their application and evaluation. Future studies to improve the performance of these models in the context of articulatory speech synthesis will be proposed.

1 Introduction

Every articulatory or parametric speech synthesizer needs a voice source model. These models can be roughly organized into three categories: parametric glottal flow models (e.g. [15]), kinematic vocal fold models (e.g. [49]), and self-oscillating biomechanical vocal fold models (e.g. [23]). Glottal flow models directly specify the shape of the glottal flow waveform (or its derivative) assuming that the glottal source and the vocal tract are linearly separable according to the source-filter theory of speech production [14]. However, it is now well established that the voice source and the vocal tract filter are generally not independent, in particular for female and child speech [51]. Kinematic vocal fold models *do* account for source-filter interaction as they specify the glottal area waveform that is then used to simulate the glottal flow. Self-oscillating biomechanical vocal fold models go even further and simulate the flow-induced oscillations of the vocal folds which in turn determine the glottal area. Thereby, they naturally reproduce effects that result from the interaction between the vocal folds, the glottal flow, and the vocal tract, such as intrinsic coarticulatory F_0 variation, vocal breaks at register transitions [53], and irregular vibration under certain conditions [42].

Within the category of self-oscillating biomechanical models are models of different (structural) complexity, ranging from rather simple lumped-mass models over distributed-mass models to complex finite element models. This study reviews the class of low-dimensional lumped-mass models, i.e. those models similar to the well-known two-mass model by Ishizaka and Flanagan [23]. Thereby we restrict the scope to those models where the mass elements are *not* divided in the dorso-ventral direction. Despite their simplicity, these models were shown to be able to reproduce many characteristics of real vocal fold oscillations [23]. Furthermore, they can produce natural-sounding voices [29] and have low computational costs. Since the first models were developed in the late 1960s, a large number of variants were proposed for different goals. This study organizes the models and compares them in terms of their mechanical design, aerodynamic simulation, and with respect to their application and evaluation. Based on this we identify open problems for future research that might further improve these models for speech applications.

2 Characterization of models

To effectively compare the existing low-dimensional self-oscillating vocal fold models, we decided to characterize each model in terms of its biomechanical characteristics, its aerodynamic characteristics, and its application and evaluation according to Tab. 1. The biomechanical model components were characterized by the following attributes:

Model type refers to the mechanical design of the models as depicted in Figures 1, 2, and 3. Most models were designed in the coronal plane only and are accordingly illustrated.

Masses per side indicates the number of lumped-mass elements per vocal fold.

Degrees of freedom (DOF) per side indicates the degrees of freedom of each vocal fold.

Spring type denotes the type of spring used to represent the tension of the vocal folds. They were either modeled as linear springs (“lin”), nonlinear springs (“nl”) with an additional cubic force-displacement relation ($F \propto x + ax^3$), or nonlinear springs (“A”) with an additional velocity-dependent term ($F \propto x + ax\dot{x}$).

Contact handling denotes how the restoring force was modeled, when the left and right vocal folds deform due to collision. “lin” and “nl” indicate that an additional linear or nonlinear spring was used according to the spring types above. “H” denotes the Hertz model of impact forces [32], and “hyp” and “tanh” denote approximations by a hyperbolic or hyperbolic tangent function.

Bilateral model indicates whether the two vocal folds were modeled independently or asymmetrically (mostly to study irregular oscillations).

The aerodynamic component was characterized by the following attributes:

Vena contracta refers to a flow separation effect at the entrance of the glottis that results in a pressure drop in excess of the drop predicted by the Benoulli equation, and hence in an energy loss. Van den Berg et al. [54] obtained a loss coefficient of 1.375 in their early experiments, which was subsequently used in many vocal fold models. Later the vena contracta effect was seriously doubted to occur in the glottis [44] and afterwards not considered at all by many models. A recent study indicates that the loss coefficient actually varies considerably depending on the glottal width and the subglottal pressure [19].

Pressure recovery indicates whether the dynamic pressure in the glottis was assumed to be partly recovered at the glottal exit. Similar to the vena contracta effect, the degree of pressure recovery at the glottal exit is controversial. A recent study by Fulcher et al. [19] shows that very little dynamic pressure (below 10%) is recovered under most conditions, and that the theoretically-derived equations used in most studies to calculate the pressure recovery do not actually agree with their experimental results.

Variable flow separation indicates whether the point where the airflow through the glottal channel separates from the walls was allowed to move (either explicitly or implicitly, e.g. in terms of an orifice discharge coefficient as in [57]). In early models, this point was implicitly assumed to be fixed at the exit of the glottis. However, Pelorson et al. [44] derived a more realistic fluid mechanical description of the airflow through the glottis and demonstrated the importance of a moving flow separation point by model simulations.

Source loading indicates whether the simulation of glottal flow was acoustically coupled to a model of the vocal tract ("v") and the trachea ("t"). This also includes low-frequency approximations in terms of a purely inductive load or one-formant approximations.

Viscous friction indicates whether air viscosity effects were considered in the glottis. They increase the pressure drop in the glottis and can have a pronounced effect on the flow waveform for small glottal openings.

There are many applications and possibilities to evaluate vocal fold models. We labeled the models and the corresponding studies with respect to the following possibilities:

Glottal flow indicates whether the model-generated glottal flow waveforms were analyzed with respect to different phonatory conditions (e.g. different degrees of abduction or different subglottal pressure) or compared or adapted to glottal flow waveforms of real speakers.

Glottal area/displacement indicates the analysis of the glottal area or displacement waveforms in the same way as for the glottal flow above. Glottal area waveforms of real speakers are becoming increasingly available due to advances of high-speed videoendoscopy [36].

Different or real subjects indicates whether attempts were made to model different voices using the same model, for example a male voice *and* a female voice, or whether the model was adapted to the voice of a real speaker, e.g. by glottal flow waveform matching.

Different voice qualities indicates whether different voice qualities or vocal registers were simulated, for example modal voice, breathy voice, and falsetto.

Onset, offset, bifurcations indicates whether the voice onset and offset was examined or compared with real data, or whether the model was analyzed in the context of the theory of nonlinear dynamics with respect to bifurcations (onset, offset, register transitions).

Perception tests indicates whether the synthetic voice was judged perceptually.

Irregular or disordered voices indicates whether such voices were produced and analyzed.

The above list is obviously not complete, but it contains some of the most important distinctive features of the class of models under consideration. With regard to the evaluation, an additional possibility is, e.g., to compare model-generated EGG waveforms, which reflect the contact area between the left and the right vocal fold, with real EGG waveforms. However, of all considered studies, this was only done in [45] and [50] for the models 2c and 3b.

With respect to their mechanical design, the models were divided into four groups: one-mass models, multi-mass models (with no body-cover separation), body-cover models, and pseudo one-mass models. A brief overview over the four groups with will be given below.

2.1 One-mass models

The first one-mass model was developed in 1968 [18] as a single linear mass-spring system (Fig. 1a). Although it could produce acceptable voiced sounds and simulate many properties

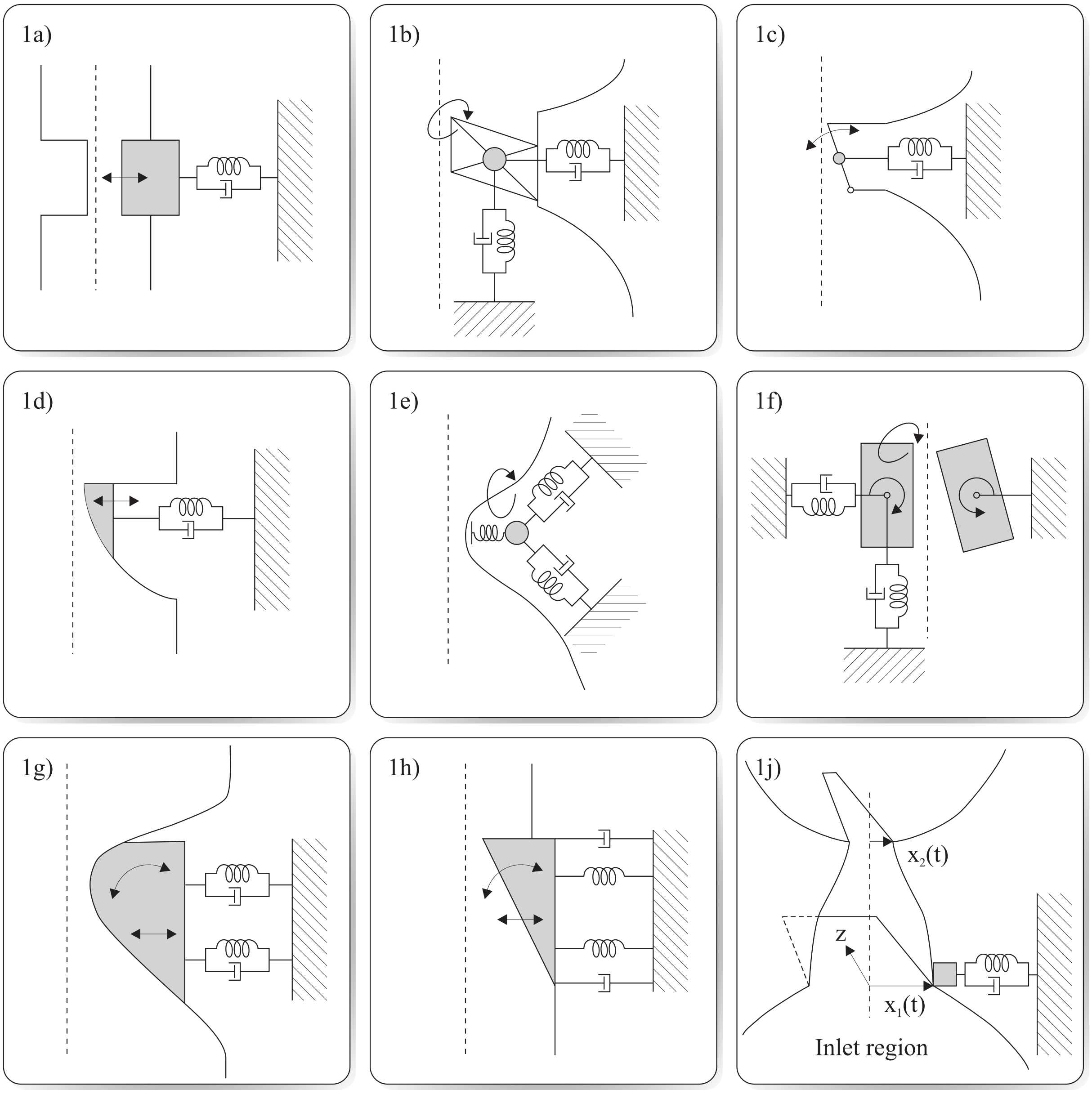

Figure 1 - Variants of vocal fold models with one mass element per vocal fold.

of glottal flow, it was incapable of sustained oscillations for a capacitive input load of the vocal tract, i.e. at frequencies just above a formant. Furthermore, like most one-mass models, it could not simulate a phase difference between the upper and lower vocal cord edges that occurs during chest register phonation. Adachi and Yu [2] proposed the model 1b, where each vocal fold was allowed to vibrate both in parallel and perpendicular to the airflow. This model was not only capable of self-sustained oscillations with a capacitive vocal tract load, but the transition between oscillations with an inductive and capacitive load was even smoother than for the classical two-mass model [23]. Liljencrants developed the one-mass model 1f with both a translational and a rotational DOF per vocal fold, which could, compared to the above models, simulate the wave-like motion in the mucosal vocal fold cover and oscillate over a wide range of frequencies [33]. This model was later extended with an additional DOF for vertical translational movement and applied to the reproduction of real glottal flow waveforms and for the synthesis of different voice qualities [27, 34]. The models 1e, 1g, and 1h are other approaches with one oscillating mass but two DOF per vocal fold according to [6, 46, 1]. An “effective” one-mass model with only one DOF (Fig. 1d) was introduced by Zañartu et al. [57] to analyze the relative importance of acoustic loading (due to the subglottal system and the vocal tract) and the time-varying glottal flow resistance to the oscillations. The time-varying flow resistance was based on the implementation of a time-varying discharge coefficient to model flow-related effects like the vena contracta effect and pressure recovery. Little et al. [35] and Bäckström [7] applied the one-mass models 1c and 1a to estimate the model parameters from a glottal flow signal for applications in speech coding and re-synthesis.

2.2 Multi-mass models

The multi-mass models discussed here have more than one mass element per vocal fold but do not distinguish between the vocal fold body and cover. Figure 2a shows the design of the classical two-mass model by Ishizaka and Flanagan [23]. Compared to the one-mass model 1a it can account for phase differences between the upper and lower vocal fold edges and hence oscillate over a wider range of frequencies. It was extended by a vertical DOF (Fig. 2b) in [24, 17] to examine the acoustic significance of *vertical* vocal fold motion compared to horizontal motion. Koizumi et al. [29] proposed the model 2e with a second vertical DOF, which was found to produce a more natural-sounding voice than model 2a. Later they introduced model 2h with a third mass element added on top of *one* vocal fold to simulate a polyp. This model proved to be an effective noninvasive diagnostic tool to discriminate (real) hoarse voices caused by a polyp from hoarse voices due to other laryngeal diseases. Tokuda et al. [52] added a third mass on top of *both sides* of the two-mass model (Fig. 2j) and showed that it could simulate the coexistence of modal voice and falsetto during the transition between the two registers corresponding to experimental data. Imagawa et al. [22] combined the two-mass model 2a with additional oscillators for the ventricular folds (Fig. 2g) and found that both the vocal and ventricular folds play an important role in determining the voice quality in singing. Figure 2f shows an alternative model design including the ventricular folds. In contrast to the above models, the models 2c, 2d, and 2f have a smooth, as opposed to step-wise, area transition between the lower and the upper vocal fold edges, which allows to incorporate a *continuously* moving point of flow separation in the glottis for improved aerodynamic modeling. Finally, the models 2k and 2m account for shape variations in the dorso-ventral dimension of the glottis, while all other models assume a fixed shape along this dimension. While Kröger [30] and McGowan et al. [41] modeled the third dimension in terms of an anterior oscillating part and a non-oscillating posterior chink, Birkholz et al. [9, 8] aligned the mass elements inclined with respect to the dorso-ventral axis as function of the degree of gottal abduction. This allowed controlling the abruptness of glottal closure and hence to simulate different voice qualities.

2.3 Body-cover models

The main drawback of all lumped parameter models above is that their discretization of tissue in the coronal plane does not capture the layered body-cover structure of the vocal folds. This limitation motivated the introduction of the three-mass model 3a that essentially adds a “body” mass lateral to the two cover masses [48]. Figure 3b shows a “bar-plate” version of the body-cover model that was patterned after the one-mass model 1f and used to study the regulation of glottal airflow in phonation [50]. Tokuda et al. [53] recently introduced the four-mass body-cover polygon model 3c with a more detailed representation of the cover geometry and used it to study register transitions and the influence of sub- and supraglottal resonances on phonation.

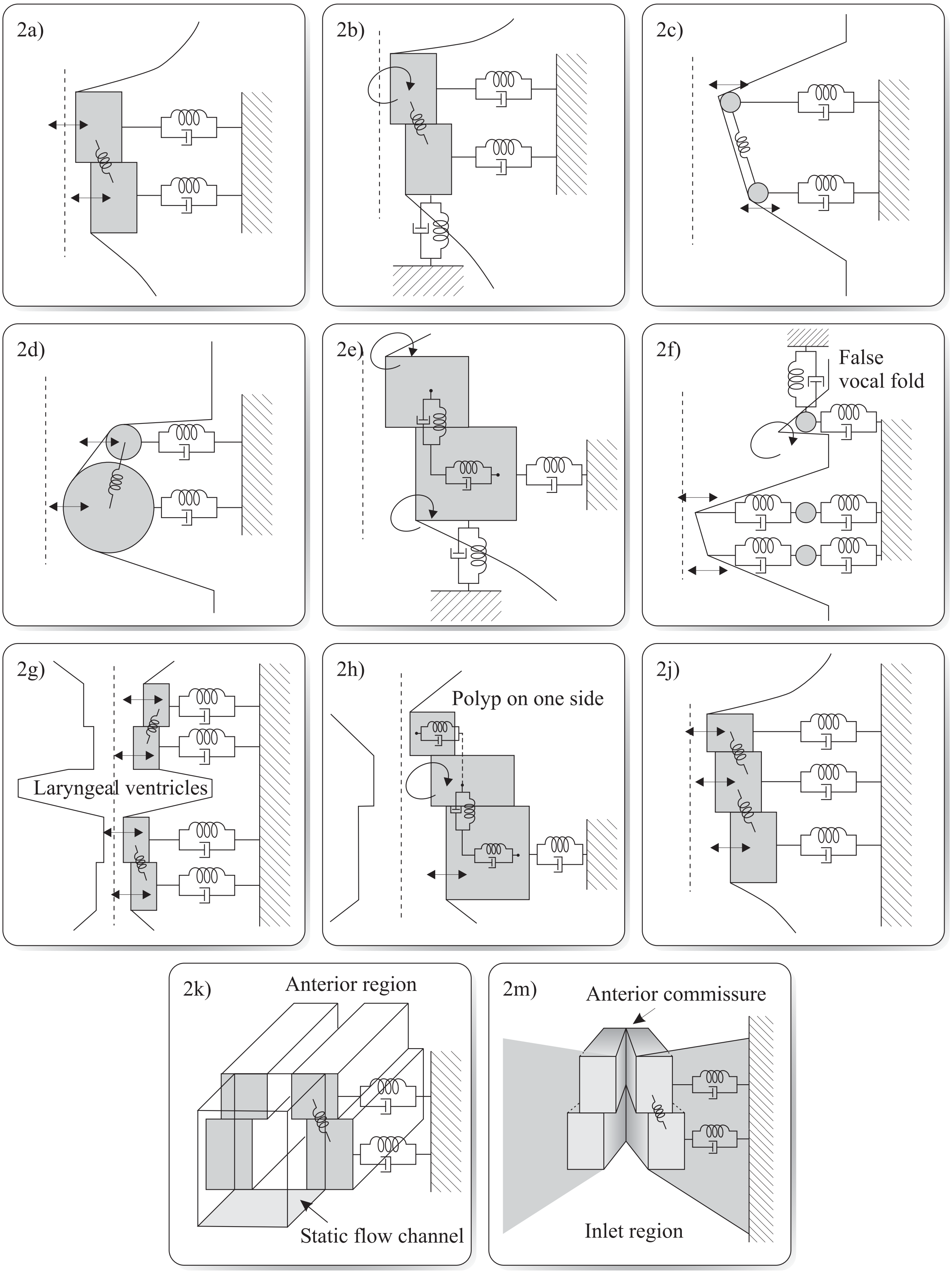

Figure 2 - Lumped multi-mass models of the vocal folds without a body-cover separation.

2.4 Pseudo one-mass models

The more detailed the biomechanical structure of a vocal fold model is, the more parameters are typically needed to control it. Therefore, complex models are difficult to fit to observed

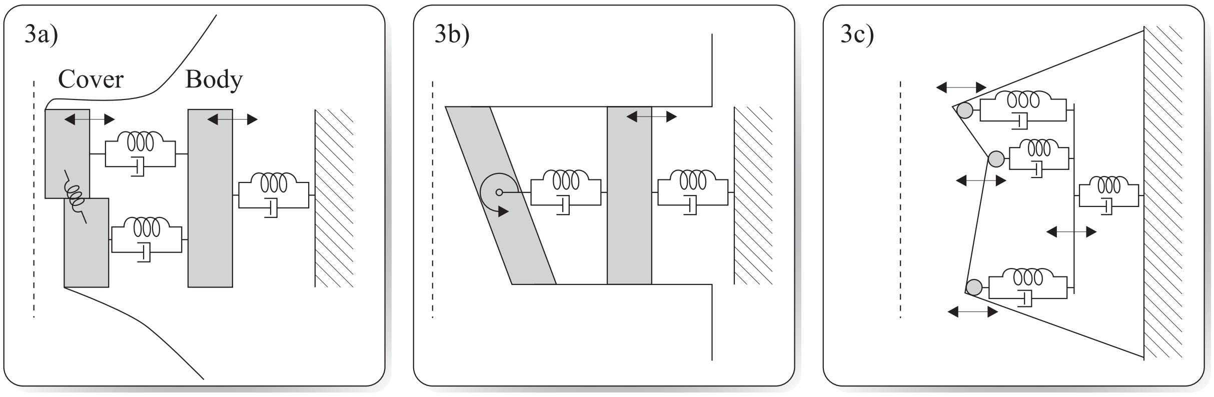

Figure 3 - Lumped body-cover models of the vocal folds.

data and they are rarely adopted in practical applications like speech coding or synthesis [11]. This motivated the development of pseudo one-mass models, i.e. models that combine a highly simplified mechanical design with phenomenological knowledge to model the glottal shape and aerodynamics. Figure 1j shows such a model where the lower edge of each fold is represented by a single mass-spring system, and a delay line models the propagation of the oscillatory motion along the inferior-superior dimension of the vocal folds to reproduce the wave-like motion of the cover [12, 11]. This model was later extended by a delay line along the dorso-ventral direction to model the gradual opening and closing of the vocal folds [13]. Avanzini et al. [4] described a pseudo one-mass model based on the two-mass model 2a, where just the motion of the lower mass was simulated by a spring-mass system, and the upper mass was assumed to follow the lower mass with a fixed delay. This idea was later adopted for the two-mass model 2c [3].

3 Discussion

In the past four decades, a remarkable variety of low-dimensional lumped-mass models of the vocal folds was proposed. According to Tab. 1, each model can be characterized by its biomechanical design and aerodynamic features. In each study, the design decisions for the vocal fold model were guided by the intended goals, but the design rationale was not always made explicit. Only very few studies evaluated how different biomechanical or aerodynamic characteristics contribute to the performance of a model for a given task. In other words, the optimal design decisions and the required level of detail of a lumped vocal fold model for a particular task remain mostly unclear. In fact, only one of the considered studies [29] actually compared different biomechanical model designs with respect to the goal of the study (in this case the naturalness of the synthetic voice).

For future studies about lumped vocal fold models, we therefore suggest analyzing the effect of individual model characteristics on the performance for a given task in more detail. For example, what level of detail is really needed for natural-sounding voice synthesis with regard to the degrees of freedom or the type of springs used to represent the vocal fold tension? Or what is gained by using the more sophisticated Hertz model of impact forces to handle vocal fold contact instead of linear springs or nonlinear springs with a cubic force term? With regard to the aerodynamic simulation, there is now consensus that a variable flow separation point within the glottal channel and an attached vocal tract and trachea are important features for realistic voice simulations. However, the treatment of the pressure drop at the glottal inlet and the pressure recovery at the glottal outlet are still controversial. The consideration of the recent experimental

	Biomechanics						Aerodynamics					Application and evaluation						
Publication	Model type	Masses per side	DOF per side	Spring type	Contact handling	Bilateral model	Vena contracta	Pressure recovery	Var. flow separation	Source loading	Viscous friction	Glottal flow	Glottal area/disp.	Diff. or real subj.	Diff. voice qualities	Onset, offset, bifurc.	Perception tests	Irreg. or disord. v.
[16, 18]	1a	1	1	lin	lin	-	✓		-	v	✓	✓	✓	-	-	-	-	-
[7]	1a	1	1	nl	nl	-			-	-	-	✓	-	✓	-	-	-	-
[21]	1a	1	1	lin		-			✓	-	✓	-	-	-	-	-	-	-
[2]	1b	1	2	nl	lin	-	-	✓	-	v	✓	-	✓	✓	✓	✓	-	-
[35]	1c	1	1	A	nl	-	-	-	-	-	-	-	-	-	-	-	-	-
[57]	1d	1	1	lin	H	-	✓	✓	✓	t,v	-	✓	✓	-	-	-	-	-
[6]	1e	1	2	nl	hyp	-				-	-	-	-	-	-	✓	-	✓
[33]	1f	1	2	lin	lin	✓	✓	✓	✓	t,v	-	✓	-	-	-	-	-	-
[27, 34]	1f	1	3	lin	lin	✓	✓	✓	✓	t,v	-	✓	-	✓	✓	-	-	✓
[46]	1g	1	2	lin	H	-	-	-	✓	-	✓	-	-	-	-	-	-	-
[1]	1h	1	2	lin	H	✓			-	v	✓	✓	-	-	-	-	-	-
[23]	2a	2	2	nl	nl	-	✓	✓	-	v	✓	✓	✓	-	-	-	-	-
[25]	2a	2	2	nl		✓	✓	✓	-	t,v	✓	-	✓	-	-	-	-	✓
[38]	2a	2	2	lin	lin	-	✓	-	-	-	-	-	-	-	-	✓	-	-
[20]	2a	2	2	lin	hyp	-	✓	-	-	-	-	-	-	-	-	✓	-	✓
[47]	2a	2	2	lin	lin	✓	-	-	✓	-	-	-	-	-	-	✓	-	✓
[39]	2a	2	2	lin	lin	-	-	-	✓	-	-	-	✓	-	✓	-	-	-
[56]	2a	2	2	lin	lin	✓	N. Stokes			t,v	✓	✓	✓	-	-	-	-	✓
[42]	2a	2	2	lin	lin	✓	-	-	✓	-	-	-	✓	✓	-	-	-	✓
[31]	2a	2	2	nl	nl	-	-	-	✓	v	✓	✓	-	✓	-	-	-	✓
[40]	2a	2	2	nl	nl	-	-	-	✓	v	✓	✓	-	✓	-	✓	-	-
[5]	2a	2	2	lin	lin	-	-	✓	✓	v	-	✓	-	✓	✓	-	-	-
[24, 17]	2b	2	3	nl	nl	-	✓	✓	-	t,v	✓	✓	✓	-	-	-	✓	-
[55]	2c	2	2	lin	nl	-	-		✓	t,v		✓	-	✓	-	-	-	-
[37]	2c	2	2	lin	lin	-	-	-	✓	t,v	-	✓	-	-	-	-	-	-
[45]	2c	2	2	lin	lin	-	-	-	✓	t,v	✓	✓	-	-	✓	-	-	-
[44]	2d	2	2	lin	lin	-	-	-	✓	-	✓	✓	-	✓	-	-	-	-
[26]	2d	2	2	lin	lin	-	✓	-	✓	t,v	✓	✓	✓	-	-	-	-	-
[29]	2e	2	4	nl	nl	-	✓	✓	-	t,v	✓	✓	✓	-	-	-	✓	-
[43]	2f	2+1	2+2	lin		-	✓		✓	-	✓	✓	✓	-	-	✓	-	-
[22]	2g	2+2	2+2	nl	nl	-	✓	✓	-	t,v	✓	-	-	-	✓	-	-	-
[28]	2h	2,3	3,4			✓				t,v		-	-	-	-	-	-	✓
[52]	2j	3	3	lin	tanh	-	-	-	✓	-	-	-	-	-	✓	✓	-	-
[30]	2k	2	2	nl	nl	-	✓	✓	-	v	✓	✓	✓	-	✓	-	-	-
[41]	2k	2	2			-				v	✓	✓	-	✓	-	✓	-	-
[9, 8]	2m	2	2	lin	lin	-	-	-	✓	t,v	✓	✓	✓	-	✓	-	✓	-
[48]	3a	3	3	nl	nl	-	-	-	✓	t,v	-	✓	✓	-	-	-	-	-
[50]	3b	2	3	lin		-	-	✓	✓	t,v	-	✓	-	-	-	-	-	-
[53]	3c	4	4	lin	lin	-	-	-	✓	t,v	-	-	-	-	✓	✓	-	-
[12, 11]	1j	1	1	lin	lin	✓	-			v		✓	-	✓	✓	-	✓	-
[3]	(2c)	2	1	lin	lin	-	-	-	✓	v	-	✓	✓	-	-	-	-	-
[4]	(2a)	2	1	lin		-	✓	✓	-	v	✓	✓	-	-	-	-	-	-

Table 1 - Characteristics of the vocal fold models. For a description of the labels refer to Sec. 2. Empty cells indicate that the specific feature was not apparent from the model description. From top to bottom, the four blocks of rows present one-mass model, multi-mass models (without body-cover separation), body-cover models, and pseudo-one-mass models.

measurements in [19] might be an important step towards a more realistic treatment of these effects. Furthermore, we propose more extensive evaluations of the models. While many studies analyzed the behavior of a model qualitatively, few studies compare the behavior *quantitatively* with data of real speakers, for example in terms of glottal flow or area waveform properties. Also valuable is the comparison of computer-simulated models with identical physical models, as in [46].

One of the most challenging tasks for a vocal fold model is certainly high-quality articulatory speech synthesis. In this context, it is important that the vocal fold model not only generates natural self-sustained oscillations for a wide variety of conditions, but also realistically simulates the oscillatory behavior during the onset and offset of voicing and during register transitions. These topics were in the focus of some modeling studies (e.g. [38, 40, 41, 53]) but still leave many open questions for future research. High-quality articulatory synthesis also needs a vocal fold model that can convincingly simulate different voice qualities, e.g. the continuum of pressed over modal to breathy phonation, because people use voice quality in a controlled way to signal paralinguistic information, similar to fundamental frequency [10]. Most of the low-dimensional lumped-mass models cannot simulate these voice quality variations, because they cannot produce different gradations of glottal closure or opening that distinguish, e.g., pressed and breathy voice. A few studies addressed this problem [9, 30, 13] but more research is needed in this direction. Finally, the voices produced by the vocal fold models were rarely evaluated perceptually, which makes it difficult to assess models with regard to speech synthesis applications.

4 Acknowledgments

I am deeply grateful to Ingmar Steiner and Bernd J. Kröger for their detailed comments on the paper.

References

[1] Aalto, A., P. Alku and J. Malinen: *A LF-pulse from a simple glottal flow model*. In *International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA'09)*, pp. 199–202, Florence, Italy, 2009.

[2] Adachi, S. and J. Yu: *Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing*. Journal of the Acoustical Society of America, 117(5):3213–3224, 2005.

[3] Avanzini, F.: *Simulation of vocal fold oscillation with a pseudo-one-mass physical model*. Speech Communication, 50(2):95–108, 2008.

[4] Avanzini, F., P. Alku and M. Karjalainen: *One-delayed-mass model for efficient synthesis of glottal flow*. In *Eurospeech 2001*, pp. 51–54, Aalborg, Denmark, 2001.

[5] Avanzini, F., S. Maratea and C. Drioli: *Physiological Control of Low-Dimensional Glottal Models with Applications to Voice Source Parameter Matching*. Acta Acustica united with Acustica, 92(5):731–740, 2006.

[6] Awrejcewicz, J.: *Numerical Analysis of the Oscillations of Human Vocal Cords*. Nonlinear Dynamics, 2(1):35–52, 1991.

[7] Bäckström, T.: *Fitting Mass-Spring Models to Glottal Flow Estimates*. In *ISCA ITRW, Speech Analysis and Processing for Knowledge Discovery*, Aalborg, Denmark, 2008.

[8] BIRKHOLZ, P., B. J. KRÖGER and C. NEUSCHEAFER-RUBE: *Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds*. In *First International Workshop on Performative Speech and Singing Synthesis (p3s 2011)*, Vancouver, BC, Canada, 2011.

[9] BIRKHOLZ, P., B. J. KRÖGER and C. NEUSCHEAFER-RUBE: *Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis*. In *Interspeech 2011*, Florence, Italy, 2011.

[10] CAMPBELL, N. and P. MOKHTARI: *Voice quality: the 4th prosodic dimension*. In *15th International Congress of Phonetic Sciences (ICPhS'03)*, pp. 2417–2420, Barcelona, Spain, 2003.

[11] DRIOLI, C.: *A flow waveform-matched low-dimensional glottal model based on physical knowledge*. Journal of the Acoustical Society of America, 117(5):3184–3195, 2005.

[12] DRIOLI, C. and F. AVANZINI: *Non-modal voice synthesis by low-dimensional physical models*. In *International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA'03)*, pp. 151–154, Florence, Italy, 2003.

[13] DRIOLI, C. and F. AVANZINI: *Improved fold closure in mass-spring low-dimensional glottal models*. In *International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA'07)*, pp. 51–54, Florence, Italy, 2007.

[14] FANT, G.: *Acoustic theory of speech production*. Mouton, The Hague, 1960.

[15] FANT, G., J. LILJENCRANTS and Q. LIN: *A four-parameter model of glottal flow*. STL-QPSR, 26(4):1–13, 1985.

[16] FLANAGAN, J. L. and L. CHERRY: *Excitation of Vocal-Tract Synthesizers*. Journal of the Acoustical Society of America, 45(3):764–769, 1969.

[17] FLANAGAN, J. L. and K. ISHIZAKA: *Computer model to characterize the air volume displaced by the vibrating vocal cords*. Journal of the Acoustical Society of America, 63(5):1559–1565, 1978.

[18] FLANAGAN, J. L. and L. L. LANDGRAF: *Self-Oscillating Source for Vocal-Tract Synthesizers*. IEEE Transactions on Audio and Electroacoustics, AU-16(1):57–64, 1968.

[19] FULCHER, L. P., R. C. SCHERER and T. POWELL: *Pressure distributions in a static physical model of the uniform glottis: Entrance and exit coefficients*. Journal of the Acoustical Society of America, 129(3):1548–1553, 2011.

[20] HERZEL, H. and C. KNUDSEN: *Bifurcations in a Vocal Fold Model*. Nonlinear Dynamics, 7(1):53–64, 1995.

[21] HOWE, M. S. and R. S. MCGOWAN: *On the single-mass model of the vocal folds*. Fluid Dynamics Research, 42:1–16, 2010.

[22] IMAGAWA, H., K.-I. SAKAKIBARA, N. TAYAMA and S. NIIMI: *The effect of the hypopharyngeal and supra-glottic shapes on the singing voice*. In *Stockholm Music Acoustics Conference (SMAC'03)*, pp. 1–4, Stockholm, Sweden, 2003.

[23] ISHIZAKA, K. and J. L. FLANAGAN: *Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords*. The Bell System Technical Journal, 51(6):1233–1268, 1972.

[24] ISHIZAKA, K. and J. L. FLANAGAN: *Acoustic Properties of Longitudinal Displacement in Vocal Cord Vibration*. The Bell System Technical Journal, 56(6):889–918, 1977.

[25] ISHIZAKA, K. and N. ISSHIKI: *Computer simulation of pathological vocal-cord vibration*. Journal of the Acoustical Society of America, 60(5):1193–1198, 1976.

[26] KABURAGI, T.: *Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling*. Journal of the Acoustical Society of America, 129(3):1554–1567, 2011.

[27] KARLSSON, I. and J. LILJENCRANTS: *Wrestling the two-mass model to conform with real glottal wave forms*. In *3rd International Conference on Spoken Language Processing (ICSLP'94)*, pp. 151–154, Yokohama, Japan, 1994.

[28] KOIZUMI, T. and S. TANIGUCHI: *A novel model of pathological vocal cords and its application to the diagnosis of vocal cord polyp*. In *1st International Conference on Spoken Language Processing (ICSLP'90)*, pp. 73–76, Kobe, Japan, 1990.

[29] KOIZUMI, T., S. TANIGUCHI and S. HIROMITSU: *Two-mass models of the vocal cords for natural sounding voice synthesis*. Journal of the Acoustical Society of America, 82(4):1179–1192, 1987.

[30] KRÖGER, B. J.: *Zur artikulatorischen Realisierung von Phonationstypen mittels eines selbstschwingenden Glottismodells*. Sprache-Stimme-Gehör, 21:102–105, 1997.

[31] KUO, H.-K. J.: *Voice source modeling and analysis of speakers with vocal-fold nodules*. PhD thesis, Massachusetts Institute of Technology, 1998.

[32] LANDAU, L. D. and E. M. LIFSHITZ: *Theory of Elasticity*. Pergamon Press, 1970.

[33] LILJENCRANTS, J.: *A translating and rotating mass model of the vocal folds*. STL-QPSR, 1:1–18, 1991.

[34] LILJENCRANTS, J.: *Experiments with analysis by synthesis of glottal airflow*. In *4th International Conference on Spoken Language Processing (ICSLP'96)*, pp. 1289–1292, Philadelphia, PA, USA, 1996.

[35] LITTLE, M., P. MCSHARRY, I. MOROZ and S. ROBERTS: *A simple, quasi-linear, discrete model of vocal fold dynamics*. In FAUNDEZ-ZANUY, M., L. JANER, A. ESPOSITO, A. SATUE-VILLAR, J. ROURE and V. ESPINOSA-DURO (eds.): *Nonlinear Analyses and Algorithms for Speech Processing*, pp. 348–356. Springer, Berlin, 2005.

[36] LOHSCHELLER, J., H. TOY, F. ROSANOWSKI, U. EYSHOLDT and M. DÖLLINGER: *Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos*. Medical Image Analysis, 11(4):400–413, 2007.

[37] LOUS, N. J. C., G. C. J. HOFMANS, R. N. J. VELDHUIS and A. HIRSCHBERG: *A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prothesis Design*. Acta Acustica united with Acustica, 84(6):1135–1150, 1998.

[38] LUCERO, J. C.: *Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region*. Journal of the Acoustical Society of America, 94(6):3104–3111, 1993.

[39] LUCERO, J. C.: *Chest- and falsetto-like oscillations in a two-mass model of the vocal folds*. Journal of the Acoustical Society of America, 100(5):3355–3359, 1996.

[40] LUCERO, J. C. and L. L. KOENIG: *Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control*. Journal of the Acoustical Society of America, 117(3):1362–1372, 2005.

[41] MCGOWAN, R. S., L. L. KOENIG and A. LÖFQVIST: *Vocal tract aerodynamics in /aCa/ utterances: Simulations*. Speech Communication, 16(1):67–88, 1995.

[42] MERGELL, P., H. HERZEL and I. R. TITZE: *Irregular vocal fold vibration – high-speed observation and modeling*. Journal of the Acoustical Society of America, 108(6):2996–3002, 2000.

[43] MILLER, J. A., J. C. PEREIRA and D. W. THOMAS: *Fluid flow through the larynx channel*. Journal of Sound and Vibration, 121(2):277–290, 1988.

[44] PELORSON, X., A. HIRSCHBERG, R. R. VAN HASSEL, A. P. J. WIJNANDS and Y. AUREGAN: *Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model*. Journal of the Acoustical Society of America, 96(6):3416–3431, 1994.

[45] SCIAMARELLA, D. and C. D'ALESSANDRO: *On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters*. Acta Acustica united with Acustica, 90(4):746–761, 2004.

[46] ŠIDLOF, P. and J. HORÁČEK: *Vocal fold motion and voice production: Mathematical modelling and experiment*. In *Forum Acusticum 2005*, pp. 2737–2742, Budapest, Hungary, 2005.

[47] STEINECKE, I. and H. HERZEL: *Bifurcations in an asymmetric vocal-fold model*. Journal of the Acoustical Society of America, 97(3):1874–1884, 1995.

[48] STORY, B. H. and I. R. TITZE: *Voice simulation with a body-cover model of the vocal folds*. Journal of the Acoustical Society of America, 97(2):1249–1260, 1995.

[49] TITZE, I. R.: *A Four-Parameter Model of the Glottis and Vocal Fold Contact Area*. Speech Communication, 8(3):191–201, 1989.

[50] TITZE, I. R.: *Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model*. Journal of the Acoustical Society of America, 111(1):367–376, 2002.

[51] TITZE, I. R.: *Nonlinear source-filter coupling in phonation: Theory*. Journal of the Acoustical Society of America, 123(5):2733–2749, 2008.

[52] TOKUDA, I. T., J. HORAČEK, J. G. SVEC and H. HERZEL: *Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments*. Journal of the Acoustical Society of America, 122(1):519–531, 2007.

[53] TOKUDA, I. T., M. ZEMKE, M. KOB and H. HERZEL: *Biomechanical modeling of register transitions and the role of vocal tract resonators*. Journal of the Acoustical Society of America, 127(3):1528–1536, 2010.

[54] VAN DEN BERG, J., J. T. ZANTEMA and P. DOORNENBAL: *On the air resistance and the Bernoulli effect of the human larynx*. Journal of the Acoustical Society of America, 29(5):626–631, 1957.

[55] VESCOVI, C., E. CASTELLI and X. PELORSON: *Adaptation of a two-mass model of the vocal cords to a particular speaker*. In *4th European Conference on Speech Communication and Technology (Eurospeech'95)*, pp. 1933–1936, Madrid, Spain, 1995.

[56] XUE, Q., R. MITTAL and X. ZHENG: *A computational study of the effect of vocal-fold asymmetry on phonation*. Journal of the Acoustical Society of America, 128(2):818–827, 2010.

[57] ZAÑARTU, M., L. MONGEAU and G. R. WODICKA: *Influence of acoustic loading on an effective single mass model of the vocal folds*. Journal of the Acoustical Society of America, 121(2):1119–1129, 2007.