International Journal of Applied Mechanics
Vol. 9, No. 5 (2017) 1750064 (18 pages)

DOI: 10.1142/S1758825117500648

World Scientific
www.worldscientific.com

# High-Speed Imaging to Study an Auto-Oscillating Vocal Fold Replica for Different Initial Conditions

A. Van Hirtum* and X. Pelorson
*Gipsa-lab, UMR CNRS 5216*
*Grenoble University, France*
**annemie.vanhirtum@gipsa-lab.grenoble-inp.fr*

Received 16 September 2016
Revised 13 December 2016
Accepted 14 December 2016
Published 26 July 2017

Experiments on mechanical deformable vocal folds replicas are important in physical studies of human voice production to understand the underlying fluid–structure interaction. At current date, most experiments are performed for constant initial conditions with respect to structural as well as geometrical features. Varying those conditions requires manual intervention, which might affect reproducibility and hence the quality of experimental results. In this work, a setup is described which allows setting elastic and geometrical initial conditions in an automated way for a deformable vocal fold replica. High-speed imaging is integrated in the setup in order to decorrelate elastic and geometrical features. This way, reproducible, accurate and systematic measurements can be performed for prescribed initial conditions of glottal area, mean upstream pressure and vocal fold elasticity. Moreover, quantification of geometrical features during auto-oscillation is shown to contribute to the experimental characterization and understanding.

*Keywords*: Fluid–structure interaction; initial vocal folds conditions; high-speed imaging; asymmetrical vocal folds; vocal fold pathology; experimental setup.

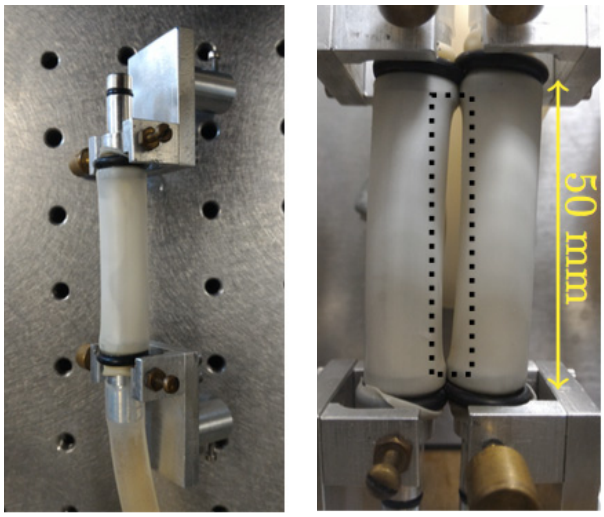
## 1. Introduction

The fluid–structure interaction in the human larynx between the vocal folds and airflow coming from the lungs is the main sound source of the human voice. In contrast to experiments on human speakers, experiments on mechanical replicas allow a controlled and systematic assessment of experimental parameters in a repeatable and reproducible way. Therefore, experiments on mechanical vocal folds replicas are an inherent part of physical studies aiming to understand the fluid–structure interaction driving human voiced speech production. As a result, an increasing number of mechanical replicas is presented in the literature [van den Berg *et al.*, 1957; Ruty *et al.*, 2007; Barney *et al.*, 1999; Murray and Thomson, 2012; Zhang *et al.*, 2006; Titze, 1988; Zhang and Luu, 2012].

*Corresponding author.

Whereas the design of these replicas is guided from observations on human speakers, the real-life complexity is simplified in order to facilitate systematic measurements for a wide range of physiologically meaningful initial conditions. Nevertheless, in general, variation of initial conditions is cumbersome and requires the researcher to intervene manually. Indeed, in order to study the effect of elastic properties, synthetic silicone-based vocal folds [Murray and Thomson, 2012; Pickup and Thomson, 2009; Murray *et al.*, 2014] need to be replaced which involves remounting of the replica which hinders repeatability of experimental results and is prone to introduce experimental errors. Mechanical replicas representing a vocal fold as a liquid-filled deformable cylinder in a rigid setting [Ruty *et al.*, 2007; Cisonni *et al.*, 2011] allow to adapt the quantity of liquid in order to study the effect of elasticity, so that compared to synthetic replicas errors related to remounting of the replica are avoided. Nevertheless, changing the amount of liquid will inflate or deflate the deformable cylinder so that besides the elasticity, the geometry of the area between both vocal folds, i.e., the glottis, will be affected since both its area and outer contour will vary.

In the current work, an existing liquid-filled vocal folds replica is considered as illustrated in Fig. 1(a) [Haas *et al.*, 2016]. It is sought to impose in an automated and accurate way: (1) the water pressure inside each vocal fold and (2) the glottal area. This way, initial conditions related to the elasticity and glottal geometry are no longer correlated and manual intervention is avoided which benefits the reproducibility and accuracy of experimental results.




(a) Single, general view (b) Two, front view

Fig. 1. Illustration of vocal folds replica consisting of a liquid-filled deformable cylinder in a metal setting: (a) general view of a single vocal fold and (b) front view of two vocal folds illustrating the glottis, i.e., area between both vocal folds (dotted rectangle).

In Sec. 2, the vocal fold's replica and setup are described and new features that are required to automate the variation of (1) the internal water pressure inside each vocal fold and (2) the glottal area are outlined. Image processing is used to accurately determine geometrical glottal features and subsequent control of the glottal area is discussed in Sec. 3. Examples of measurements obtained during auto-oscillation of the vocal folds using the improved replica and setup are discussed in Sec. 4. In addition, it is shown how the proposed setup and image processing can contribute to the experimental study of vocal folds during auto-oscillation for different initial conditions of glottal area, mean upstream pressure and vocal fold elasticity. The conclusion is formulated in Sec. 5.

## 2. Vocal Fold's Replica and Setup

The existing vocal fold's replica and setup is described in Sec. 2.1 (prior features). Next, improvements to the setup are outlined in Sec. 2.2 (new features) in order to realize the objectives, i.e., variation of initial conditions in an automated and quantified way. For clarity, an overview of prior and new features is presented in Table 1.

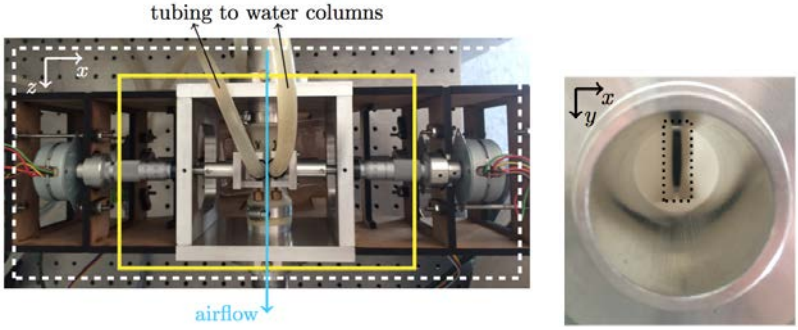
### 2.1. *Prior features*

The existing vocal folds replica is detailed in [Haas *et al.*, 2016]. Briefly, elastic properties of each vocal fold (Fig. 1) are varied by changing the water pressure inside each deformable cylinder. This is done in a quantitative way by connecting each vocal fold to a water column whose height can be adapted manually. Note that asymmetrical elastic properties between both vocal folds can be assessed since the water pressure for each vocal fold can be set independently. Next, vocal folds are connected to an upstream and downstream rigid tube representing, respectively, the trachea exit and vocal tract inlet. An upper view of mounted vocal folds is illustrated in Fig. 2(a). Note that the resulting tubing is sealed thanks to a deformable tube spanned between the trachea exit and the vocal tract inlet so

Table 1. Overview of prior and new features of the vocal folds replica and setup during experiments.

| Setting | Physical Property | Prior | New |
|---|---|---|---|
| Height water column (2) | Elasticity, geometry | Quantified (M) | Motor (A)[a] |
| Micrometer screws (4) | Geometry | Poorly quantified (M) | Motor (A) |
| Front-view imaging[b] | — | Not quantified | Camera,[c] lights[d] |

*Notes*: (A) Automated, (M) Manual.
[a]Positioning water column on a vertical glider.
[b]Of glottis, i.e., geometrical features of gap between both vocal folds.
[c]Integrated in measurement system.
[d]LED strips and fixed position of main light sources.




(a) Top view (b) Front view

Fig. 2. Mounted vocal folds replica indicating prior features (full rectangle) and new features (dashed rectangle): (a) top view ($(x, z)$-plane) and (b) front view ($(x, y)$-plane) illustrating area between both vocal folds (inside dotted rectangle).

that the vocal folds are clamped against its exterior. Each vocal fold is equipped with two micrometer screws (Mitutoyo 153-101, accuracy 0.01 mm, range 5 mm), one towards each outer edge (Fig. 2(a)), so that its position in the $(y, z)$-plane can be varied and the aperture between both vocal folds subsequently varies as well. Moreover, the vocal folds can be positioned asymmetrically with respect to the center line of the upstream and downstream tube. Next, the upstream tube (trachea) is then mounted to a pressure reservoir representing the lungs, which is supplied by an airflow facility. The fluid–structure interaction can then be studied for different upstream pressures (0 up to 3000 Pa), thanks to a pressure regulator (Norgren 11-818-987) followed by a manual valve. A calibrated pressure sensor (Kulite XCS-0.93-0.35-Bar-G with power supply EA-3005S) mounted upstream from the glottal replica is available to measure upstream pressure $P$ during the experiments [Ruty *et al.*, 2007].

A measurement chain is available for signal generation and data acquisition. Concretely, a multiple input output data acquisition card (PC/DAQ, National Instruments NPCI-MIO-16XE-10) and connector block (BNC 2110) together with LabView (National Instruments) is used.

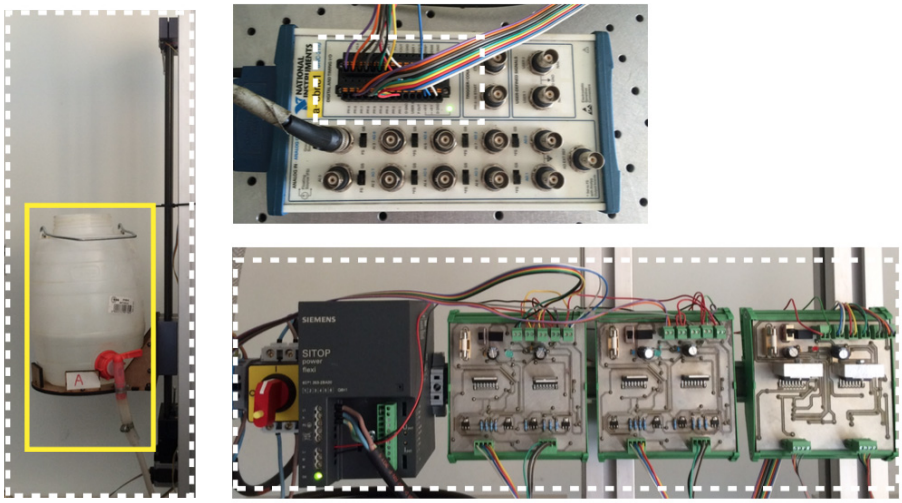
In addition, a high-speed digital camera (Mikrotron, MotionBlitz EoSens Cube7, Gigabit Ethernet interface, CMOS sensor) equipped with a macro lens (Tokina, AT-X100 mmf/2.8 PRO D) and two light blocks (ST Mirl N2) can be used to gather front-view images as shown in Fig. 2(b). Nevertheless, the camera is not integrated in the measurement system so that it can be used as a still photo camera in order to gather information of the glottal area prior to the start of the experiment, i.e., in absence of flow, using dedicated operator software (MotionBlitz, director 2). Further extraction of geometrical features is then done by post-processing of the captured images.

## 2.2. *New features*

The existing setup (Sec. 2.1) is altered in order to enable automated variation and quantification of initial conditions, i.e., elasticity and glottal area. To this goal, two main adaptations are made. Firstly, the setup is motorized so that the height of each water column (elasticity) as well as the positioning of each micrometer screw (glottal area) can be automated [Delepine *et al.*, 2016]. Secondly, the high-speed camera is integrated in the measurement chain so that glottal geometrical features can be captured throughout experiments [Delepine *et al.*, 2016].

### 2.2.1. *Motorization of the setup*

Each micrometer screw is equipped with a two-phase step-motor (McLennan 42M0-48C–1U, velocity $< 0.5\,\text{mm/s}$) as illustrated in Fig. 2(a). This way, each vocal fold can be displaced along the $x$-dimension with a maximal stroke of 7 mm. Next, each water column is mounted on a vertical (along the $y$-dimension) one-dimensional positioning system (IGUS SHT-12-AWM-500, maximum stroke of 500 mm) with a platform as illustrated in Fig. 3(a). Each positioning system is again equipped with a two-phase step-motor (Sanyo Denik 103h5208, accuracy 0.01 mm/step, velocity $< 5\,\text{mm/s}$). The useful range of the positioning system along which the water column height can be varied yields 45 cm. This means that the internal water pressure inside the vocal folds can be varied up to 4500 Pa.



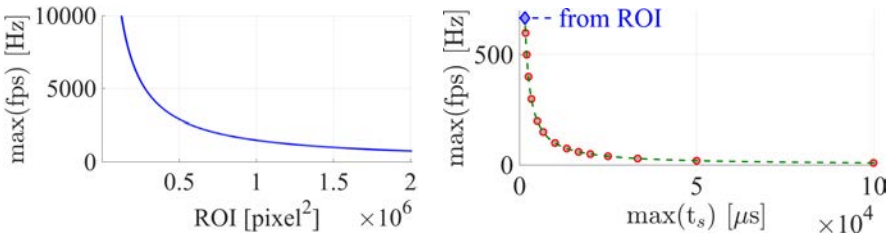

(a) Water column (b) Step motors PCB-chain (bottom) and DAQ connection (top)

Fig. 3. Illustration of prior features (full rectangle) and new features (dashed rectangle) concerning: (a) vertical $y$-positioning system to vary water column height for a single vocal fold and (b) step motors rail PCB-chain (bottom) and DAQ connection (top).

Dedicated phase circuit boards (PCB), shown in Fig. 3(b), are realized in order to power and command the step motors driving the water column heights (two motors) and micrometer screws (four motors) [Delepine *et al.*, 2016]. Each circuit board is designed to drive 2 step motors. Power is supplied using a transformator (Siemens SITOP) which is regulated to give 6.3 V at its outlet. An additional DC/DC converter (Traco Power TSR-1) is added on the PCB's to convert this input voltage to 5 V as required for its active components. The two main components on each PCB are two dual full-bridge drivers (ST Microelectronics L298N) to relay input commands to each of the step motors. Each step motor needs four input signals, i.e., one for each coil. Nevertheless, since coils have $2 \times 2$ opposite phase, the number of required input signals is halved by adding two transistors (Mosfet, TYPE MERK) for each driver to inverse the input. This reduction is not only more elegant, but allows to use the same data-acquisition block (connector card DAQ-board BNC 2110, National Instruments, 16 digital outputs) to command the motors as well as to acquire the data, which benefits the measurement procedure.

### 2.2.2. *High-speed imaging*

The high-speed camera is integrated in the measurement chain so that usage of the dedicated software is no longer required and the camera can be used as any other measurement instrument for both still and rapid imaging [Delepine *et al.*, 2016]. To this end, camera parameter settings are added to the measurement interface (labView, National Instruments): shutter time ($2\,\mu\text{s} < t_s < 1\,\text{s}$), region-of-interest (ROI) ($128 \times 128 < \text{ROI} < 1696 \times 1710$, [pixel$^2$]) and frames per second (523 <fps< 45000 [Hz]). Note that the ROI is defined by its area value as well as by its resolution $N_x$ along the $x$-dimension and $N_y$ along the $y$-dimension. The ROI (area and resolution), maximum frame rate per second max(fps) and maximum shutter time max($t_s$) are correlated (Fig. 4(a)) to ensure sufficient light exposure on the camera sensor. Therefore, the relationship between the maximum fps and region-of-interest (ROI) is non-trivial. An upper bound of max(fps) is found from the ROI-area for the




(a) Upper bound for ROI (b) Influence of $t_s$

Fig. 4. Characterization of correlated camera settings. (a) upper bound of maximum frames per second (max(fps)) as a function of the ROI for the smallest maximum shutter time. (b) relationship between max(fps) and maximum shutter time (max($t_s$)) for constant ROI ($1500 \times 1500$) corresponding to an upper bound of 666 Hz ($\diamond$).

smallest maximum shutter time so that the relationship ROI$\times$ max(fps) $\approx 1.53e9$ holds as shown in Fig. 4(a). Below this upper bound due to the ROI, the maximum fps decreases as the maximum shutter time is increased following the relationship $\max(\text{fps}) \times \max(t_s) \approx 998580$, as illustrated in Fig. 4(b).

Next, the position of the main light blocks is fixed at either side of a vocal fold so that only its distance along the $x$-dimension can be varied from 12 cm up to 6 cm. Concretely, experiments are performed with the lights as close as possible to each vocal fold (6 cm). This way, the variation in light conditions between experiments is reduced and the illuminance is maximum, which benefits the subsequent image processing either during or after the measurements. In addition, small (34 mm $\times$ 5 mm) LED strips are developed as detailed in Appendix A which can be positioned around the glottal area, e.g., to avoid shadow zones within the ROI.

## 3. Control of Setup Parameters

The control of the elasticity of each vocal fold is straightforward since it is set by the height of each water column which is directly related to commanding the vertical positioning system by means of its step motors as outlined in Sec. 2.2. The control of the geometrical glottal features on the other hand is less obvious. Indeed, a first requirement is to quantify geometrical features of interest during the measurement process and next to command the micrometer screws by means of its step motors in order to adapt the feature of interest. Note, that by firstly imposing the wanted elasticity and secondly the wanted geometrical glottal feature, both parameters are decorrelated, which is an important advantage compared to previously used setups. In this section, firstly, geometrical feature extraction is outlined (Sec. 3.1) and secondly the control of glottal parameters is detailed (Sec. 3.2).

### 3.1. *Geometrical glottal feature extraction*

Once the high-speed camera is integrated in the measurement system, it is positioned so that front-view images from the glottal aperture are quantified, i.e., concretely at 29 cm in front of the glottis. Geometrical features are extracted using basic image processing techniques to detect the contour of the glottal aperture and derive the features of interest [National Instruments, 2004; Kumar *et al.*, 2012; Yan *et al.*, 2006]. Concretely, the glottal area $A$, maximum aperture following the $y$-dimension $\max(a_y)$ and $x$-dimension $\max(a_x)$ are quantified. In addition, for each $y$-position of the image ($1 \leq i \leq N_y$), the corresponding number of pixels $a_x^i$ associated with the aperture is quantified as well. Consequently, the area $A$ can be obtained as $A = \sum_{i=1}^{N_y} a_x^i$ as well. A captured image and a processed image indicating main geometrical features are illustrated in Fig. 5.

The applied segmentation process uses a threshold value to converse the gray-scale image to a binary image and a number of erosions to remove spurious spots. Optimal values for these parameters varies from one image to the next. Therefore,

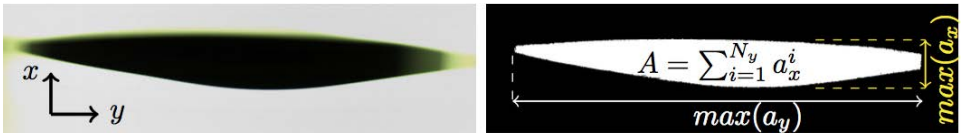
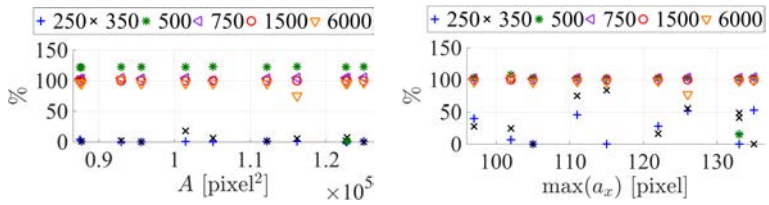



Fig. 5. Illustration of basic image processing. Original captured image (left) and processed image (right). Main geometrical features of glottal front-view images: area $A$, maximum aperture following the vertical $y$-dimension $\max(a_y)$ and maximum aperture following the vertical $x$-dimension $\max(a_x)$. Note that area $A$ can be obtained as the sum of the number $a_x^i$ of pixels associated with the aperture at each $y$-position, i.e., $1 \leq i \leq N_y$.

in order to improve the robustness of the feature detection and to avoid the tedious task of manual segmentation parameter adaptation for each image, these parameters are updated automatically using a simple adaptive approach [Delepine *et al.*, 2016]. Concretely, the threshold value is reduced until a non-zero area is obtained whereas the number of erosions is increased until a single object is detected. The influence of these parameters is further illustrated in Appendix B. Note that the iterative process does not significantly change the time required for image processing which is of the order of seconds.

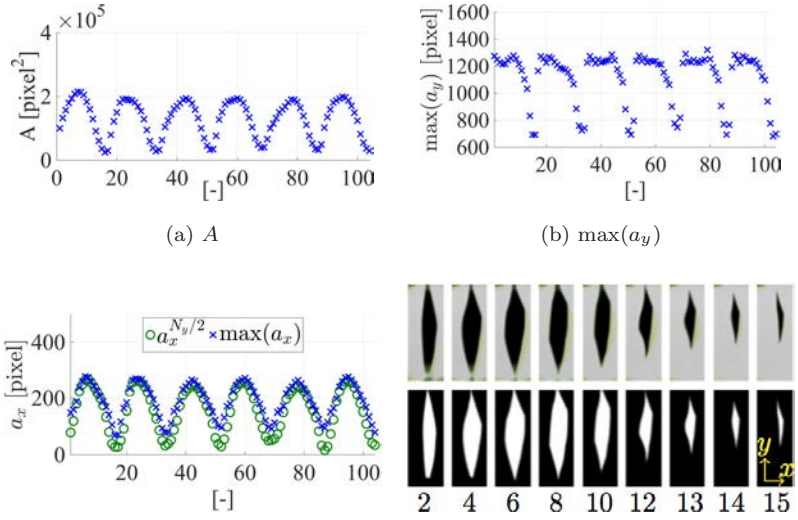
Besides the segmentation parameters, light conditions will influence the image processing as well. Light conditions will be influenced by the setup as well as by the camera. Section 2.2 details the setup which is adapted to fix the position of main light sources. Therefore, the influence of the light position on the image processing is not considered. The exposure time of the camera is determined by the chosen shutter. From Fig. 6, it is seen that a shutter time $\geq 750\,\mu$s is sufficiently long to ensure accurate feature extraction so that its value does not affect the extracted features. Therefore, all experiments are performed with a shutter time set to $750\,\mu$s since this value permits to set the frame rate to 1315 Hz whereas further increasing the shutter time would decrease the frame rate and hence limit the frequency interval for which useful information can be obtained from the extracted geometrical features (see Sec. 4).

The resulting geometrical feature extraction of images obtained on an auto-oscillating replica is illustrated in Fig. 7. From the images, it is observed that the




(a) $A$ (b) $\max(a_x)$

Fig. 6. Relative influence of shutter time ($\mu$s) on extracted geometrical features: (a) area $A$ and (b) maximum aperture along the $x$-dimension $\max(a_x)$. Values obtained for $1500\,\mu$s are taken as a reference.

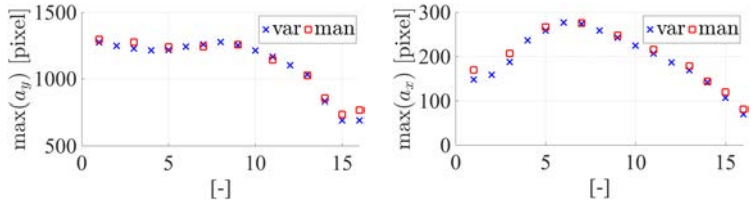



(a) $A$ (b) $\max(a_y)$

(c) $a_x^{N_y/2}$, $\max(a_x)$ (d) Typical images

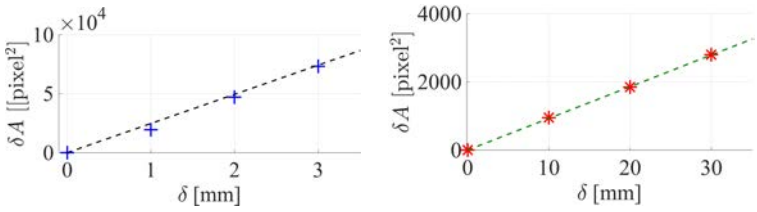
Fig. 7. Example of geometrical glottal feature extraction during auto-oscillation of the replica: (a) area $A$, (b) maximum aperture along the $y$-dimension $\max(a_y)$, (c) aperture at half the total resolution along the $y$-dimension $a_x^i$ for $i = N_y/2$ (○), maximum aperture along the $x$-dimension $\max(a_x)$ (×) and (d) captured (top) and processed (bottom) images with associated image number.

shape of the glottal contour is not rectangular so that the quantified feature $a_x$ varies depending on the associated $y$-position. This is illustrated by comparing $a_x^{N_y/2}$ (value of aperture $a_x$ at half $y$-resolution) and $\max(a_x)$ (maximum aperture value of $a_x$ for all possible $y$-positions) in Fig. 7(c), for instance, during closure, differences up to 50 pixels are observed. Nevertheless, in both cases, $a_x$ varies quasi-periodically as does the estimated area $A$ and $\max(a_y)$ due to the auto-oscillation of the vocal folds. The maximum aperture along the $y$-dimension $\max(a_y)$ is nearly constant when the glottal area is opening. It would be interesting to investigate this glottal shape asymmetry between opening and closing in future work using the current setup. Furthermore, from Fig. 8, it is seen that extracted features using the adaptive




(a) $\max(a_y)$ (b) $\max(a_x)$

Fig. 8. Comparison between manually extracted features (man) and features extracted using the adaptive approach (var) for a portion of the images shown in Fig. 7: (a) maximum aperture along the $y$-dimension $\max(a_y)$ and (b) maximum aperture along the $x$-dimension $\max(a_x)$.




(a) Micrometer screw (b) Water column height

Fig. 9. Static (no air is supplied) glottal area change ($\delta A$) as a function of displacement $\delta$ (symbols) and associated linear fit (dashed lines) for step motors driving: (a) micrometer screws with slope 25000 pixel$^2$/mm, (b) 1D vertical positioning systems or water height columns with slope 100 pixel$^2$/mm.

approach are comparable to manually extracted features for $\max(a_y)$ and $\max(a_x)$. The overall relative error is smaller than 10%.

### 3.2. *Control of glottal setup parameters*

Driving the micrometer screws and water column heights with the step motors will affect the static glottal area in the absence of flow to an extent which can be quantified from the observed glottal features, i.e., area $A$ in this case. Figure 9 illustrates the relationship between displacement $\delta$ of micrometer screws and water column height and the resulting change of area $\delta A$. A linear relationship between $\delta$ and $\delta A$ is observed in both the cases. The slopes yield 25000 pixel$^2$/mm for the micrometer screws and 100 pixel$^2$/mm for the water column height.

The estimated slopes are used to impose the glottal area at the start of the experiment using either one of the motorized systems, i.e., micrometer screws or water column height. Concretely, an area set point as well as a tolerance level (typically $< 2\%$) are given by the user. Next, a loop is executed until the tolerance criterium for the set point is reached: firstly, the current area is calculated; secondly, the difference between the current area and area set point is evaluated; thirdly, a displacement command is send to the step motors based on the estimated slopes. The tolerance criterium is generally satisfied in 3 up to 5 iterations so that the loop provides a simple and effective control algorithm for the glottal area.

The initial glottal area control algorithm can then be used to study the influence of the initial glottal area $A_0$ on the fluid–structure interaction driving the vocal folds auto-oscillation while maintaining the height of each water column and hence the elasticity of each vocal fold constant, i.e., decorrelating elasticity and glottal area. An example of a systematic increase of initial glottal area for constant conditions of the water column heights is given in Fig. 10. It is seen that the initial glottal area $A_0$ resulting from the control algorithm is well within the tolerance level for each area set point. In addition, the area estimated at the end of a fluid–structure interaction driving vocal folds auto-oscillation for each initial area $A_0$ is indicated. It is seen that the fluid–structure interaction can cause the area to vary up to 30%

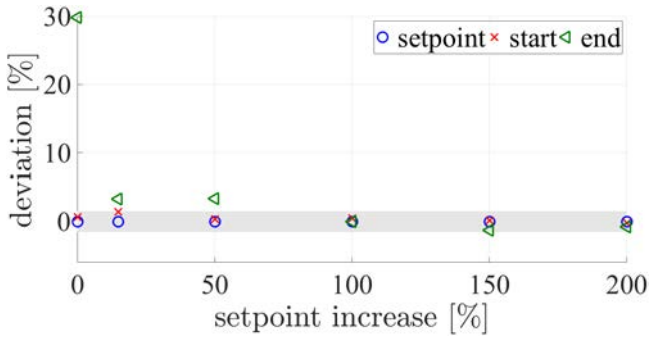



Fig. 10. Illustration of systematic usage of glottal area control for constant elasticity of each vocal fold (asymmetrical condition, constant water column heights of 63.5 cm and 43.5 cm). The initial area set point (set point, ○) is systematically increased from 0% up to 200% using the area control algorithm commanding the micrometer screws of each vocal folds. The tolerance level (grey area) was set to 1.5%. The actual initial area resulting from the control algorithm $A_0$ (start, ×) is shown as the deviation (%) from its set point. In addition, the estimated area after a fluid–structure interaction (end, ◁) is given.

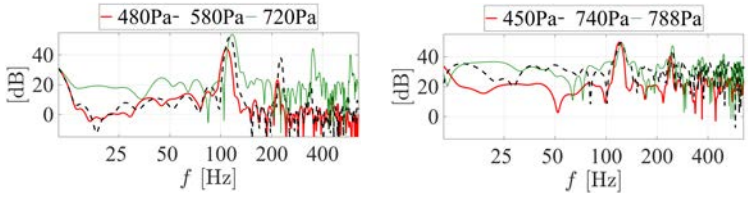
from the initial area which is outside the imposed tolerance level of the initial glottal area. This shows that the accuracy of experiments studying the effect of the initial glottal area set point needs to be reinitialized after each fluid–structure interaction. The current setup allows to do so.

Note that a similar approach can be applied to impose other geometrical features when this is of interest.

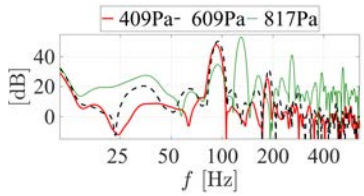
## 4. Imaging of Asymmetrical Vocal Folds Auto-Oscillation

Measurements are performed using the developed setup with automated control of the height of each of the water columns, i.e., elasticity of each vocal fold, as well as with automated control of the initial glottal area $A_0$ prior to auto-oscillation of the vocal folds. The integration of high-speed imaging in the measurement setup enables to capture images during auto-oscillation of the vocal folds at a constant mean upstream pressure $P_u$. The image processing is then also applied to extract geometrical glottal features of the images obtained during vocal folds auto-oscillation as shown in Sec. 3.1. Extracted image features can then be analyzed as a function of the imposed experimental conditions, i.e., height of each water column, initial glottal area $A_0$ and mean upstream pressure $P_0$.

In the following, experiments are discussed which are obtained for three different asymmetry conditions of the vocal folds elasticity. Three different conditions (no asymmetry, medium asymmetry and large asymmetry) are obtained by imposing a constant water column height for one vocal fold (87 cm) and by imposing respectively 100%, 69% and 47% of this height to the other vocal fold (87 cm, 67 cm and 47 cm). Note that 100% corresponds to symmetrical elasticity properties in both vocal folds (no asymmetry), while the asymmetry increases as the




(a) 100%, 87 cm and 87 cm (b) 69%, 87 cm and 67 cm



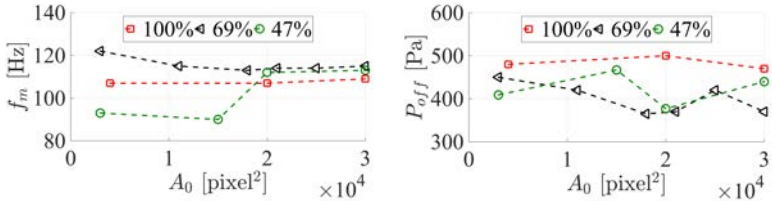

(c) 47%, 87 cm and 47 cm

Fig. 11. Illustration of area spectra estimated on images obtained during vocal fold auto-oscillation for a constant initial area $A_0 \approx 3500\,\text{pixel}^2$ as a function of upstream pressure $400 < P_0 < 820\,\text{Pa}$ and vocal fold elasticity asymmetry degree: (a) symmetrical condition 100% (87 cm and 87 cm), (b) medium asymmetry condition 69% (87 cm and 67 cm) and (c) large asymmetry condition 47% (87 cm and 47 cm).

water column height is decreased to 69% (medium asymmetry) and 47% (large asymmetry), respectively.

The frequency content of the estimated areas $A$ of consecutive images is analyzed for each set of initial conditions (asymmetry condition, $A_0$ and $P_0$). Obtained spectra are illustrated in Fig. 11 for assessed asymmetry conditions and for three upstream pressures in the range $400 < P_0 < 820\,\text{Pa}$. The initial glottal area $A_0 \approx 3500\,\text{pixel}^2$ is held constant. It is seen that the frequency peak $f_m$ associated with maximum spectral energy is of the order of 100 Hz whereas its actual value shifts depending on imposed initial conditions such as the asymmetry condition. Compared to $f_m$ observed for symmetrical conditions (100%), the frequency peak for a moderate asymmetry (69%) shifts to a slightly higher frequency regardless of the upstream pressure. For a large asymmetry (47%), the influence of the pressure is more complex, since the frequency peak for 409 Pa and 609 Pa shifts to lower frequencies than the peak observed for 100% while for 817 Pa a new frequency peak appears at a higher frequency. Shown spectra suggest that the coupling between both vocal folds during auto-oscillation is affected by the degree of asymmetry between both water column heights as well as by the applied upstream pressure.

The influence of imposed initial glottal area $A_0 \leq 30000\,\text{pixel}^2$ as a function of asymmetry (100%, 69% and 47%) on auto-oscillation features is assessed near the offset pressure of auto-oscillation $P_0 \approx P_{\text{off}}$, i.e., the upstream pressure below which the auto-oscillation ceases. Figure 12(a) illustrates frequency $f_m$ associated with maximum spectral energy and Fig. 12(b) shows the associated offset pressure



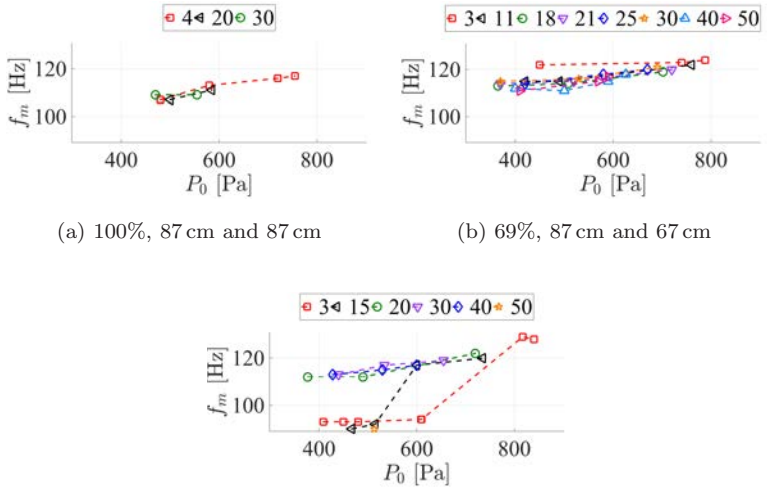

(a) $f_m$ (b) $P_{\text{off}}$

Fig. 12. Influence of imposed initial glottal area $A_0$ (pixel$^2$) as a function of asymmetry condition (100%, 69% and 47%) on auto-oscillation features near auto-oscillation offset pressure $P_0 \approx P_{\text{off}}$: (a) spectral frequency $f_m$ associated with maximum spectral energy, (b) offset pressure $P_{\text{off}}$.

$P_{\text{off}}$. It is seen that offset pressure $P_{\text{off}}$ decreases for all assessed initial glottal areas $A_0$ when an asymmetry between both vocal folds is imposed. Therefore, current data suggest that introducing an asymmetry by lowering the water column of one vocal fold facilitates auto-oscillation. The overall variation of the offset pressure in the assessed range $A_0 \leq 30000\,\text{pixel}^2$ is small ($< 30\,\text{Pa}$) for symmetrical vocal folds (100%), whereas the extent of the variation increases for asymmetrical vocal folds ($< 100\,\text{Pa}$). These findings suggest that asymmetrical vocal fold configurations are more sensible to the precise value of initial area $A_0$ than symmetrical vocal fold configurations. When considering peak frequency $f_m$, the variation in the range $A_0 \leq 30000\,\text{pixel}^2$ increases as the asymmetry is increased ($< 5\,\text{Hz}$ for 100%, $< 10\,\text{Hz}$ for 69% and $< 25\,\text{Hz}$ for 47%). It is seen that the variation occurs for $A_0 < 20000\,\text{pixel}^2$ and in this range the degree of asymmetry influences the peak frequency $f_m$, suggesting that it affects the coupling between both vocal folds during auto-oscillation. In the range $A_0 > 20000\,\text{pixel}^2$, the observed $f_m$ is determined by the presence ($f_m \approx 114\,\text{Hz}$) or absence ($f_m \approx 108\,\text{Hz}$) of asymmetry and less by the degree of asymmetry.

Peak frequency $f_m$ as a function of upstream pressure $P_0$ for different initial glottal areas $A_0$ is shown in Fig. 13 for each asymmetry condition (100%, 69% and 47%). As a general tendency, $f_m$ increases with increasing $P_0$ for all conditions. The overall effect of the initial glottal area $A_0$ on $f_m$ increases as the asymmetry degree increases so that it is most pronounced for the large asymmetry condition ($< 10\,\text{Hz}$ for 69% and $< 20\,\text{Hz}$ for 47%). The influence of initial area on $f_m$ is more pronounced for small initial glottal areas ($A_0 < 12000\,\text{pixel}^2$ for 69% and $A_0 < 21000\,\text{pixel}^2$ for 47%). Note that these observations correspond with observations for pressures $P_0 \approx P_{\text{off}}$ (Fig. 12). Therefore, it is of interest for future studies to extend the range of assessed asymmetry conditions and initial glottal areas in order to obtain definite conclusions.

Besides a spectral analysis of extracted geometrical features, it is of interest to analyze the asymmetry of geometrical features in order to understand the movement of each vocal fold due to the imposed asymmetry conditions. This is illustrated by considering the left–right asymmetry of $a_x^y$ with respect to the point with coordinate



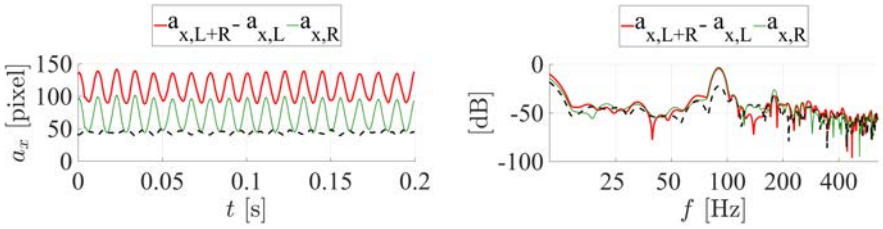

(a) 100%, 87 cm and 87 cm

(b) 69%, 87 cm and 67 cm

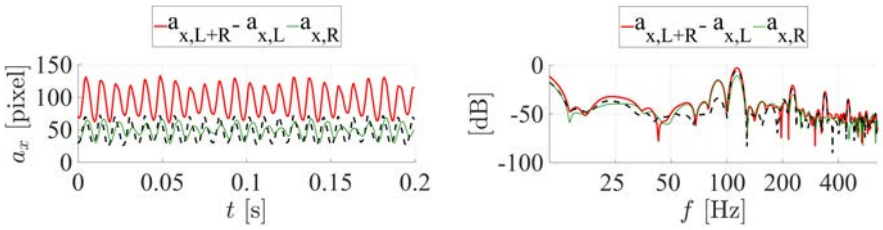
(c) 47%, 87 cm and 47 cm

Fig. 13. Influence of upstream pressure $P_0$ and initial glottal area $A_0$ (pixel$^2$) (symbols denote $A_0/1000$) for each asymmetry condition: (a) symmetrical condition 100% (87 cm and 87 cm), (b) medium asymmetry condition 69% (87 cm and 67 cm) and (c) large asymmetry condition 47% (87 cm and 47 cm).

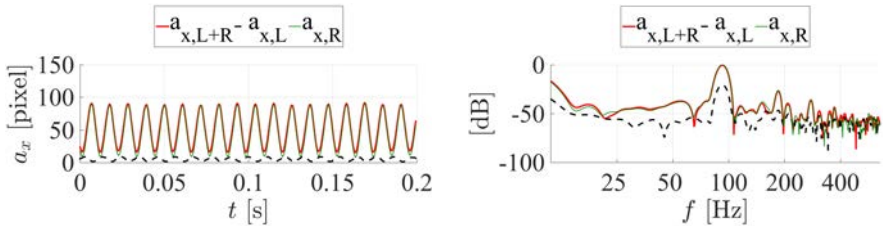
$(q, y)$, where $y$ indicates the $y$-position of the line scan and the $q$-position cuts the line in two, i.e., left half ($a_{x,L}$) associated with the left vocal fold with constant elasticity (water column height at 87 cm), right half $a_{x,R}$ associated with the right vocal fold for which the elasticity is varied (water column height at 87 cm, 67 cm or 47 cm) and total $a_{x,L+R} = a_{x,L} + a_{x,R}$. As an example, the point $(q, y)$ is taken as the center of the aperture with the smallest area. Figure 14 illustrates temporal and spectral features for the large asymmetry condition (47%, left at 87 cm and right at 47 cm) for different initial glottal areas and upstream pressures ($300 < P_0 < 520$ Pa). For the first ($A_0 = 3000$ pixel$^2$ and $P_0 = 513$ Pa) and third case ($A_0 = 50000$ pixel$^2$ and $P_0 = 409$ Pa) the amplitude of $a_{x,L}$ is nearly constant suggesting that the left vocal fold almost does not move during oscillation. This suggests that the oscillation of $a_{x,L}$ is a byproduct of the oscillation observed for $a_{x,R}$. The frequency associated with maximum spectral energy is almost similar in both cases (90 Hz and 92 Hz) and few spectral differences are observed between $a_{x,L+R}$ and $a_{x,R}$ for the main spectral peak. Temporal signals $a_{x,L}$ and $a_{x,R}$ observed for the second case ($A_0 =$ 40000 pixel$^2$ and $P_0 = 428$ Pa in Fig. 14(b)) move with a similar amplitude so that both vocal folds contribute to the same extent to the total movement and none can be considered a byproduct of the other. The frequency peak at 90 Hz still appears in the spectra due to the movement of the right vocal fold associated with $a_{x,R}$, in addition to a frequency peak at 114 Hz appears due to the movement of the left vocal fold $a_{x,L}$. These examples show that such an analysis can contribute to the understanding of spectra shown in Fig. 11(c) and that initial conditions largely




(a) $A_0 = 50000\,\text{pixel}^2$ and $P_0 = 513\,\text{Pa}$




(b) $A_0 = 40000\,\text{pixel}^2$ and $P_0 = 428\,\text{Pa}$




(c) $A_0 = 3000\,\text{pixel}^2$ and $P_0 = 409\,\text{Pa}$

Fig. 14. Illustration of asymmetry analysis of line scan $a_x$ during auto-oscillation of the vocal folds replica for a large asymmetry condition 47% (87 cm and 67 cm). Temporal (left-hand side, 0.2 s) and spectral (right-hand side) data for the left portion $a_{x,L}$ (dashed line), right portion $a_{x,R}$ (thin full line) and the total line scan $a_{x,L+R}$ (thick full line) for: (a) $A_0 = 50000\,\text{pixel}^2$ and $P_0 = 513\,\text{Pa}$, (b) $A_0 = 40000\,\text{pixel}^2$ and $P_0 = 428\,\text{Pa}$, (c) $A_0 = 3000\,\text{pixel}^2$ and $P_0 = 409\,\text{Pa}$.

affect vocal folds auto-oscillation (no movement or movement for similar elasticity conditions) so that detailed studies using a suitable setup are needed.

The given example of line-scan asymmetry is comparable to high-speed videokymography of vocal fold vibration on a human subject [Svec, 2000] or high-speed imaging based techniques [Booth and Childers, 1979; Yan *et al.*, 2006; Aichinger, 2014]. It is noted that the applied approach can be easily extended to other line positions or to multiple lines.

## 5. Conclusion

An existing vocal fold replica is automated and high-speed image processing is integrated in the measurement chain. The new setup allows to study the fluid–structure

interaction driving vocal fold auto-oscillation for a quantified and systematic variation of initial conditions, i.e., initial glottal area, mean upstream pressure and elasticity of each vocal fold. In contrast to previous studies, the initial glottal area and vocal fold elasticity are no longer correlated so that both parameters can be studied independently. Furthermore, the integration of high-speed imaging allows not only to impose initial conditions, but compared to previous studies, additional physical quantities related to the glottal geometry can be analyzed together with other physical quantities, e.g., pressure. As an example, some measurements for asymmetrical vocal folds elasticity are analyzed for different initial glottal areas and upstream pressures. It is shown that geometrical features of the time-varying glottal area during auto-oscillation allow to understand spectral features in terms of the motion of each vocal fold. In addition, it is shown that vocal fold movement can be inhibited or favored depending on the initial glottal area. Therefore, the current setup is an important new tool for the study of vocal folds pathologies (such as vocal fold paralysis) as well as the development and validation of theoretical vocal fold auto-oscillation models. Indeed, the current setup allows to account for the actual initial glottal area which is an important model input parameter. Furthermore, the current setup allows to take into account the actual shape of the glottis instead of assuming a rectangular glottal shape for the flow model.

## Acknowledgments

Clément Leduque, Benoit Delepine and Thomas Themelin are acknowledged for their contribution to the development of the measurement chain. This work was partly supported by the ArtSpeech project (ANR-15-CE23-0024).

## Appendix A. LED Strip

Small (34 mm×5 mm) LED strips are developed as shown in Fig. A.1 [Delepine *et al.*, 2016]. The LED strips are composed of equally spaced six high power LEDs (ProLight, 1 W High CRI Power LED PK2N-1LWE-R8, 110 lm at 350 mA). The performance of the LED strip is compared to a single LED by measuring the illuminance (VoltCraft MS-1300) in similar conditions (dark space, 84 mm between centers of the light source and sensor). It is seen (Fig. A.1) that the resulting illumination can be approximated by an additive law since the overall loss is smaller than 10% ($\Delta < 10\%$) and this reduces further to less than 5% ($\Delta<5\%$) for currents smaller than 0.25 mA.

## Appendix B. Influence of Segmentation Parameters

The image processing and hence the obtained geometrical glottal features depend on the segmentation parameters, i.e., threshold and number of erosions. In particular, the value of the threshold parameter used to obtain a binary image affects the

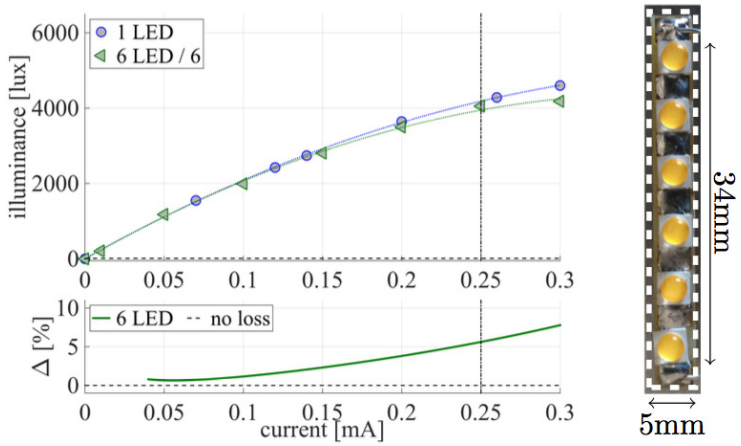
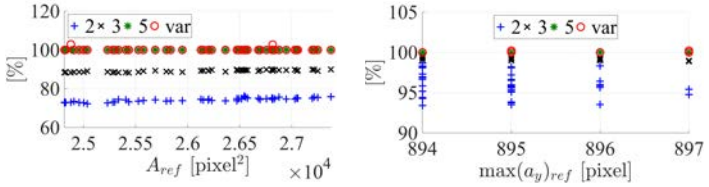
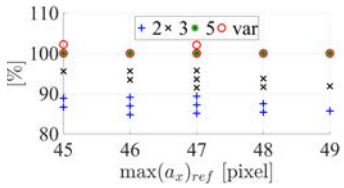



Fig. A.1. LED strip (right) and its illuminance performance (top left) as a function of supplied current compared to a single LED. The background level is indicated (horizontal dashed line). The loss $\Delta$ (bottom left) in comparison to a single LED is quantified. The current corresponding to 5% loss is indicated (vertical dashed line).




(a) $A$ (b) $\max(a_y)$




(c) $\max(a_x)$

Fig. A.2. Influence of threshold value on feature extraction, manual constant value compared to adaptive approach (var) where a constant value of 5 is taken as a reference value: (a) $A$ compared to $A_{\mathrm{ref}}$, (b) $\max(a_y)$ compared to $\max(a_y)_{\mathrm{ref}}$ and (c) $\max(a_x)$ compared to $\max(a_x)_{\mathrm{ref}}$.

feature extraction. This is illustrated in Fig. A.2. The area is varied by a small variation of the height of each water column. Feature values of $A$, $\max(a_y)$ and $\max(a_x)$ obtained using manually set constant threshold values are compared against feature values obtained with the adaptive approach. It is seen that the manual setting of the threshold value might result in an underestimation of the features of interest to a degree which is determined by its value. In particular, the value of the area is affected since differences up to more than 20% are observed.

## References


Aichinger, P. [2014] Diplophonic voice: Definitions, models and detection, Ph.D. thesis, Graz University of Technology, Austria.

Barney, A., Shadle, C. H. and Davies, P. O. A. L. [1999] "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. measurements and theory," *Journal of the Acoustical Society of America* **105**, 444–455.

Booth, L. R. and Childers, D. G. [1979] "Automated analysis of ultra high-speed laryngeal films," *IEEE Transactions on Biomedical Engineering* **26**, 185–193.

Cisonni, J., Van Hirtum, A., Pelorson, X. and Lucero, J. C. [2011] "The influence of geometrical and mechanical input parameters on theoretical models of phonation," *Acta Acustica* **97**, 291–302.

Delepine, B., Leduque, C. and Themelin, T. [2016] Automatisation et contrôle d'un système permettant d'étudier la production de parole humaine, Technical report, Univ. Grenoble Alpes, France, 2016.

Haas, J., Luizard, P., Pelorson, X. and Lucero, J. C. [2016] "Study of the effect of a moderate asymmetry on a replica of the vocal folds," *Acta Acustica* **102**, 230–239.

Kumar, R. A. V., Nataraj, K. R. and Rekha, K. R. [2012] "Morphological real time video edge detection in Labview," *International Journal of the Computer Science and Information Technologies* **3**, 3808–3811.

Murray, P. R. and Thomson, S. L. [2012] "Vibratory responses of synthetic, self-oscillating vocal fold models," *Journal of Acoustical Society of America* **132**, 3428–3438.

Murray, P. R., Thomson, S. L. and Smith, M. E. [2014] "A synthetic self-oscillating vocal fold model platform for studying augmentation injection," *Journal of Voice* **28**, 133–143.

National Instruments, *IMAQ Vision for LabVIEW*, 2004.

Pickup, B. A. and Thomson, S. L. [2009] "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal folds models," *Journal of Biomechanics* **42**, 2219–2225.

Ruty, N., Pelorson, X., Van Hirtum, A., Lopez, I. and Hirschberg, A. [2007] "An in-vitro setup to test the relevance and the accuracy of low-order models of the vocal folds," *Journal of the Acoustical Society of America* **121**, 479–490.

Svec, J. [2000] On vibration properties of human vocal folds: Voice registers, bifurcations, resonance characteristics, development and application of videokymography, Ph.D. thesis, University of Groningen, The Netherlands.

Titze, I. R. [1988] "The physics of small-amplitude oscillation of the vocal folds," *Journal of the Acoustical Society America* **83**, 1536–1552.

van den Berg, J., Zantema, J. T. and Doornenbal, P. [1957] "On the air resistance and the bernoulli effect of the human larynx," *Journal of the Acoustical Society of America* **29**, 626–631.

Yan, Y., Chen, X. and Bless, D. [2006] "Automatic tracing of vocal-fold motion from high-speed digital images," *IEEE Transactions on Biomedical Engineering* **53**, 1394–1400.

Zhang, Z. and Luu, T. H. [2012] "Asymmetric vibration in a two-layer vocal fold model with left-right stiffness asymmetry: Experiment and simulation," *Journal of the Acoustical Society of America* **132**, 1626–1635.

Zhang, Z., Neubauer, J. and Berry, D. A. [2006] "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," *Journal of the Acoustical Society of America* **120**, 2841–2848.