Biomedical Signal Processing and Control 64 (2021) 102309

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Modelling sagittal and vertical phase differences in a lumped and distributed elements vocal fold model

Carlo Drioli [a,*], Philipp Aichinger [b]

[a] Department of Mathematics, Computer Science and Physics, University of Udine, Udine 33100, Italy
[b] Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Vienna, Austria

ARTICLE INFO

Keywords:
High-speed video analysis
Vocal folds dynamical modelling
Voice quality characterization
Voice disorders

ABSTRACT

The quality and timbre of disordered voices heavily rely on the vibration properties of the vocal folds. We discuss the representation of sagittal phase differences in vocal fold oscillations through a numerical biomechanical model involving lumped elements as well as distributed elements, i.e., delay lines. A dynamic glottal source model is proposed in which the fold displacement along the vertical and the sagittal dimensions is modelled using delay lines. In contrast to other models, with which the reproduction of sagittal phase differences is impossible (e.g., in two-mass models) or not easy to control (e.g., in 3D 16-mass and multi-mass models in general), the one proposed here provides direct control over the amount of phase delay between folds' oscillations at the posterior and anterior part of the glottis, i.e., the sagittal axis, and at the superior and inferior part of the glottis, i.e., the vertical axis, while keeping the dynamic model simple and computationally efficient. The model is assessed by addressing the reproduction of oscillatory patterns observed in high-speed videoendoscopic data, in which sagittal phase differences are observed. Also, timing asymmetry parameters observed in hemi glottal area waveforms (GAWs) are used for fitting.

1. Introduction

Model-based descriptions of vocal fold vibration are central to the understanding of the human voice's function. In particular, voice quality depends on the properties of vocal fold vibration. An important property of vocal fold vibration is the spatial difference of the vibratory phases, which commonly occurs in pathological, but also healthy phonation [1–5]. The potential clinical application of such model-based descriptions are to find normative and nonnormative ranges of mechanical vocal fold parameters, which enable distinction of normal and abnormal voice to support the indication, selection, evaluation as well as optimization of medical treatment techniques. These treatment techniques include voice therapy conducted by speech-language pathologist/logopedists, or phonosurgery, which includes surgery that aims at improving voice quality.

When observing vocal fold oscillatory patterns, vertical, sagittal, and lateral phase differences are distinguished. Vertical phase differences are known to play a pivotal role in the biomechanics of the self-sustained oscillation of the vocal folds, as they promote the transfer of energy from the transglottal airflow to the vocal fold kinetics. The inferior parts of the vocal folds are normally ahead of the superior parts, which results in convergence of vocal folds during vibratory opening, and divergence during closing. Lateral phase differences, i.e., phase differences between the left and the right vocal fold, are primarily caused by differences of mass and tension of the vocal folds, and thus typical for laryngeal paralyses, as well as for vocal fold lesions. However, a lack of understanding vocal fold vibration patterns involving sagittal phase differences is observed. In this work, the pathologies for which we observe phase differences are functional voice disorders (3 subjects), Reinke's edema (2 subjects), bamboo nodes (1 subject), and laryngitis (1 subject). Data of two healthy subjects involving phase differences are also included.

Video data acquisition and processing are already recognized as essential tools for voice quality assessment and medical diagnosis. Processing strategies that are particularly relevant include videokymography, which enables visualization of lateral differences [6], multislice videokymography, which extends the former to enable visualization of sagittal differences [7], and phonovibrography, which enables visualization based on graphical segmentation of the glottal gap [8]. Recent research discussing connections between biomechanical modelling of the folds and high-speed videoendoscopic or videokymographic techniques can be found in [9–11]. A multi-parameter approach that aims at indicating zipper-like sagittal phase differences during glottal opening was proposed recently [12].

[*] Corresponding author.
E-mail addresses: carlo.drioli@uniud.it (C. Drioli), philipp.aichinger@meduniwien.ac.at (P. Aichinger).

https://doi.org/10.1016/j.bspc.2020.102309
Received 31 May 2020; Received in revised form 9 October 2020; Accepted 24 October 2020
Available online 5 November 2020

Voice source analysis through numerical models of the vocal folds oscillatory patterns is nowadays an established research field, and reliable glottal models of different accuracy and complexity are available that mimic the mechanics of the folds [13–17]. An efficient kinematic model is the phase-delayed overlapping sinusoids (PDOS) model which enables modelling vertical, but not sagittal phase differences in an explicit way [18,19]. Also, existing biomechanical models of the folds based on lumped elements (masses, springs and dampers) either cannot reproduce sagittal phase differences (as the two-mass model) or do not enable control of the phase difference in an easy and direct way (as, for example, in the 3D 16-mass model, in which the phase difference depends on a number of parameters). Another model enabling sagittal differences but no direct control thereof uses empirical eigenfunctions, which relate to modes of vocal fold vibration [20,21]. Conversely, we propose an approach to fold edge modelling that allows direct control over the amount of phase delay between oscillations at posterior and anterior parts of the glottis. Control of lateral phase differences is enabled via unbalancing the left and right vocal folds' natural frequencies. Thus, our edge displacement model is driven by a low-dimensional lumped-element scheme, previously introduced in [22] and extended here to enable sagittal phase differences using delay lines. This allows to keep the dynamic model simple and computationally efficient. In summary, the novelty and aim of this work is to extend the previous model by adding delay lines to enable direct control of sagittal phase differences, proposing automatic parameter estimation, and testing the model using 30 snippets of in vivo high-speed videos (HSVs). We recently presented the principle of our approach and results of 5 of the 30 snippets [23,24], and provide more thorough explanations and analyses here.

The remainder of the article is structured as follows. The lumped/distributed elements mechanical model is described in Section 2. In Section 3, a description of the HSVs used to assess the model is given, and in Section 4, a set of GAW asymmetry parameters are defined. In Section 5, the reproduction of oscillatory patterns observed in the data is performed by manual tuning the model parameters, and in Section 6 an automatic parameter optimization algorithm is illustrated and assessed on the data. In Section 7 results are discussed and conclusions are drawn.

2. Description of the model

The modelled vocal fold vibration is driven by mass–spring systems with stiffness k, damping r and mass m. Possible lateral asymmetry is taken into account by using two different single-mass systems, one for each fold. The displacements x of the masses from its resting position are obtained via computing the exerted pressure P_m and the exerted force F from the transglottal airflow U and the inferior glottal area A_i, using Newton's and Bernoulli's laws.

$$\left\{\begin{array}{l} m^{\alpha}\ddot{x}^{\alpha}(t) + r^{\alpha}\dot{x}^{\alpha}(t) + k^{\alpha}x^{\alpha}(t) = F^{\alpha}(t) \\ F^{\alpha}(t) = P_m(t)\cdot S^{\alpha} \\ P_m(t) = P_l - \frac{1}{2}\rho\frac{U^2(t)}{A_i^2(t)}, \end{array}\right. \tag{1}$$

where $S = Y \cdot Z$ is the equivalent fold surface on which the pressure is exerted, ρ is the air density, P_l is the lung pressure, the superscripts $\alpha \in \{l, r\}$ are used to indicate the left and the right elements respectively, and Y and Z are the length and the thickness of the folds respectively. The force F^l is perpendicular to S^l and oriented to the left, whilst F^r is perpendicular to S^r and oriented to the right. The control parameters in (1) are the ones that are independent of t, except ρ, which is a constant. In particular, the nine control parameters are two masses m^{α}, two dampings r^{α}, two stiffnesses k^{α}, the vocal fold length Y, the vocal fold thickness Z, and the lung pressure P_l.

Two additional control parameters are the vertical and sagittal delays, resulting in a total of 11 control parameters. The vertical phase difference of the vibration of the fold edges is essential for the modelling of self-sustained oscillations. It is represented by a delay line using a delay parameter τ_{vert} of the displacement of the fold along the vertical axis. The propagation of the displacement along the sagittal axis is represented by a distributed element introducing a delay τ_{sag} (see Fig. 1).

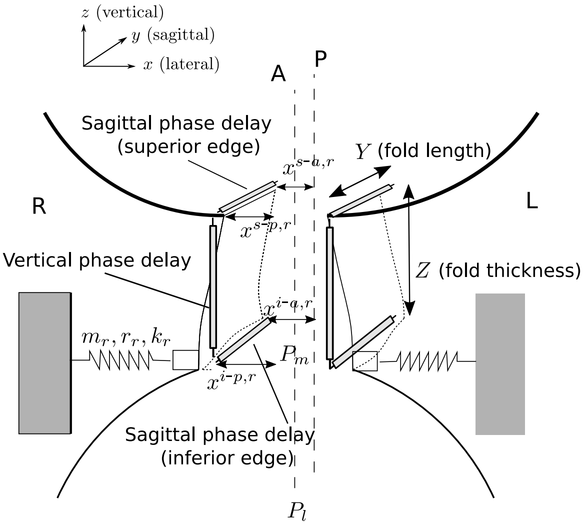

Fig. 1. Schematic view of the model: the vertical (inferior–superior) and sagittal (anterior–posterior) phase differences of the fold displacement are modelled using three propagation lines for each fold.

Inspired by the modelling approach in [15], and coherently with the previous investigation in [22], we align the source of the displacement with a given point on the fold surface, and propagate the motion along the surface like a travelling wave. Let $x^{i\text{-}p}$ be the posterior displacement of the fold at the entrance of the glottis (inferior, posterior edge), and $x^{s\text{-}p}$ the displacement at the exit (superior, posterior edge). A collision model f_X distorts the folds displacement and adds the offsets $x_0^l = -x_0$ and $x_0^r = x_0$, which correspond to the resting positions.[1] The displacements are computed using the vertical delay parameter τ_{vert} as

$$\left\{\begin{array}{l} x^{i\text{-}p,\alpha}(t) = f_X(x^{\alpha}(t), x_0) \\ \qquad\qquad = \left\{\begin{array}{ll} x^{\alpha}(t) + x_0 & \text{if } x^l(t) - x_0 < x^r(t) + x_0 \\ (x^l(t) + x^r(t))/2 & \text{otherwise} \end{array}\right. \\ x^{s\text{-}p,\alpha}(t) = f_X(x^{\alpha}(t - \tau_{vert}), x_0). \end{array}\right. \tag{2}$$

The sagittal phase difference of the displacement is also modelled here by using a distributed element. The delay of the displacement can either be oriented from the posterior end (P) to the anterior end (A) along the sagittal axis (to model P→A opening and closing), or from the anterior to the posterior end (to model A→P opening and closing). The sagittal delay parameter τ_{sag} is set according to the desired sagittal phase difference. The displacements of the anterior edges $x^{i\text{-}a}$ and $x^{s\text{-}a}$ are given by:

$$\left\{\begin{array}{l} x^{i\text{-}a,\alpha}(t) = x^{i\text{-}p,\alpha}(t - \tau_{sag}^{\alpha}) \\ x^{s\text{-}a,\alpha}(t) = x^{s\text{-}p,\alpha}(t - \tau_{sag}^{\alpha}) \end{array}\right. \tag{3}$$

We allow τ_{vert} and τ_{sag} to be unequal for the left and the right vocal fold. As shown in Fig. 2, the most anterior edge of the distributed element is connected to the anterior commissure (amplitude 0, fixed position) via straight lines, whereas the most posterior parts of the distributed elements move freely. This reflects the anatomy of

[1] Displacements are considered negative on the left of the sagittal plane, and positive on the right.

Table 1
The set of parameters used in the simulation of Fig. 2.

Symbol	Parameter	Value
f_0^l	L-fold natural frequency	210 Hz
f_0^r	R-fold natural frequency	180 Hz
$\tau_{sag}^l, \tau_{sag}^r$	Sagittal delay	1.81 ms
$\tau_{vert}^l, \tau_{vert}^r$	Vertical delay	0.13 ms
x_0^l, x_0^r	Rest positions	2.24 mm
Y	Vocal fold length	14 mm
Z	Vocal fold thickness	3 mm
S	Vocal fold medial surface	42 mm^2
ρ	Air density	1.15 kg m^{-3}
P_l	Lung pressure	1000 N m^{-2}

the vocal folds, which are connected at the anterior commissure, but disconnected at their posterior ends. Finally, a flow model converts the glottal area into the transglottal airflow. The length of glottis along the sagittal axis is sliced into N_Y coronal sections, each one of length $\delta_Y = Y/N_Y$. Due to the occasional hiding of the lower edges underneath the upper edges, the glottal width $w_j(t)$ at slice j is obtained from the inferior and superior edges' minima. The cross-sectional glottal area is computed as the length δ_Y times the glottal width. The GAW denoted as $A_g(t)$ is computed as the sum of all cross-sectional areas along the sagittal dimension. The flow is assumed to be proportional to the total glottal area.

$$\begin{cases} x_j^{i,\alpha}(t) = x^{i\text{-}p}(t - \frac{j \cdot \tau_{sag}^{\alpha}}{N_Y}),\ j = 1 \ldots N_Y \\ x_j^{s,\alpha}(t) = x^{s\text{-}p}(t - \frac{j \cdot \tau_{sag}^{\alpha}}{N_Y}),\ j = 1 \ldots N_Y \\ w_j(t) = \min\{x_j^{i,r}(t), x_j^{s,r}(t)\} - \min\{x_j^{i,l}(t), x_j^{s,l}(t)\} \\ a_j(t) = \delta_Y \cdot w_j(t) \\ A_g(t) = \sum_{j=1}^{N_Y} a_j(t) \\ U_g(t) = \sqrt{\frac{2P_l}{\rho}} \cdot A_g(t) \end{cases} \tag{4}$$

During collision, the cross-sectional areas $a_j(t) = 0$.

The Eqs. (1)–(4) are solved numerically after time-discretization to obtain an estimate of the glottal flow $U_g(nT)$, and of the folds' displacements $x_j^{i,\alpha}(nT)$ and $x_j^{s,\alpha}(nT)$, where n is the discrete time and the sampling interval $T = 1/22050$ Hz.

By comparing the last equation in (1) and the last equation in (4), it can be noticed that the inferior glottal area is

$$A_{gi}^2(t) = \frac{P_l}{P_l - P_m(t)} \cdot A_g^2(t). \tag{5}$$

From that it follows that $P_m(t) = 0$ if the glottis is divergent in the Z direction, whereas $P_m(t) > 0$ if it is convergent, given that $P_m(t) < P_l$ and $P_l > 0$. This means that the vocal folds are pushed laterally during convergence only, and that only expiratory phonation is considered.

The oscillatory patterns are visualized as if the folds were observed from above, so that they can be visually compared to HSV data. Fig. 2 illustrates the simulation from an example setup in which the natural frequencies of the left and right fold are different, i.e., $f_0^l = 210$ Hz, and $f_0^r = 180$ Hz. The two sagittal phase delays are $\tau_{sag}^l = \tau_{sag}^r = 1.8$ ms (the whole set of parameters of this configuration is reported in Table 1). Given that the resulting glottal cycle length is approximately 5 ms, the maximal sagittal phase difference is 0.36 cycles, or approximately 130 degrees. The resulting oscillation is characterized by paramedian collision caused by lateral phases differences. Note that the natural frequency of a mass–spring system is $f_0 = \sqrt{k/m}/2\pi$, but the resulting observed vibration frequency may happen to be different, due to coupling of the two vocal folds via the airstream. The direct control of delays, masses, and spring constants enables convenient simulation of various asymmetric vibratory patterns.

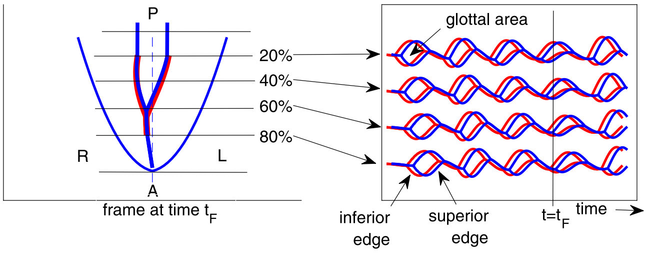

Fig. 2. Example of oscillatory pattern obtained by lateral mass unbalancing and laterally symmetric sagittal phase differences (blue stroke represent the superior edge of the folds, red stroke the inferior edge). Left panel: the vocal folds' contours as seen from above; Right panel: displacements of the vocal folds' contour points at four equidistant locations along the sagittal axis.

Table 2
Overview of the subjects, their pathologies, and the used HSV snippets.

Subjects	Pathologies	Snippets
1	Functional voice disorder	S1
2	Oedema	S2
3	Healthy	S3
4	Healthy	S4
5	Oedema	S5
6	Bamboo nodes	S5–S12
7	Laryngitis	S3–S16
8	Functional voice disorder	S17–S18
9	Functional voice disorder	S19–S30

3. Corpus description

A total of 30 video snippets are selected from the Laryngeal High-Speed Video Database of Pathological and Non-Pathological Voices [25]. The videos were recorded with a Richard Wolf Endocam 5562, and audio signals were recorded simultaneously. Video snippets are selected which (i) are auditory rough, i.e., R of GRBAS >= 1 [26,27], and (ii) have time-invariant sagittal phase differences. R is defined as the 'audible impression of irregular glottic pulses' [27]. The video snippets are between 11 and 40 ms long, corresponding to 45 and 160 samples at the frame rate of 4 kHz. For five video snippets, model parameters are adjusted manually and qualitative comparisons are made, including visual comparisons of vocal fold displacements. For the remaining 25 video snippets, automatic parameter optimization is carried out via comparing the observed and modelled vibration frequencies and time delays. The first five snippets S1–S5 are from five individual subjects, and the remaining snippets S6–S30 are from four additional subjects. Table 2 gives an overview of the subjects, pathologies, and snippets.

4. Timing asymmetry parameters

To quantitatively assess the model's capability of fitting sagittally asymmetric vibration, a set of measures related to the GAW are computed.[2] To this aim, we refer to the left and the right hemi-GAWs as $A_g^L(t)$ and $A_g^R(t)$, which are defined as the time-varying areas of the left and the right half of the glottis respectively, and satisfy $A_g^R(t) - A_g^L(t) = A_g(t)$. Similarly, we refer to the anterior and the posterior hemi-GAWs as $A_g^A(t)$ and $A_g^P(t)$, which are defined as the time-varying areas of the anterior and the posterior half of the glottis respectively, and satisfy $A_g^A(t) + A_g^P(t) = A_g(t)$. A schematic illustration of the hemi-GAWs is shown in Fig. 3. In each cycle, the time instants corresponding to maximum areas, i.e., T_c^R, T_c^L, T_c^A, T_c^P with cycle index c, are obtained by picking peaks as illustrated in Fig. 4. Finally, lateral and sagittal timing differences are obtained as $\Delta T_c^{LR} = T_c^R - T_c^L$ and $\Delta T_c^{AP} = T_c^A - T_c^P$

[2] Videos are graphically segmented to obtain the spatio-temporal vibration patterns, i.e., phonovibrograms, as described in [28].

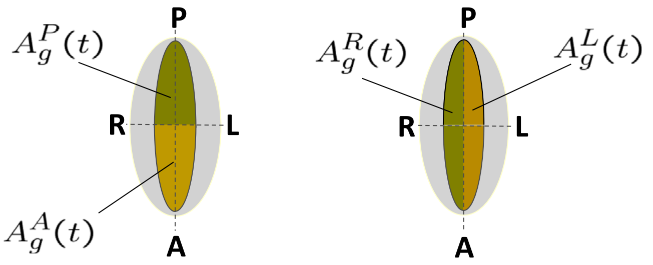

Fig. 3. Schematic representation of sagittal hemi-GAWs (left), and lateral hemi-GAWs (right).

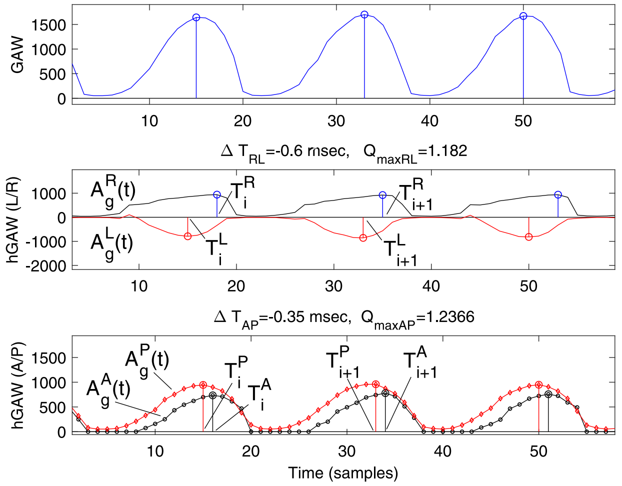

Fig. 4. Result of peak picking procedure on GAW (top), lateral hemi-GAWs (centre), and sagittal hemi-GAWs (bottom).

respectively. In a nutshell, the introduced parameters provide means of comparing times of maxima of hemi-GAWs, which reflect lateral and sagittal phase differences.

5. Manual parameter adjustment: Case studies

In this section, the model is assessed qualitatively by empirically tuning its parameters to replicate some special oscillatory patterns observed in high-speed videoendoscopic data, in which phase differences and paramedian collisions due to the mass or stiffness/tension unbalancing are observed. First, we aim at reproducing observed oscillatory patterns of the folds observed in the HSV qualitatively, hence we do not aim to achieve copy-synthesis yet. An estimate of the period is obtained from the video data by first obtaining the GAW signals from HSVs and then performing an automatic search for the GAW peaks. The parameters of the mass–spring system are then tuned to match the estimated GAW period and the resulting oscillation period of the model. Where not specified differently, we use $Y = 14$ mm and $Z = 3$ mm in the simulations. The parameter tuning strategy consists essentially in a set of simple rules: for symmetric patterns, 1. increase(decrease) the fold natural frequencies $f_0^l = f_0^r$ to increase(decrease) the resulting pitch, 2. increase(decrease) the sagittal delay parameter $\tau_{sag}^l = \tau_{sag}^r$ to increase(decrease) the sagittal phase delay, 3. use negative or positive τ_{sag} values to get a P→A or A→P phase delay, respectively; for asymmetric patterns, 1. decrease f_0^l with respect to f_0^r to obtain a positive ΔT^{LR} (or increase it for a negative ΔT^{LR}), 2. increase(decrease) τ_{sag}^l with respect to τ_{sag}^r if a visual pattern is observed in the HSV, with faster(slower) propagation on the left fold than on the right fold, 3. use same criterion as above to set the sign of the τ_{sag} parameters.

In Fig. 5, P→A sagittal opening is simulated by using laterally symmetric sagittal phase differences. The left and right folds' oscillating frequencies are $f_0^l = f_0^r = 210$ Hz, and the two sagittal phase delays are $\tau_{sag}^l = \tau_{sag}^r = 1.5$ ms, resulting in a sagittal phase delay of 0.315 cycles, or approximately 113.4 degrees.

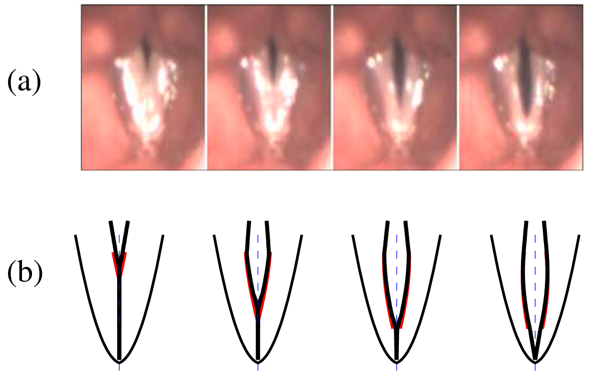

Fig. 5. Selections of frames within the opening phase from (a) a high-speed video (snippet: $S1$), and (b) the model simulation. A posterior to anterior opening pattern is observed due to a laterally symmetric sagittal phase difference.

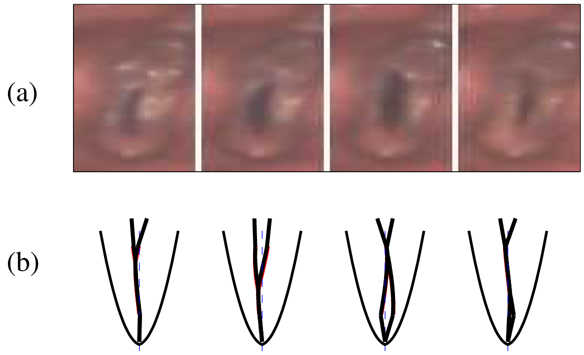

Fig. 6. (a) A selection of frames within one cycle from an high-speed video (subject: $S2$), and (b) a selection from the model simulation. The vibration pattern is characterized by both lateral asymmetry, as well as sagittal phase delay.

Fig. 6 shows a complicated vibration pattern affected by lateral asymmetry, as well as sagittal phase delay (upper panel), and its imitation provided by the model (lower panel). The left and right natural frequencies are $f_0^l = 125.0$ Hz and $f_0^r = 115.0$ Hz respectively, and the two sagittal phase delays $\tau_{sag}^l = \tau_{sag}^r = 1.36$ ms. The observed frequency resulted in 114.5 Hz.

A selection of five recordings (S1–S5) from the Laryngeal High-Speed Video Database of Pathological and Non-Pathological Voices described in [25] is used. The HSVs contain sagittal and lateral phase differences. The peak picking based hemi-GAW analysis is applied to both the natural video data, and the model simulations thereof. Results are reported in Table 3 (average over 10 periods).

Snippet $S1$ is characterized by a small lateral asymmetry and a small sagittal phase difference (double pulsing is also observable as a small secondary peak in the sagittal hemi-GAWs, however we disregard this component in the present investigation). In the model, the left and right folds' natural frequencies are slightly unbalanced and the two sagittal phase delays are $\tau_{sag}^l = \tau_{sag}^r = 0.32$ ms.

In snippet $S2$, shown in Fig. 6, a complicated vibration pattern due to lateral mass differences and sagittal phase differences is observed. The left and right folds' natural frequencies are $f_0^l = 125$ Hz and $f_0^r = 115$ Hz respectively to obtain a negative ΔT^{LR} value, and the two sagittal phase delays $\tau_{sag}^l = \tau_{sag}^r = 1.36$ ms.

Snippet $S3$ is characterized by a moderate $P \rightarrow A$ sagittal delay, resulting in a negative ΔT^{AP}, and a moderate lateral asymmetry. The natural frequencies are $f_0^l = 294$ Hz and $f_0^r = 346$ Hz to obtain a

Table 3
Hemi-GAW-based asymmetry analysis. The parameters of the model were tuned manually. The vibration frequency of the model is reported as $\overline{F_0}$, the natural frequencies are reported as f_0^l, and f_0^r, and the sagittal delay parameters are reported as τ_{sag}^l and τ_{sag}^r. The delays between the left and right vocal folds, as well as between their anterior and posterior ends are reported as ΔT^{LR} and ΔT^{AP} respectively.

Snippet	HSV data GAW analysis			Model output GAW analysis			Model parameters		
	F_0 (Hz)	ΔT^{LR} (ms)	ΔT^{AP} (ms)	$\overline{F_0}$ (Hz)	$\overline{\Delta T^{LR}}$ (ms)	$\overline{\Delta T^{AP}}$ (ms)	f_0^l (Hz)	f_0^r (Hz)	$\tau_{sag}^l = \tau_{sag}^r$ (ms)[a]
S1	181.0	0.20	0.15	185.4	0.30	0.16	195.0	205.0	0.32
S2	117.0	−0.98	0.64	114.5	−0.78	0.70	125.0	115.0	1.36
S3	285.0	0.65	−0.26	284.3	0.43	−0.25	294.0	346.0	−0.49
S4	222.0	0.01	−0.40	232.9	0.00	−0.45	240.0	240.0	−0.91
S5	222.0	−0.60	−0.35	218.2	−0.42	−0.32	230.0	230.0	−0.05(L),−0.95(R)

[a] Note: a minus sign in this column means P→A sagittal propagation.

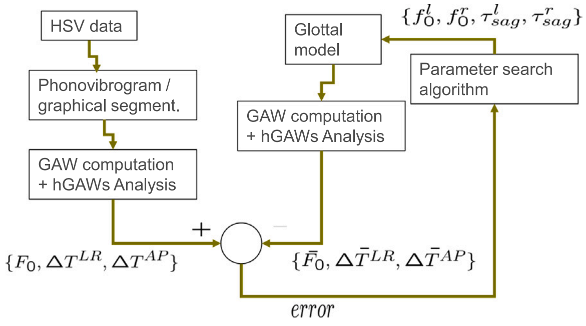
HSV data
Phonovibrogram / graphical segment.
GAW computation + hGAWs Analysis
Glottal model
GAW computation + hGAWs Analysis
Parameter search algorithm
$\{f_0^l, f_0^r, \tau_{sag}^l, \tau_{sag}^r\}$
+
−
$\{F_0, \Delta T^{LR}, \Delta T^{AP}\}$
$\{\bar{F}_0, \Delta\bar{T}^{LR}, \Delta\bar{T}^{AP}\}$
error

Fig. 7. Scheme of the automatic parameter estimation.

positive ΔT^{LR} value, and the two sagittal phase delays are $\tau_{sag}^l = \tau_{sag}^r = 0.45$ ms. A negative value of ΔT^{AP} is obtained by propagating the fold displacement from posterior to anterior.

Snippet $S4$ is characterized by a $P \rightarrow A$ sagittal delay and lateral symmetry. The left and right folds' oscillating frequencies are $f_0^l = f_0^r = 240$ Hz, and the two sagittal phase delays are $\tau_{sag}^l = \tau_{sag}^r = -0.9$ ms.

Finally, snippet $S5$ has a laterally asymmetric negative sagittal phase delay, reproduced by setting $\tau_{sag}^l = -0.05$ ms, and $\tau_{sag}^r = -0.95$ ms.

6. Automatic parameter optimization

The results obtained by empirical tuning of the model suggest that estimates of the vocal folds' natural frequencies f_0^l and f_0^r, as well as of the sagittal phase delay parameters τ_{sag}^l and τ_{sag}^r, might be obtained by minimizing a suitable error criterion based on the difference of hemi-GAW parameters ΔT^{LR} and ΔT^{AP}. The proposed parameter optimization scheme is illustrated in Fig. 7.

In the scheme and in what follows, the symbols F_0, ΔT^{LR}, and ΔT^{AP} represent quantities observed in the data through video analytics procedures, whereas the same symbols with an overline represent quantities resulting from the numerical simulation of the model. For what concern oscillation frequencies, f_0^l and f_0^r are the natural frequencies of the mechanical oscillators representing the folds, $\overline{F_0}$ is the resulting model oscillation frequency, and F_0 is the desired oscillation frequency as observed in the data.

The parameter optimization algorithm performs an iterative search in which the error terms $err_{F0} = (F_0 - \overline{F_0})^2$, $err_{LR} = (\Delta T^{LR} - \overline{\Delta T^{LR}})^2$, and $err_{LR} = (\Delta T^{AP} - \overline{\Delta T^{AP}})^2$, are summed into the error criterion using a linear scalarization model, i.e. $err = w_1 \cdot err_{F0} + w_2 \cdot err_{LR} + w_3 \cdot err_{AP}$, where w_1, w_2, and w_3 are weights that allow to balance the importance of the error components. In our experiments, we have set $w_1 = 1/F_0$, so that the pitch component is a relative error, and we have empirically set $w_2 = w_3 = 1/0.1$, since 0.1 is a typical value assumed for the ΔT parameters. With this choice, the range of the two unbalancing error components and the range of the relative pitch error become comparable. In each iteration, the three parameters are minimized one after the other according to:

$$
\begin{aligned}
\hat{f}_0 &= \underset{f_0}{\operatorname{argmin}}\ err \\
\hat{\delta f}_0 &= \underset{\delta f_0}{\operatorname{argmin}}\ err \\
\hat{\tau}_{sag} &= \underset{\tau_{sag}}{\operatorname{argmin}}\ err
\end{aligned}
\tag{6}
$$

where f_0^l and f_0^r are initialized as $f_0^l = f_0^r = f_0$. Afterwards, f_0^l and f_0^r differ by $2\delta f_0$, i.e., $f_0^l = f_0 - \delta f_0$ and $f_0^r = f_0 + \delta f_0$; and τ_{sag}^l and τ_{sag}^r are assumed to be equal during the minimization of err_{AP}, i.e., $\tau_{sag}^l = \tau_{sag}^r = \tau_{sag}$. The pseudocode of the iterative procedure is listed in Algorithm 1. In summary, first the natural frequency is tuned to minimize the error, since the frequency is the parameter that has the most influence on the error. Second, the natural frequencies of the vocal folds are unbalanced by offsetting them, which primarily effects the lateral phase difference. Third, the sagittal delay is tuned. It was chosen to tune the model parameters one by one instead of in parallel to keep the search fast. This works particularly well with our model because the individual parameters mostly effect individual components of the model error in a decoupled way. In particular, the average of the left and the right vocal fold's natural frequencies primarily effected the vibration frequency, the difference between the left and the right vocal folds' natural frequencies primarily effected the lateral phase difference, and the sagittal delay parameter primarily effected the sagittal phase difference. The tuning is repeated in a loop until convergence. The chosen strategy has proven to be effective as compared to a few alternatives we also experimented with.

Algorithm 1 Iterative parameter estimation algorithm

```
Initialization:
Set $f_0 = F_0$,
Set $\delta f_0 = 0$,
Set $\tau_{sag} = 0$
while $err > \epsilon$ do
    Tune the folds: $f_0 = \underset{f_0^{l,r}}{\operatorname{argmin}}\ err$
    Unbalance the folds: $\delta f_0 = \underset{\delta f_0}{\operatorname{argmin}}\ err$
    Set A-P phase delay: $\tau_{sag} = \underset{\tau_{sag}^{l,r}}{\operatorname{argmin}}\ err$
end while
```

A comparison between manual and automatic parameter tuning performed on subjects S1–S5 is shown in Fig. 8. The number of iteration for the tuning was limited to a maximum of 7, and the value of the error threshold parameter ϵ for early stop was set to 0.1. With this settings, the iterations required for the five cases varied from 2 to 7. In general automatic tuning produced comparable or better results, except for one case (i.e., for subject S3, automatic pitch tuning did not reach the desired accuracy within the maximum number of iterations).

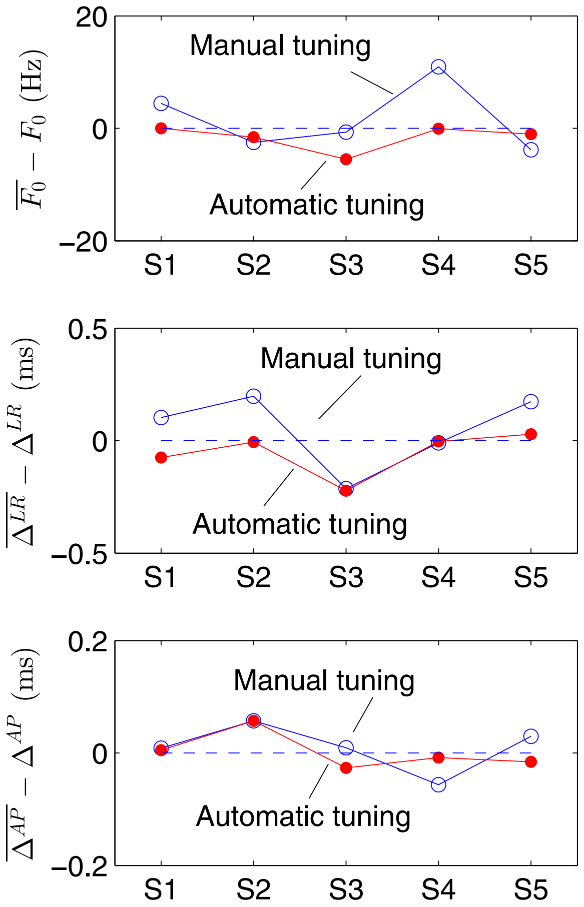

Fig. 8. Parameter tuning error: comparison between manual tuning (circle marker) and automatic tuning (dot marker), performed on subjects S1–S5.

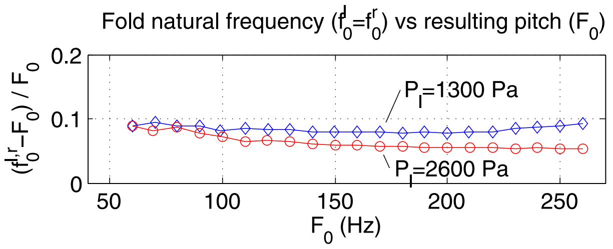

Fig. 9. Tuning of the natural frequency of the folds for different vibration frequencies in the range [60–260] Hz, and different subglottal pressures.

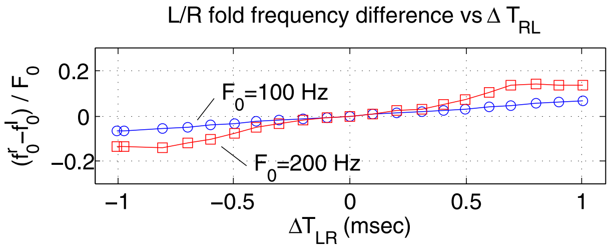

Fig. 10. Computation of the natural frequency unbalancing of the two folds to obtain different ΔT_{LR} values, in the range [−1, 1] ms.

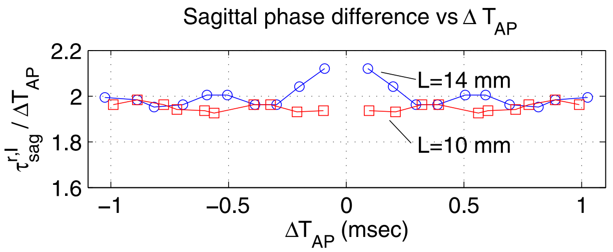

Fig. 11. Computation of the sagittal phase delays $\tau^{l,r}_{sag}$ to obtain different ΔT_{AP} values, in the range [−1, 1] ms.

6.1. Exploration of the model behaviour

In this section, the model's behaviour is explored in terms of natural frequencies f_0^α, vibration frequency F_0, delay parameters τ^α_{sag}, and observed delay times ΔT_{AP} and ΔT_{LR}. The tuning of the model parameters was performed by the automatic parameter optimization described in Algorithm 1, which minimizes the difference between observed parameters and target parameters. In particular, in Figs. 9, 10, and 11, target values shown on the x-axes are tried to be matched automatically by only tuning single model parameters.

Fig. 9 shows the relative offset of the natural frequency of the folds for different vibration frequencies in the range $[60-260]$ Hz. Subglottal pressure was set to 1300 Pa and to 2600 Pa. It is interesting to note the difference between the natural frequency of the folds and the resulting oscillation of the model, due to the nonlinear feedback structure of the processing loop.

Fig. 10 shows the computation of the natural frequency unbalancing of the two folds to obtain different ΔT_{LR} values, in the range [−1, 1] ms and for two F_0 values, 100 and 200 Hz.

Fig. 11 shows the computation of the sagittal phase delays $\tau^{l,r}_{sag}$ to obtain different ΔT_{AP} values, in the range [−1, 1] ms and for two different fold length values. In this case, it is to note that the relation between the τ_{sag} parameters and the resulting ΔT_{AP} is roughly proportional, and depending on the vocal fold length Y only partially.

Figs. 10 and 11 relate the asymmetry parameters to the model parameters to develop a more intuitive and mechanistic understanding of the parameters and their relationship.

6.2. Evaluation using the corpus

The automatic parameter optimization described in Algorithm 1 was carried out on the 30 video snippets (S1–S30) from nine different subjects, by comparing the observed and modelled vibration frequencies and time delays. The snippets were first analysed to extract the hemi-GAW parameters of timing and amplitude asymmetry, and the model parameters were tuned by the iterative algorithm. The results are reported in Table 4.

Figs. 12, 13, and 14 show Bland Altman Plots of the estimated and reference vibration frequency F_0, and sagittal phase asymmetry parameters ΔT^{LR} and ΔT^{AP}. The x-axes show the means of the pairs of estimated and reference values, and the y-axes show their differences. The blue lines and shaded areas reflect the means of the differences, and its confidence intervals, reflecting the estimation bias. The red lines and shaded areas are the upper and lower 95%-limit of agreement, and their confident intervals, reflecting random estimation errors. The blue areas do not include 0 difference for the frequency, but include 0 difference for the timing asymmetry parameters. Hence, the estimate of frequency is biased by approximately 9.28 Hz, but the timing asymmetry estimates appear to be unbiased. The limits of agreement reflect the intervals within which 95% of the estimation errors are expected. They are $[-32.17, +50.72]$ Hz, $[-0.54, +0.47]$ ms, and $[-0.85, +0.88]$ ms for the frequency, lateral phase difference, and sagittal phase difference.

Table 4
Model tuned on hGAW-based asymmetry data.

Snippet	HSV data GAW analysis			Model output GAW analysis			Model parameters		
	F_0 (Hz)	ΔT^{LR} (ms)	ΔT^{AP} (ms)	$\overline{F_0}$ (Hz)	$\overline{\Delta T^{LR}}$ (ms)	$\overline{\Delta T^{AP}}$ (ms)	f_0^l (Hz)	f_0^r (Hz)	$\tau^l_{sag} = \tau^r_{sag}$ (ms)[a]
$S1$	181.0	0.200	0.150	181.0	0.125	0.155	191.3	194.8	0.319
$S2$	117.0	−0.980	0.640	115.4	−0.986	0.697	130.6	116.1	1.390
$S3$	285.0	0.650	−0.260	279.5	0.427	−0.286	287.1	309.1	−0.568
$S4$	222.0	0.010	−0.400	221.9	0.006	−0.408	230.4	230.5	−0.794
$S5$	222.0	−0.600	−0.350	220.9	−0.571	−0.366	246.9	225.6	−0.713
$S6$	267.0	−0.188	−0.375	261.0	−0.189	−0.37	270.0	267.0	−0.732
$S7$	364.0	0.167	−0.917	369.0	0.159	−0.601	368.0	370.0	−1.16
$S8$	267.0	−0.0625	−0.25	262.0	−0.0756	−0.257	277.0	276.0	−0.481
$S9$	250.0	−0.667	−0.917	243.0	−0.612	−0.741	255.0	244.0	−1.47
$S10$	364.0	0.286	−0.0714	352.0	0.272	−0.187	372.0	379.0	−0.368
$S11$	333.0	0.575	−0.15	274.0	0.302	−1.12	348.0	371.0	−2.26
$S12$	333.0	0.325	0.175	320.0	0.324	0.181	341.0	348.0	0.352
$S13$	308.0	0.361	0.278	302.0	0.35	0.285	316.0	323.0	0.568
$S14$	235.0	0.722	0	224.0	0.499	0.172	242.0	249.0	0.319
$S15$	286.0	0.477	−0.5	267.0	0.491	−0.454	281.0	292.0	−0.891
$S16$	286.0	0.438	−0.125	277.0	0.295	−0.181	298.0	305.0	−0.35
$S17$	308.0	0.306	0.417	350.0	0.801	1.03	324.0	329.0	1.79
$S18$	308.0	0.417	0.222	289.0	0.596	0.46	302.0	310.0	0.891
$S19$	308.0	−0.4	0.05	290.0	−0.305	0.168	312.0	304.0	0.298
$S20$	250.0	−0.917	0	236.0	−0.925	0.461	247.0	237.0	0.891
$S21$	308.0	−0.325	0.025	277.0	−0.302	0.181	303.0	287.0	0.344
$S22$	286.0	−0.611	−0.0833	264.0	−0.438	−0.34	281.0	271.0	−0.66
$S23$	235.0	−1.0	−0.125	199.0	−0.975	−0.306	221.0	205.0	−0.568
$S24$	286.0	−0.575	0.275	266.0	−0.567	0.454	272.0	265.0	0.891
$S25$	250.0	−1.22	0	235.0	−0.454	0.287	250.0	241.0	0.568
$S26$	308.0	−0.531	0.156	280.0	−0.469	0.28	303.0	291.0	0.535
$S27$	267.0	−1.08	−0.0556	257.0	−0.537	−0.272	281.0	268.0	−0.534
$S28$	308.0	−0.556	0.278	372.0	−1.28	−1.63	318.0	305.0	2.26
$S29$	286.0	−0.95	0.3	265.0	−0.627	0.559	281.0	265.0	1.14
$S30$	308.0	−0.675	0.275	308.0	−0.627	0.454	315.0	308.0	0.891

[a]Note: a minus sign in this column means P→A sagittal propagation.

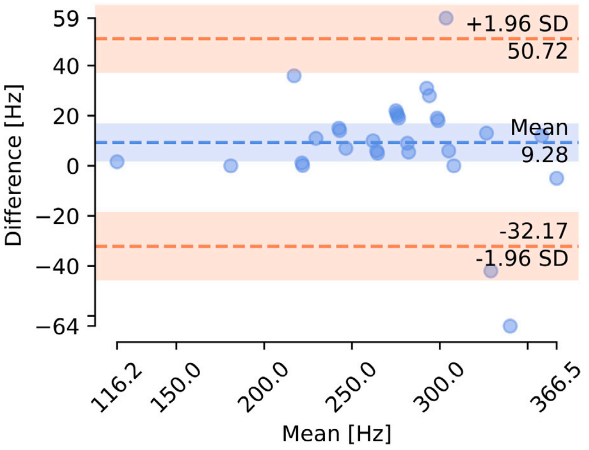

Fig. 12. Bland Altman plot related to F_0 matching. The y-axis shows the difference of the vibration frequency F_0 observed in the high-speed video, and the vibration frequency $\overline{F_0}$ observed in the output of the proposed model. The x-axis shows the pairwise means of F_0 and $\overline{F_0}$.

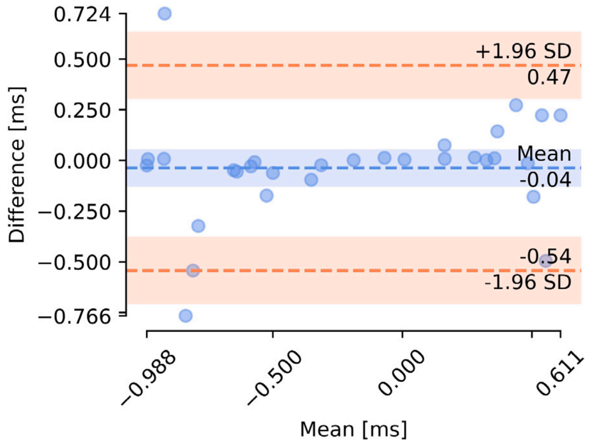

Fig. 13. Bland Altman plot related to ΔT_{LR} matching. The y-axis shows the difference of the lateral timing asymmetry parameter ΔT_{LR} observed in the high-speed video, and the lateral timing asymmetry parameter $\overline{\Delta T_{LR}}$ observed in the output of the proposed model. The x-axis shows the pairwise means of ΔT_{LR} and $\overline{\Delta T_{LR}}$.

7. Discussion and conclusions

We discussed the modelling of sagittal phase differences in vocal folds oscillations by means of a lumped and distributed elements vocal fold model. The lumped components of the model represent the fold mass and the aerodynamic interaction with the glottal airflow, whilst the distributed elements (delay lines) represent the vertical and the sagittal propagation of the fold displacement. Although the model retains the main features of a physical model, it also relies on several simplifications and modelling solutions which are not strictly physically justified, and it should therefore be classified as a mathematical model. We address the reproduction of the oscillatory patterns observed in high-speed video recordings of the folds, including vertical and sagittal phase differences and left–right fold mass unbalancing. The model was assessed qualitatively, by empirically tuning its parameters to replicate some oscillatory patterns observed in high-speed videoendoscopic data. Kinematic asymmetry measures regarding timing were derived from the peak analysis of the hemi-GAWs and were compared to those obtained from the HSV data. The comparisons suggest that it is possible to independently control the lateral and sagittal phase differences by tuning the left and right mass unbalancing, and the sagittal propagation delay respectively. Also, the model parameters were shown to be adjustable automatically, using an iterative search procedure.

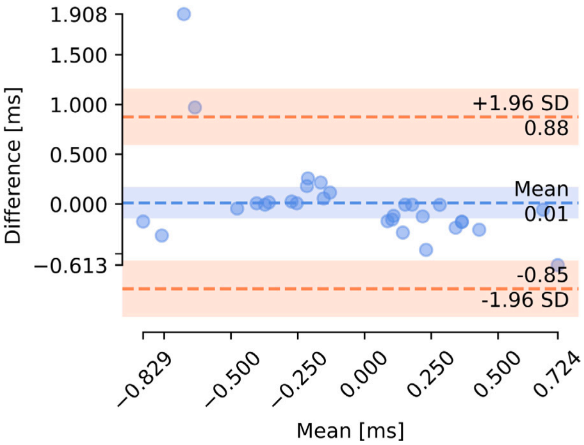

Fig. 14. Bland Altman plot related to ΔT_{AP} matching. The y-axis shows the difference of the sagittal timing asymmetry parameter ΔT_{AP} observed in the high-speed video, and the sagittal timing asymmetry parameter $\overline{\Delta T_{AP}}$ observed in the output of the proposed model. The x-axis shows the pairwise means of ΔT_{AP} and $\overline{\Delta T_{AP}}$.

Limitations of the study and suggestions for future work include the following. First, it was assumed that model parameters do not change over time. In particular, frequencies and timing differences often change slowly, reflecting intonation-like variations, but also fast, reflecting perturbation of the voice source. Regarding the latter, modulations may exist even on a pulse-to-pulse time scale. Since voice quality may be strongly affected by fast modulations, they should be considered in future studies. Second, we assumed that the vocal fold contours are reflected by upper and lower margin only. This restricts modelled GAWs to have sharp peaks only, whereas smooth vocal fold contours would allow the simulation of rounded lateral peaks. Third, the mass–spring system is driven as if there was no vocal tract. The basic assumption to achieve this is that the pressure P_l, which is given relative to the supraglottal threshold, is assumed to be constant. This is a simplification which nevertheless allowed the creation of a quite realistically looking motion that is driving the delay-lines included in the model. As a consequence, our model is not capable of rippling and skewing of the glottal pulse shape caused by the feedback of the vocal tract. Fourth, one would have expected from first principles that the speed of the travelling waves depend on the vocal folds' natural frequencies. Thus, a lateral unbalancing of the natural frequencies could result in laterally different delay parameters. However, for the purpose of automatic parameter estimation, we assumed for now that the sagittal delay parameters are equal for the two vocal folds. Finally, one could argue that using the RMS difference of phonovibrograms instead of timing and frequency differences only would in the future enable a more complete evaluation of the vibration patterns produced by the model.

CRediT authorship contribution statement

Carlo Drioli: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Investigation, Visualization. **Philipp Aichinger:** Conceptualization, Methodology, Data curation, Writing - original draft, Resources, Writing - review & editing, Visualization, Funding acquisition.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.bspc.2020.102309.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF): KLI 722-B30. The authors want to thank Jean Schoentgen for a fruitful discussion and the Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Germany, for providing the Glottis Analysis Tools (GAT).

References

[1] J. Švec, F. Šram, H. Schutte, Videokymography in voice disorders: What to look for? Ann. Otol. Rhinol. Laryngol. 116 (3) (2007) 172–180, http://dx.doi.org/10.1177/000348940711600303.

[2] H.S. Bonilha, D. Deliyski, T. Gerlach, Phase asymmetries in normophonic speakers: visual judgments and objective findings, Am. J. Speech-Lang. Pathol. 17 (4) (2008) 367–376, http://dx.doi.org/10.1044/1058-0360(2008/07-0059), Phase.

[3] R. Orlikoff, M.E. Golla, D. Deliyski, Analysis of longitudinal phase differences in vocal-fold vibration using synchronous high-speed videoendoscopy and electroglottography, J. Voice 26 (6) (2012) 816.e13–816.e20, http://dx.doi.org/10.1016/j.jvoice.2012.04.009.

[4] A. Yamauchi, H. Imagawa, K.I. Sakakibara, H. Yokonishi, T. Nito, T. Yamasoba, N. Tayama, Characteristics of vocal fold vibrations in vocally healthy subjects: Analysis with multi-line kymography, J. Speech Lang. Hearing Res. 57 (2) (2014) http://dx.doi.org/10.1044/2014_JSLHR-S-12-0269.

[5] A. Yamauchi, H. Yokonishi, H. Imagawa, K.I. Sakakibara, T. Nito, N. Tayama, Quantitative analysis of vocal fold vibration in vocal fold paralysis with the use of high-speed digital imaging, J. Voice 30 (6) (2016) 766.e13–766.e22, http://dx.doi.org/10.1016/j.jvoice.2015.10.015.

[6] J. Švec, H. Schutte, Videokymography: high-speed line scanning of vocal fold vibration, J. Voice 10 (2) (1996) 201–205.

[7] T. Wittenberg, M. Tigges, P. Mergell, U. Eysholdt, Functional imaging of vocal fold vibration: Digital multislice high-speed kymography, J. Voice 14 (3) (2000) 422–442, http://dx.doi.org/10.1016/S0892-1997(00)80087-9.

[8] J. Lohscheller, U. Eysholdt, H. Toy, M. Dollinger, Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics, IEEE Trans. Med. Imaging 27 (3) (2008) 300–309, http://dx.doi.org/10.1109/TMI.2007.903690.

[9] A.P. Pinheiro, D.E. Stewart, C.D. Maciel, J.C. Pereira, S. Oliveira, Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling, Digit. Signal Process. 22 (2) (2012) 304–313, http://dx.doi.org/10.1016/j.dsp.2010.11.002.

[10] M. Döllinger, P. Gómez, R.R. Patel, C. Alexiou, C. Bohr, A. Schützenberger, Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videoendoscopy, PLoS One 12 (11) (2017) 1–26, http://dx.doi.org/10.1371/journal.pone.0187486.

[11] C. Drioli, G.L. Foresti, Accurate glottal model parametrization by integrating audio and high-speed endoscopic video data, Signal Image Video Process. 9 (2015) 1451–1459.

[12] T. Murtola, P. Alku, Indicators of anterior – posterior phase difference in glottal opening measured from natural production of vowels, J. Acoust. Soc. Am. 148 (2) (2020) EL141–EL146, http://dx.doi.org/10.1121/10.0001722.

[13] K. Ishizaka, J.L. Flanagan, Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell Syst. Tech. J. 51 (6) (1972) 1233–1268.

[14] T. Koizumi, S. Taniguchi, S. Hiromitsu, Two-mass models of the vocal cords for natural sounding voice synthesis, J. Acoust. Soc. Am. 82 (4) (1987) 1179–1192.

[15] I.R. Titze, The physics of small-amplitude oscillations of the vocal folds, J. Acoust. Soc. Am. 83 (4) (1988) 1536–1552.

[16] X. Pelorson, A. Hirschberg, R.R. van Hassel, A.P.J. Wijnands, Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model, J. Acoust. Soc. Am. 96 (6) (1994) 3416–3431.

[17] J.C. Lucero, J. Schoentgen, J. Haas, P. Luizard, X. Pelorson, Self-entrainment of the right and left vocal fold oscillators, J. Acoust. Soc. Am. 137 (4) (2015) 2036–2046, http://dx.doi.org/10.1121/1.4916601.

[18] I.R. Titze, F. Alipour, The Myoelastic Aerodynamic Theory of Phonation, National Center for Voice and Speech, Iowa City, 2006.

[19] J.J. Jiang, C.I. Chang, J.R. Raviv, S. Gupta, F.M. Banzali, D.G. Hanson, Quantitative study of mucosal wave via videokymography in canine larynges, Laryngoscope 110 (9) (2000) 1567–1573.

[20] D. Berry, H. Herzel, I.R. Titze, K. Krischer, Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, J. Acoust. Soc. Am. 95 (6) (1994) 3595–3604.

[21] J. Neubauer, P. Mergell, U. Eysholdt, H. Herzel, Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes, J. Acoust. Soc. Am. 110 (6) (2001) 3179–3192, http://dx.doi.org/10.1121/1.1406498.

[22] C. Drioli, A flow waveform-matched low-dimensional glottal model based on physical knowledge, J. Acoust. Soc. Am. 117 (5) (2005) 3184–3195.

[23] C. Drioli, P. Aichinger, Aerodynamics and lumped-masses combined with delay lines for modeling vertical and anterior-posterior phase differences in pathological vocal fold vibration, in: Proceedings of the Annual Conference of the International Speech Communication Association, Vol. 2019-Septe, INTERSPEECH, 2019, pp. 2503–2507, http://dx.doi.org/10.21437/Interspeech.2019-2338.

[24] C. Drioli, P. Aichinger, Modeling vertical and longitudinal phase differences observed in pathological vocal fold vibrations by means of a biomechanical lumped model, in: Models and Analysis of Vocal Emissions for Biomedical Applications, 2019, pp. 137–140.

[25] P. Aichinger, I. Roesner, M. Leonhard, D. Denk-Linnert, W. Bigenzahn, B. Schneider-Stickler, A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices, in: Proc. Int. Conf. Lang. Resour. Eval., Vol. 10, 2016, pp. 767–770, http://dx.doi.org/10.13140/RG.2.2.15467.34088.

[26] M. Hirano, Clinical Examination of Voice, Springer, New York, 1981.

[27] P.H. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. Van De Heyning, M. Remacle, V. Woisard, A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques, Eur. Arch. Oto-Rhino-Laryngol. 258 (2) (2001) 77–82.

[28] J. Lohscheller, U. Eysholdt, Phonovibrogram visualization of entire vocal fold dynamics, Laryngoscope 118 (4) (2008) 753–758.