22

Support Vector Machine Classification of Vocal Fold Vibrations Based on Phonovibrogram Features

Michael Döllinger[1], Jörg Lohscheller[2],
Jan Svec[3], Andrew McWhorter[4] and Melda Kunduk[5]
[1]*University Hosptital Erlangen,*
[2]*University of Applied Sciences Trier,*
[3]*Palacky University Olomouc,*
[4]*Our Lady of the Lake Voice Center,*
[5]*Louisiana State University,*
[1,2]*Germany*
[3]*Czech Republic*
[4,5]*USA*

1. Introduction

Voice is invaluable for our livelihood, as it takes place in humans everyday lives, like talking, laughing, crying, singing, screaming, shouting etc. Over the past 200 000 years, humans use the lung, larynx, tongue, and lips, to produce and modify the highly intricate arrays of voice (Titze, 2006) for realizing verbal communication and emotional expression. Among the participating tissues, the vocal folds within the human larynx have evolved to be a key organ in the creation of human voice. Their vibrations serve as origin of the primary voice signal. The process of voice production is called phonation (Titze, 2006), and is the preliminary stage for speech.

In our knowledge-based societies, communication skills have become more and more important. Communication disorders became a socio economic factor: A study in the year 2000 estimated losses within the Gross National Product of the USA being up to $186 billion annually (Ruben, 2000), on the basis that approx. 10% of the entire population suffers from communication disturbances. To increase the quality of life of the people concerned on one hand and to keep the economic costs under control on the other, appropriate technologies have to be developed to disclose all factors conducive to communication disorders. Also, analysis methods have to be applied to objectively quantify grades of disease, document therapy, and to guide surgical interventions. A high number of communication disorders are due to a disturbance in voice, i.e. disturbed vocal fold vibrations.

Examination of vocal fold vibrations (*100 Hz – 300 Hz*) and the acoustic signal are the basic components of clinical voice assessment. It is widely held that vocal fold vibration irregularities lead to an impairment of the voice signal. Irregularities being present in vocal fold vibrations during sound production can be determined by direct (i.e. endoscopic

laryngeal imaging) or indirect (i.e. acoustic and aerodynamic) assessment techniques. However, detailed quantitative knowledge about interrelations between acoustic signal and vibrations of the voice generator is still in its infancies.

Currently, videostroboscopy is a commonly used clinical laryngeal imaging tool to investigate the vocal fold vibratory dynamics. However, videostroboscopy is just suitable for periodic vocal fold vibrations since the image sensor captures only one frame per oscillation cycle and thus does not fulfil the Nyquist sampling theorem (Kendall et al., 2005; Svec et al., 2008). Hence, videostroboscopy has severe limitations when it comes to investigating pathological voices which frequently exhibit non-periodic vibrations. State-of-the-art technology in investigating of vocal fold vibrations is high-speed digital imaging (HSI). Current systems are equipped with a 2D image sensor delivering images at frame rates up to *2,000-8,000 fps*, which can capture the vibration patterns of vocal folds at their usual frequencies of up to *300 Hz* along the entire visible glottal length (Schade & Mueller, 2005; Hertegard, 2005; Bonilha & Deliyski, 2008; Deliyski, et al., 2008). Thus, HSI allows visualizing regular and irregular vibration patterns which are found in normal and pathological voices (Kendall et al., 2005; Doellinger, 2009), Fig. 1.

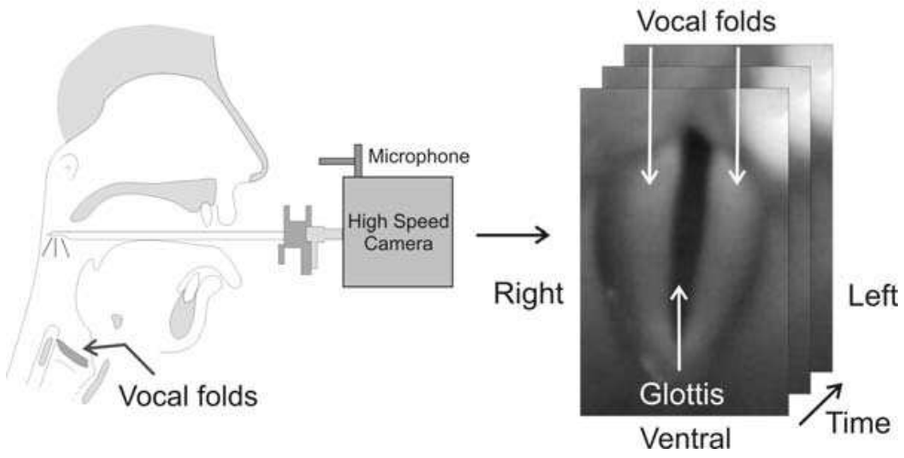


Fig. 1. Schematic representation of performing endoscopic high-speed recordings. Left, the recording situation including camera and endoscope are shown. On the left, the recorded area (vocal folds and opening and closing glottis) can be seen.

Even though high-speed videos deliver a novel insight into laryngeal vibrations, the investigation of vocal fold vibrations demands a sophisticated quantitative analysis of the video data (Doellinger, 2009). To reach this objective, different approaches have been developed to analyze vocal fold vibrations (Doellinger, 2009). Commonly, from the endoscopic HSI data the time varying opening between the vocal folds (i.e. glottis) is analyzed or trajectories are extracted at specific positions of the vocal folds (Braunschweig et al., 2008). To quantify the obtained motion data, several measures have been introduced describing the symmetry and regularity of vocal fold vibrations (Qiu et al., 2003, Yan et al., 2005). Instabilities of fundamental frequencies, amplitude and phase asymmetries as well as regularity parameters were detected in pathological voices (Bonilha & Deliyski, 2008). Other approaches automatically adapt biomechanical models to vocal fold vibrations extracted from HSI videos applying parameter optimization strategies (Doellinger et al., 2002; Doellinger et al. 2003; Tokuda et al., 2007, Yang et al., 2010). These obtained parameters

represent the degree of laryngeal asymmetry and vibration stability (Schwarz et al., 2008; Wurzbacher et al., 2006; Wurzbacher et al., 2008). However, up to the present there is still no established feature extraction strategy describing the entire vibration patterns of vocal fold dynamics adequately.

Recently, the novel Phonovibrogram (PVG) approach was suggested to quantify the entire visible vocal fold vibrations (Lohscheller et al., 2007; Lohscheller et al., 2008a) expanding formerly introduced spatio-temporal plots (Westphal & Childers, 1983; Neubauer et al., 2001). In the PVG approach, contours of the oscillating vocal folds are segmented from video data and are transformed into a single colour coded PVG image. Depending on the underlying vocal fold vibrations, characteristic geometric patterns occur within a PVG which can be used for further clinical interpretation (Lohscheller & Eysholdt, 2008). PVG images can be regarded as fingerprints of vocal fold vibrations, enabling intuitional assessment of vocal fold vibrations (Eysholdt & Lohscheller, 2008). PVG analysis demonstrates that the complex two-dimensional vibratory patterns of vocal folds can robustly be described (Eysholdt & Lohscheller, 2008). It further establishes an objective basis for novel automatic analysis and classification approaches (Doellinger et al., 2009; Lohscheller et al. 2008b, Kunduk et al. 2010).

Within this work we propose a novel approach to achieve a fully automatic analysis of PVG images for detecting even slight alterations within underlying vocal fold vibrations: After segmenting the vocal fold vibrations from HSI and computing the appropriate PVG image matrix a set of novel PVG features will be introduced which describe the main characteristics of vocal fold dynamics. For investigating the sensitivity of the proposed PVG analysis approach the following physiological conditions were considered:

Vocal fold vibrations show individual patterns for each subject and can thus be highly variable between different patients. However, during voice production for a single subject the vocal fold vibrations show at specific voice intensity and fundamental frequency a reproducible dynamical behaviour. Within a subject, alterations of the fundamental frequency and/or intensity result into slight changes within vocal fold vibrations (Rovirosa et al., 2008). To obtain clinically relevant information about the physiology of a subject's voice the changes of vocal fold vibrations need to be traced. Accordingly, a computerized analysis procedure has to be sensitive enough to capture the individual changes within a subject. Hence, the validation of sensitivity of a computerized analysis approach needs to be performed within one single subject as changes of vocal fold vibrations between different subjects are not comparable.

According to the fulfilments above the sensitivity of PVG analysis was investigated by applying the PVG approach extensively to data sets obtained from a single healthy female subject. For data acquisition the subject was instructed to phonate at nine specified combinations of fundamental frequencies (low, normal, and high) and voice intensities (soft, normal, and loud). For each of these nine phonatory tasks twelve different high-speed sequences were obtained. Totally, 108 HSI sequences from this single subject could be processed. To obtain reliable results it is further of great importance to examine a healthy subject with no signs of voice disorders. Only for a single healthy subject it can be assumed that during the repeated examinations of a phonatory task the vocal fold vibrations are reproducible and do not change. Hence, the presence of pathologically caused and thus arbitrarily induced alterations of vocal fold vibrations can be excluded between recordings.

For further validation, simultaneously to the video data the emitted acoustic signal was recorded. From the acoustic data clinically used acoustic quality measures like Jitter, Shimmer, HNR, and SNR (Murphy, 1999; Zhang & Jiang, 2008) were computed allowing indirect conclusions about the vibrational behavior of vocal folds.
The results of this work will show that using PVG features in combination with a Support Vector Machine (SVM) even minor changes of vocal fold vibrations - caused by frequency and intensity alterations - can be highly robustly detected. Comparing the classification results gained by PVG features with results obtained from conventionally applied glottal as well as acoustic features will show the superiority of the novel PVG analysis approach.

2. Methods

2.1 Data collection

The KAY Elemetrics, High–Speed Digital Video System, Model 97, was used for data collection. Recordings were performed at a *2,000 fps* rate by using a specially designed, multi-port, super sensitive camera for eight seconds of recording. Gray scaled images were captured at 384Mb/sec into high-speed video RAM with a spatial resolution of *128 x 256* pixels. Images were obtained with a rigid *70°* endoscope (KAY Elemetrics, 9106) with a 300-watt-coldlight source (Olympus CLV-U20). The rigid laryngoscope was coupled to the high-speed digital camera head and endoscopy was performed as in conventional videostroboscopy. A microphone was placed *15 cm* from the lips to obtain the acoustic signal. This signal was fed through the KAY Elemetrics System for simultaneous recording of the endoscopic and acoustic signals (*50 KHz*). KAY Elemetrics, Rhino-Laryngeal Stroboscope (RLS 9100 B) and its microphone was used to determine F_o and the volume of the voice signal. The visual display on the system directed the subject for the maintenance and consistency of the desired F_o and volume for each phonatory task.

2.2 Subject and phonatory tasks

One female subject's voice was recorded with HSI for this study. The subject was non smoker and had no known history of neurological disease, laryngeal surgery, prior/or existing laryngeal disorders, voice problems at the time of data collection nor observed neither reported speech/language impairment. The HSI and acoustic recordings were simultaneously acquired while the subject was producing the vowel /i/ at the following fundamental frequency (F_0) / intensity (I) combinations:

- low F_0 (F1) at soft (I1), normal (I2), and high (I3) intensity,
- normal F_0 (F2) at soft (I1), normal (I2), and high (I3) intensity,
- high F_0 (F3) at soft (I1), normal (I2), and high (I3) intensity,

resulting in 9 different phonatory tasks. For all F_0/I combinations four phonation trails were performed. Within each recorded trail three different intervals of phonation were present. Each interval contained a voice onset followed by sustained phonation of at least one second being divided by short periods of silence. Hence, for each F_0/I combination 12 phonation sequences were available. For later analysis purposes the following class system is introduced (F→ Frequency, I→Intensity):
3 Frequency classes

$$CF1:=\{F1I1, F1I2, F1I3\};\ CF2:=\{F2I1, F2I2, F2I3\};\ CF3:=\{F3I1, F3I2, F3I3\}. \quad (1)$$

3 Intensity classes

$$\text{CI1:=\{F1I1, F2I1, F3I1\}; CI2:=\{F1I2, F2I2, F3I2\}; CI3:=\{F1I3, F2I3, F3I3\}.} \quad (2)$$

9 Combined Frequency/Intensity classes

$$\begin{aligned} &CS1 := \{F1I1\};\ CS2 := \{F1I2\};\ CS3 := \{F1I3\}; \\ &CS4 := \{F2I1\};\ CS5 := \{F2I2\};\ CS6 := \{F2I3\}; \\ &CS7 := \{F3I1\};\ CS8 := \{F3I2\};\ CS9 := \{F3I3\}. \end{aligned} \quad (3)$$

2.3 Selection of sequences

Within the acoustic signals the intervals of sustained phonation were identified by visual inspection. Within each interval a time section of 1 second was selected. The identical section was analyzed in high speed video data. The sequence length of one second time (> 150 glottal cycles) was in accordance with previous studies who suggested approx. 130 - 190 cycles (Karnell, 1991). Thus, altogether 108 pairs of high-speed and acoustic data sets were available (Tab. 1), reflecting isochronal information about vibratory characteristics of the voice generator (high-speed data) and the acoustic outcome (voice signal). Only in four cases the video data could not be further processed due to low image quality. To ensure, that possible occurring differences between recordings were only induced by the different phonation task, the recordings were performed within a day. As far as we know these data represent the most exhaustive examination of a single subject's vocal fold dynamics using HSI.

Intensity/F0	Low(F1)	Normal(F2)	High(F3)	**CI1-CI3**
Soft(I1)	4(12)	4(12)	4(12)	**12(36)**
Normal(I2)	4(9)	4(11)	4(12)	**12(32)**
Loud(I3)	4(12)	4(12)	4(12)	**12(36)**
CF1-CF3	**12(33)**	**12(35)**	**12(36)**	**36(104)**

Table 1. Applied Data. Overview of the performed 36 recordings which equals 108 sequences. From these sequences 104 could be analysed for acoustic and dynamical data.

2.4 PVG parameters describing vocal fold dynamics

2.4.1 Image processing

The vibrating edges of both vocal folds were extracted alongside their entire glottal length to analyze the laryngeal vibrations during phonation (Lohscheller et al., 2007). Information at each specific position of vocal folds is required to obtain detailed information about the vibration characteristics at dorsal, medial and ventral parts of vocal folds. For this purpose an extensively evaluated image segmentation procedure was applied (Lohscheller et al., 2007). The procedure delivers the left/right vocal fold edge contours $c_{L/R}(t)$, the glottal area $a(t)$, the location of anterior/posterior glottal ending $A(t)$ and $P(t)$ as well as the glottal main axis $l(t)$. A typical result of a segmented high-speed image is shown in Fig. 2.

Since the segmentation accuracy highly affects the following analysis, the quality of the results was visually monitored. For this purpose, within a movie viewer the segmented vocal fold contours were displayed. Further, for identifying potential faulty segmented

images (outliers) the glottal area *a(t)* was displayed within a diagram, see Fig. 2. Thus, in case of imprecise results, a re-segmentation of the high speed videos could be performed.

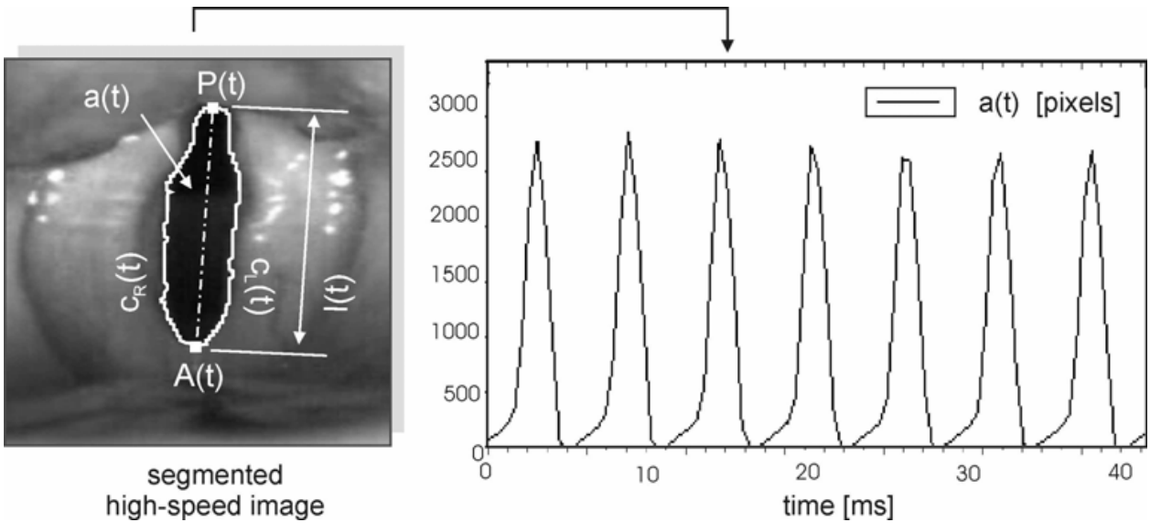


Fig. 2. Glottal area function. Left: Segmented image of a high-speed video. The extracted vocal fold edges are superimposed and are used to verify visually the accuracy of the segmentation results. Right: The glottal area waveform *a(t)* is monitored to detect faulty segmented images within a segmented video sequence.

In this study, the image processing procedure was applied only when the glottal length was fully visible during one second. From all 108 data sets 104 sequences each containing 2,000 consecutive images were successfully processed resulting in 208,000 segmented images. In all cases satisfactory segmentation accuracy were obtained, which are comparable to the example shown in Fig. 3.

2.4.2 Generation of phonovibrograms

For visualizing the entire vibration characteristics of both vocal folds the Phonovibrogram (PVG) was applied which was described in detail before (Lohscheller et al., 2008a). The principles of PVG computation are shortly summarized in Fig. 3. For each image of a high-speed video, the segmented glottal axis is longitudinally split and the left vocal fold contour is turned *180°* around the posterior end. Following, the distances $d^{L,R}(y,t)$ between the glottal axis and the vocal fold contours are computed; $y \in [1,\ldots,Y]$ with $Y=256$ denotes the spatial sampling of glottal axis. The distance values are stored as column entries of a vector and become color coded. The distance magnitudes are represented by the pixel intensities and two different colors. If vocal fold edges cross the glottal axis during an oscillation cycle the pixel is encoded by the color blue, otherwise the color red was used to indicate the distance from the glottal axis. A grayscale representation (black: vocal fold edges are at the glottal midline, white vocal fold edges have a distance to the glottal midline) of the originally colored PVG is given in Fig. 3. The entire vibration characteristics of both vocal folds are captured within one single PVG image by iterating the described procedure for an entire sequence and consecutively arranging the obtained vectors to a two-dimensional matrix. The left vocal fold is represented in the upper and the right vocal fold in the lower horizontal plane of the PVG, respectively. The PVG enables at the same time an assessment

of the individual vibration characteristics for each vocal fold and gives evidence about left/right and posterior/anterior vibration asymmetries as well as predications about the temporal stability of the vibration pattern.

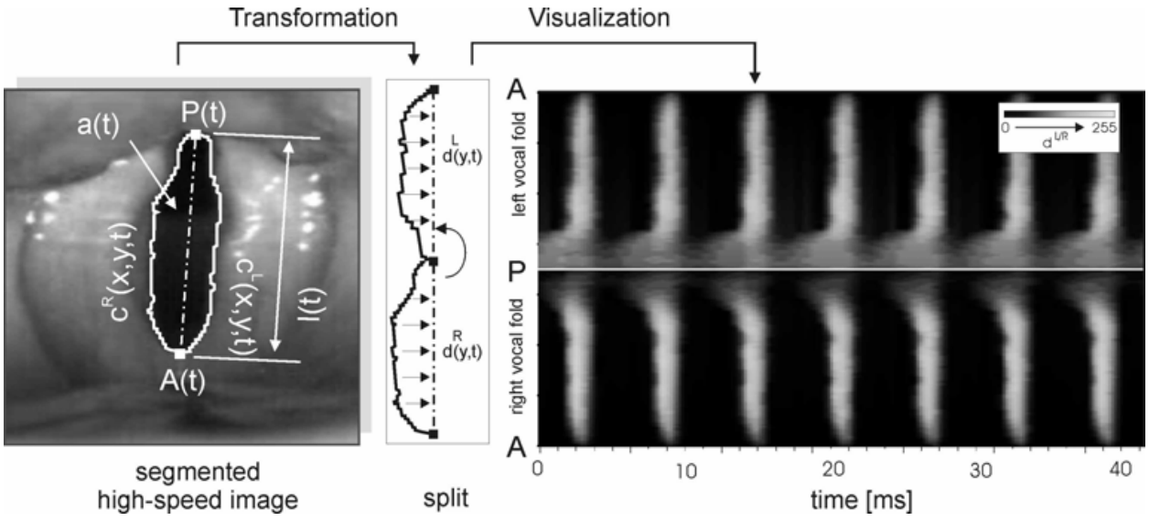


Fig. 3. PVG generation. 1) Segmentation of HS video. 2) Transformation of extracted vocal fold contours and computation of the distance values $d^{L,R}(y,t)$ which represent the distances from the vocal fold edges to the glottal midline. 3) Color coding of distance values for an entire high-speed video result into a PVG image comprising the entire vibration dynamics of both vocal folds in a single image (PVG is shown as grayscale image).

2.4.3 Analysis of vocal fold vibrations

PVG pre-processing: Phonovibrograms obtained from high speed sequences contain multiple reoccurring geometric patterns representing consecutive oscillation cycles of vocal folds. In order to describe the vibratory characteristics of vocal folds objectively, the 104 PVGs were pre-processed as follows: Firstly, for the left and right vocal fold unilateral PVGs are computed, denoted as $uPVG^{L/R}$ which are in the following regarded as two-dimensional functions $v^L(k,y)$ and $v^R(k,y)$ with $k \in \{1,\ldots,K\}$ and $K=2{,}000$ representing the number of frames within a sequence. From the unilateral PVGs the Glottovibrogram (GVG) is derived $v^G(k,y) = v^L(k,y) + v^R(k,y)$ which represents the glottal width (distances between the vocal folds) at each vocal fold position y over time, Fig. 4. In a subsequent step, the $uPVGs$ and the GVG are automatically subdivided into a set of single PVG/GVG cycles, Fig. 4 right. A frequency analysis and peak picking strategy in the image domain is performed for the cycle identification (Lohscheller et al., 2008a).

Finally, the obtained single cycle PVGs are normalized to a constant width and height which are denoted $sPVG^L_i$, $sPVG^R_i$, $sGVG_i$, with $i \in \{1,\ldots,I^{L,R,G}\}$ and $I^{L,R,G}$ representing the number of cycles within the corresponding Phonovibrogram. Hence, vocal fold vibrations can be described by a set of the three functions

$$d_i^L(t,y) \;:= sPVG_i^L \;,\; d_i^R(t,y) \;:= sPVG_i^R \;,\; g_i(t,y) \;:= sGVG_i \tag{4}$$

with $t \in \{1,\ldots,T\}$ where $T=256$ represents the normalized cycle length. In the following, the index $\alpha:=\{L,R\}$ is introduced to distinguish the functions $d^\alpha_i(t,y)$ representing the left and

right vocal fold. Both, the unilateral as well as the normalized PVGs form the basis for the following analysis to obtain detailed information about vocal fold dynamics.

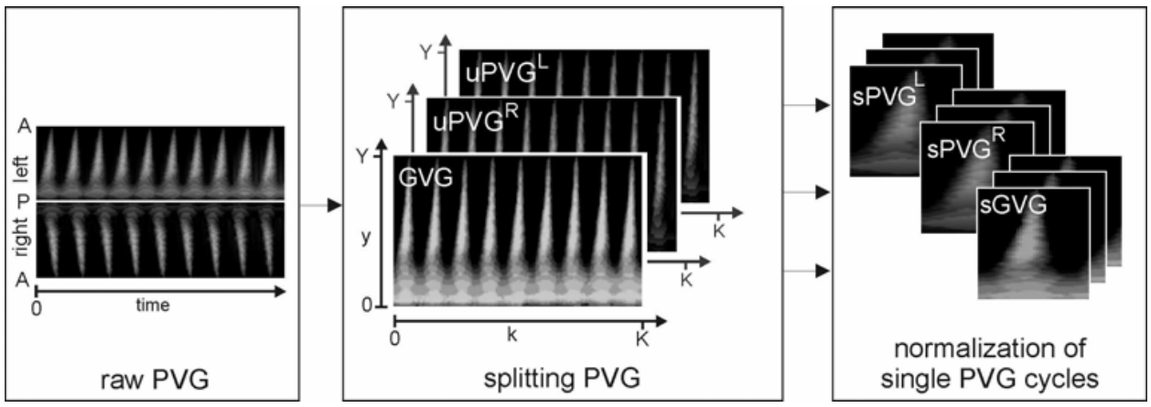


Fig. 4. Pre-Processing. From a raw PVG (left) so-called unilateral PVGs are computed (middle) which are further subdivided into a set of normalized single cycle *PVGs* (right).

Extraction of symmetry features: In order to describe the overall behavior of vocal fold dynamics the PVGs are analyzed as follows. At each glottal position y the 1D-power spectrum

$$\mathbf{P}^{\alpha}(f,y) := |\, FFT\{v^{\alpha}(k,y)\}\,| \quad \forall y \tag{5}$$

is calculated by Fast Fourier Transform algorithm (*FFT*). Due to settings, corresponding frequency resolution of the spectral components were 1 *Hz*. Fundamental frequencies $\mathbf{f}_0^{\alpha}$ are estimated by identifying the maxima within the discrete power spectra

$$\mathbf{f}_0^{\alpha} := \arg\max_f P^{\alpha}(f,y) \quad \forall y. \tag{6}$$

By defining the feature vector

$$\boldsymbol{\theta} := \theta(y) := \frac{\mathbf{f}_0^L}{\mathbf{f}_0^R} \quad \forall y \tag{7}$$

frequency differences between the left and right vocal fold as well as differences alongside the glottal axis are captured. If lateral (i.e. left/right) fundamental frequencies are identical the feature vector

$$\boldsymbol{\upsilon} := \upsilon(y) := \varphi\{\mathbf{P}^L(\mathbf{f}_0^L, y)\} - \varphi\{\mathbf{P}^R(\mathbf{f}_0^R, y)\} \quad \forall y \tag{8}$$

describes the phase delays between the left and right vocal fold.

The left/right vibration asymmetry is further described by introducing the mean relative amplitude ratios $\bar{a}(y)$ which are computed as follows. Within the $sPVG^{L,R}$ the points in time

$$\mathbf{T}_{y,i}^{\alpha\,\max} := \arg\max_t d_i^{\alpha}(t,y) \quad \forall \alpha, y, i \tag{9}$$

along the vocal fold length are identified when the maximum vocal fold deflections occur. By identifying the time points of minimal vocal fold deflection

$$\mathbf{T}_{y,i}^{\alpha\,\min} := \arg\min_{t} d_i^{\alpha}(t,y) \quad \forall \alpha, y, i \tag{10}$$

the relative peak-to-peak amplitudes

$$\mathbf{A}_{y,i}^{\alpha} := d_i^{\alpha}(\mathbf{T}_{y,i}^{\alpha\,\max}, y) - d_i^{\alpha}(\mathbf{T}_{y,i}^{\alpha\,\min}, y) \quad \forall \alpha, y, i \tag{11}$$

can be defined which are independent from the absolute position of the glottal axis. The mean relative amplitude ratios

$$\bar{\mathbf{a}} := \bar{a}(y) = \overline{\left(\frac{\mathbf{A}_{y,i}^{L}}{\mathbf{A}_{y,i}^{R}} \right)} \quad \forall y \tag{12}$$

and corresponding standard deviations $\boldsymbol{\sigma_a}$:=σ_a(y) serve as features to describe left/right asymmetries as well as the stability of vibrations at each position of the vocal folds. The obtained parameters are merged to the symmetry feature vector s (Eqs. (7),(8),(12)):

$$\mathbf{s} := [\boldsymbol{\theta}, \boldsymbol{\upsilon}, \bar{\mathbf{a}}, \boldsymbol{\sigma}_{\mathbf{a}}]. \tag{13}$$

Extraction of glottal features g: In order to capture characteristics of the glottal dynamics within the oscillation cycles, the following parameters are extracted from the normalized GVG matrices $g_i(t,y)$. Firstly, the maximum glottal area of each oscillation cycle i is determined as

$$\boldsymbol{\rho}_i = \max_{t} \sum_{y=1}^{Y} g_i(t,y) \quad \forall t, i. \tag{14}$$

The feature

$$\sigma_{\rho} = \sqrt{Var(\boldsymbol{\rho}_i)} \tag{15}$$

describes the stability of the glottal vibratory cycles over time. Subsequently, the open quotients $\boldsymbol{OQ}_{y,i}$ are defined for each glottal position i as duration of open phase divided by duration of complete glottal cycle and are computed as

$$\mathbf{OQ}_{y,i} = \left(\sum_{t} \hat{g}_i(t,y) \right) / T \quad \forall y, i; \tag{16}$$

with

$$\hat{g}_i = \begin{cases} 1 & g_i(t,y) > 0 \quad \forall t. \\ 0 & otherwise. \end{cases} \tag{17}$$

The mean values

$$\overline{\mathbf{oq}} = \frac{1}{I}\sum_{i}^{I}\mathbf{OQ}_{y,i} \quad \forall y \tag{18}$$

and standard deviations

$$\boldsymbol{\sigma}_{oq} = \sqrt{Var(\mathbf{OQ}_{y,i})} \quad \forall y \tag{19}$$

are used as features describing the stability of the glottal opening behavior at each position alongside the glottal axis (*Var* symbolizes the variance). Analogously, the mean speed quotients $\overline{\mathbf{sq}}$ and the corresponding standard deviations $\boldsymbol{\sigma}_{sq}$ are computed describing the mean glottal vibratory shape and its stability over time (Jiang et al., 1998).
Finally, the glottal closure insufficiencies

$$\mathbf{gci}_i = \frac{\min\limits_{t}\sum\limits_{y}^{Y}\hat{h}_i(t,y)}{Y} \qquad \forall t,i. \tag{20}$$

are derived using

$$\hat{h}_i = \begin{cases} 1 & g_i(t,y) > 0 \quad \forall y. \\ 0 & otherwise. \end{cases} \tag{21}$$

which are identifiable for each oscillation cycle *i*. The supplemental features $\overline{gci}$ and σ_{gci} describe the mean glottal closure insufficiency and its stability for the entire high-speed sequence. The glottal parameters are merged to the glottal feature vector (Eqs. (15),(18),(19)):

$$\mathbf{g} := [\sigma_{\rho}, \overline{\mathbf{oq}}, \boldsymbol{\sigma}_{oq}, \overline{\mathbf{sq}}, \boldsymbol{\sigma}_{sq}, \overline{gci}, \sigma_{gci}]. \tag{22}$$

Extraction of geometric PVG feature ω: Besides the conventional symmetry and glottal parameters we propose a novel way for describing vocal fold vibrations by quantifying the geometric structure within $sPVG^{\alpha}$ images. The main vibration characteristics of a vocal fold can be described by extracting representative contour lines from the $sPVG^{\alpha}$ images. This is done by determining the oscillatory states *n* during the opening ($t < \mathbf{T}_{y,i}^{\alpha\,\max}$) and closing ($t > \mathbf{T}_{y,i}^{\alpha\,\max}$) phases where vocal folds reach a certain percentage of relative deflection

$$\mathbf{A}_{y,i}^{\alpha\,n} := \frac{n}{100}\mathbf{A}_{y,i}^{\alpha}, \quad n \in [0,100]. \tag{23}$$

Hence, the set of vectors

$$\mathbf{O}_{y,i}^{\alpha\,n} := \arg_{x}(d_i^{\alpha}(x,y) = \mathbf{A}_{y,i}^{\alpha\,n}), \text{ with } t < \mathbf{t}_i^{\alpha\,\max} \quad \forall \alpha, y, i. \tag{24}$$

$$\mathbf{C}_{y,i}^{\alpha\,n} := \arg_{x}(d_i^{\alpha}(x,y) = \mathbf{A}_{y,i}^{\alpha\,n}), \text{ with } t > \mathbf{t}_i^{\alpha\,\max} \quad \forall \alpha, y, i. \tag{25}$$

describe temporal and spatial propagation of each vocal fold at different oscillation states during glottal opening $\mathbf{O}_{y,i}^{\alpha\,n}$ and closing $\mathbf{C}_{y,i}^{\alpha\,n}$. In order to get a comprehensive

understanding of the entire vibration cycle, multiple contour lines are extracted at different oscillation states. Fig. 5 shows exemplarily extracted contour lines at *n=(30,60,90)* for the left and right vocal fold during a single oscillation cycle.
The functional characteristics

$$\mathbf{PO}_{y,i}^{\alpha n} := d_i^{\alpha}(t,y)\Big|_{\mathbf{o}_i^{\alpha n}} \quad \mathbf{PC}_{y,i}^{\alpha n} := d_i^{\alpha}(t,y)\Big|_{\mathbf{c}_i^{\alpha n}} \quad \forall \alpha, y, i \tag{26}$$

of $sPVG^{\alpha}$ at positions $\mathbf{O}_{y,i}^{\alpha n}$ and $\mathbf{C}_{y,i}^{\alpha n}$ of the contour lines give precise information on actual deflection of the vocal folds. As features which describe the average vibratory pattern of vocal folds, the means for the contour lines *n=(30,60,90)*, the deflection characteristics and their time indices

$$\overline{\mathbf{O}_{y,i}^{\alpha n}}, \overline{\mathbf{PO}_{y,i}^{\alpha n}}, \overline{\mathbf{C}_{y,i}^{\alpha n}}, \overline{\mathbf{PC}_{y,i}^{\alpha n}}, \tag{27}$$

are computed for all cycles *i*. The vibration stability is captured by the corresponding standard deviations

$$\boldsymbol{\sigma}(\mathbf{O}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PO}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{C}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PC}_{y,i}^{\alpha n}). \tag{28}$$

The Euclidian-Norm $\|\ \|_2$ between the mean positions of the contour lines

$$N_{O,C}^{n} = \left\| \overline{\mathbf{O}_{y,i}^{Ln}} - \overline{\mathbf{O}_{y,i}^{Rn}} \right\|_2 \quad \forall n \tag{29}$$

describes deviations between the mean left and right vocal fold vibration patterns. Finally, all parameters (Eqs. (27),(28),(29)) are merged to the PVG feature vector

$$\boldsymbol{\omega} := [\overline{\mathbf{O}_{y,i}^{\alpha n}}, \overline{\mathbf{PO}_{y,i}^{\alpha n}}, \overline{\mathbf{C}_{y,i}^{\alpha n}}, \overline{\mathbf{PC}_{y,i}^{\alpha n}}, \boldsymbol{\sigma}(\mathbf{O}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PO}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{C}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PC}_{y,i}^{\alpha n}), N_{O,C}^{n}]. \tag{30}$$

The entire vocal fold dynamics extracted from one high speed sequence can be described by merging the introduced features for left-right symmetry, glottal and PVG characteristics (Eqs. (13),(22),(30)) to the feature vector

$$\boldsymbol{\beta} := [\mathbf{s}, \mathbf{g}, \boldsymbol{\omega}]. \tag{31}$$

The feature vector $\boldsymbol{\beta}$ represents vocal fold dynamics at each position y along the glottal axis with $y \in \{1,\ldots,Y\}$. In order to reduce the dimensionality of the parameter space for further analysis, the feature vector is reduced to $y \in \{1,\ldots,12\}$ by computing average values. Hence, for an effective vocal fold length of *1 cm* the feature vector represents the average oscillation dynamics within *0.9 mm* sections of the vocal length which constitutes sufficient accuracy.
Acoustic voice quality measures: For the nine frequency/intensity phonatory tasks also the acoustic voice signals were analyzed. The selected acoustic sequences correspond to the time intervals of the analyzed video data. From the selected intervals 10 voice quality measures were derived using Dr.Speech-Tiger-Electronics/Voice-Assessment-3.2 software (www.drspeech.com). The computed parameters describe temporal voice properties as cycle duration stability (Jitter, STD F_0, STD Period, F_0 tremor), amplitude stability (Shimmer, STD

Ampl., Amp. Tremor), harmonic to noise ratio (HNR), signal to noise ratio (SNR), and normalized noise energy (NNE). The nine different frequency/intensity classes are given by the measured sound pressure level (*SPL[dB]*) and mean fundamental frequency (Mean F_0*[Hz]*), Tab. 2.

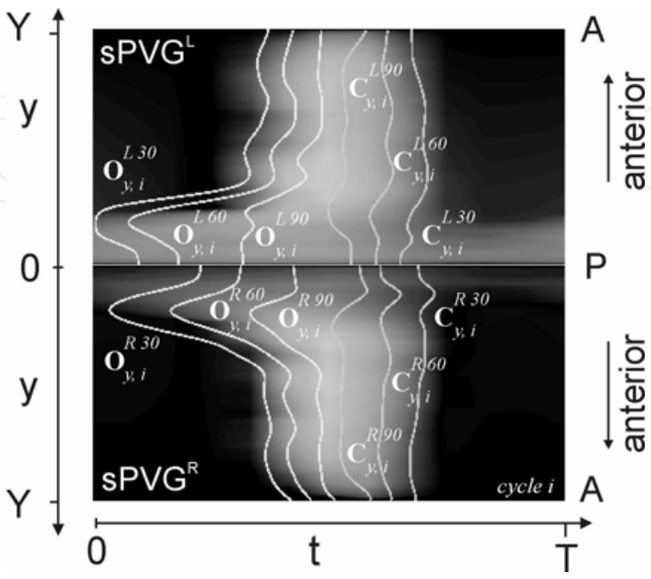


Fig. 5. The contour lines **O** (opening phase) and **C** (closing phase) describe the main characteristics of $sPVG^{\alpha}$ geometry. The contours represent the spatio-temporal positions of vocal fold edges at the oscillation states *n=(30,60,90)* for the left and right vocal fold. The *n* value corresponds to the percentage of open and closed positions.

	CS1	**CS2**	**CS3**	**CS4**	**CS5**	**CS6**	**CS7**	**CS8**	**CS9**
No.Sequ.	12	9	12	12	11	12	12	12	12
SPL(dB)	*59,0 ±0,8*	*63,3 ±0,5*	*72,5 ±1,7*	*58 ±0*	*63 ±0*	*75 ±0*	*58,3 ±0,5*	*64,3 ±1,4*	*71 ±0,9*
Mean F0 (Hz)	*153 ±3*	*160 ±4*	*201 ±2*	*182 ±4*	*193 ±4*	*231 ±8*	*318 ±5*	*328 ±8*	*328 ±5*

Table 2. Mean values and standard deviations for the different fundamental frequencies [mean F_0] and voice intensities [sound pressure level (*SPL[dB]*)] representing the nine different phonatory tasks CS1-CS9.

Classification of different phonation conditions: Due to the high number of PVG parameters conventional statistics and correlation analysis is not appropriate to identify potential parameter changes between the different phonation conditions. Thus, to explore the influence of intensity and frequency alterations within the parameter sets a nonlinear classification approach was applied (Hild et al., 2006; Selvan & Ramakrishnan, 2007; Lin, 2008).

The following hypothesis was investigated: if a classifier is capable of distinguishing between different phonatory classes it can be concluded that intensity and frequency variations are actually present within the observed vocal fold dynamics represented by the introduced feature sets.

For classification of the PVG features, a nonlinear support vector machine (SVM) was used (Duchesne et al., 2008; Kumar & Zhang, 2006). For the SVM, a Gaussian radial basis function kernel (RBF) was chosen (Vapnik, 1995). Appropriate SVM parameters were determined by an evolutionary strategy optimization procedure (Beyer & Schwefel, 2002). The parameter space of SVM, cost parameter and the width of the RBF kernel was automatically searched in order to obtain best classification results (Hsu et al., 2003). The models' classification accuracy was evaluated via 10-fold cross-validation with stratification (Kohavi, 1995).
In order to compare PVG result with conventionally used measures the classifier was also applied to traditional glottal and symmetry parameters as well as to the ten acoustic voice quality measures.

3. Results

3.1 Validation of data acquisition

For a reliable interpretation of the later classification results it is essential to verify that the data acquisition representing the nine different phonatory tasks effectively succeeded. Tab. 2 shows the means and standard deviations for the different sound pressure levels (*SPL*) and fundamental frequencies (mean F_0) for all nine phonatory tasks. Already the very small standard deviations of the *SPL* and mean F_0 within the classes CS1-CS9 prove the high consistency of the data acquisition which included the repeated recording of the different phonatory tasks. Applying statistical analysis (Kolmogorov-Smirnov-Tests following *t*-Tests or Mann-Whitney-U-Tests) it could be shown that for frequency classes LOW (CF1), NORMAL (CF2), and HIGH (CF3) (Eq. (1)) the fundamental frequencies were significantly ($p<0.05$) different. Also for intensity classes SOFT (CI1), NORMAL (CI2), and LOUD (CI3) (see Eq. (2)) the intensity values were computed significantly ($p<0.05$) different.

3.2 SVM classification of vocal fold vibrations

Exemplarily, Tab. 3 shows SVM classification results obtained for frequency classes CF1-CF3. The *Class Precision* reflects the percentage of the correct allocation: 30 out of 104 sequences were predicted as low (CF1). From these 30, three sequences were wrongly assigned to the class low (being actually in class CF2) resulting in 90% *Class Precision*. In contrast, the *Class Recall* reflects the percentage of how many members of the class were allocated towards the class. Here, 35 out of 38 normal sequences were correctly assigned to class CF2 whereas three sequences were predicted to class CF1. This results in a *Class Recall* accuracy of 92.1%. The *Overall Accuracy* for all classes is 94.18% ±6.53% which represents the mean performance of the classifier which is in the following used for interpretation purpose.

	True Low	True Normal	True High	***Class Precision***
Low (CF1)	27	3	0	90.0%
Normal (CF2)	3	35	0	92.1
High (CF3)	0	0	36	100.0%
Class Recall	90.0%	92.1%	100.0%	

Table 3. Classification result of the SMV of the intensity class problem CF1-CF3 using the entire feature vector from eq. (31). The overall classification accuracy amounts approx. 94%.

Using the parameters captured within the feature vector $\boldsymbol{\beta}:=[\mathbf{s},\mathbf{g},\boldsymbol{\omega}]$ (Eq. (31)) the SVM reached a classification accuracy of 95.1%±6.7% for the frequency class problem (CF1-3),

97.3%±4.2% for the intensity class problem (CI1-3), and 94.2%±9.1% for the nine class problem (CS1-CS9). This very high classification accuracy was obtained just by parameters describing vocal fold dynamics extracted from the high speed videos.
In order to investigate which parameters can be made responsible for the high performance of the classifier, the SVM was individually applied to components [**s**], [**g**] and [**ω**] as well as to the combinations [**s**,**g**], [**g**,**ω**], [**s**,**ω**]. The results are summarized in Fig. 6. The conventional symmetry [**s**] and glottal parameters [**g**] achieved classification accuracy of only 15.5%±4.9% and 40.5%±10.5% for the nine class problem. Likewise, the classification accuracies for the frequency and intensity class problems were significantly reduced. Contrarily, very high classification accuracy was obtained using the new introduced PVG features [**ω**]. Applying exclusively the PVG features [**ω**] a classification accuracy of 85.5%±7.7% for the nine class problem, 96.2%±4.7% for the frequency class problem, and 91.6%±7.6% for the intensity class problem was obtained.

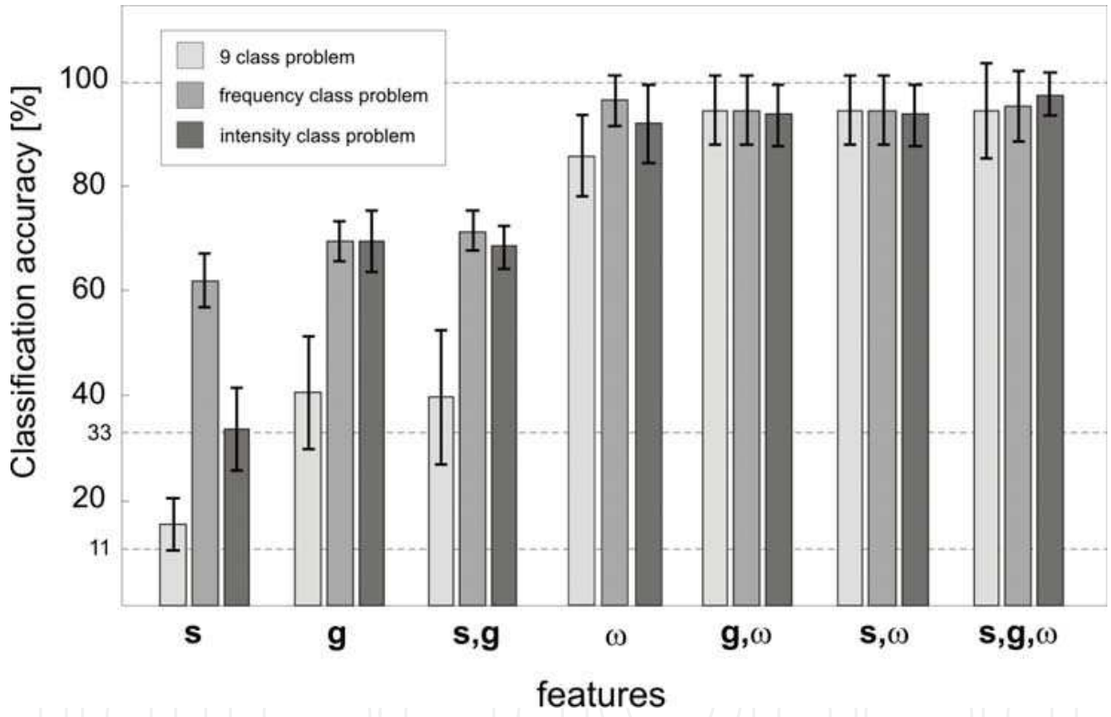


Fig. 6. Mean classification accuracies and standard deviations achieved by applying conventional symmetry [**s**], glottal [**g**] and PVG [**ω**] parameters using a support vector machine (SVM) classification approach with stratified 10-fold cross-validation. The highest classification accuracy is obtained by the new introduced PVG features [**ω**].

As the PVG feature vector contains information derived from different oscillation states ($\mathbf{O}_{y,i}^{\alpha n}$, $\mathbf{C}_{y,i}^{\alpha n}$) it was further investigated which oscillation state delivers the most valuable information needed for classifying vocal fold vibrations. For this purpose, the SVM was applied to different oscillation parts *n={30,60,90}* of the feature vector [**ω**]. Fig. 7 summarizes the achieved classification accuracies obtained by *n={[30,60],[60,90],[30,60,90]}.* Using the single oscillation states *n={[30],[60],[90]},* already a mean classification accuracy of 58.2%±9.9% could be obtained for the nine class problem which exceeds considerably the

classification rates obtained by the conventional symmetry [**s**] and glottal [**g**] parameters as shown in Fig. 6. The classification accuracies by applying combined oscillation states *n={[30,60],[60,90] ,[30,60,90]}* are significantly improved.

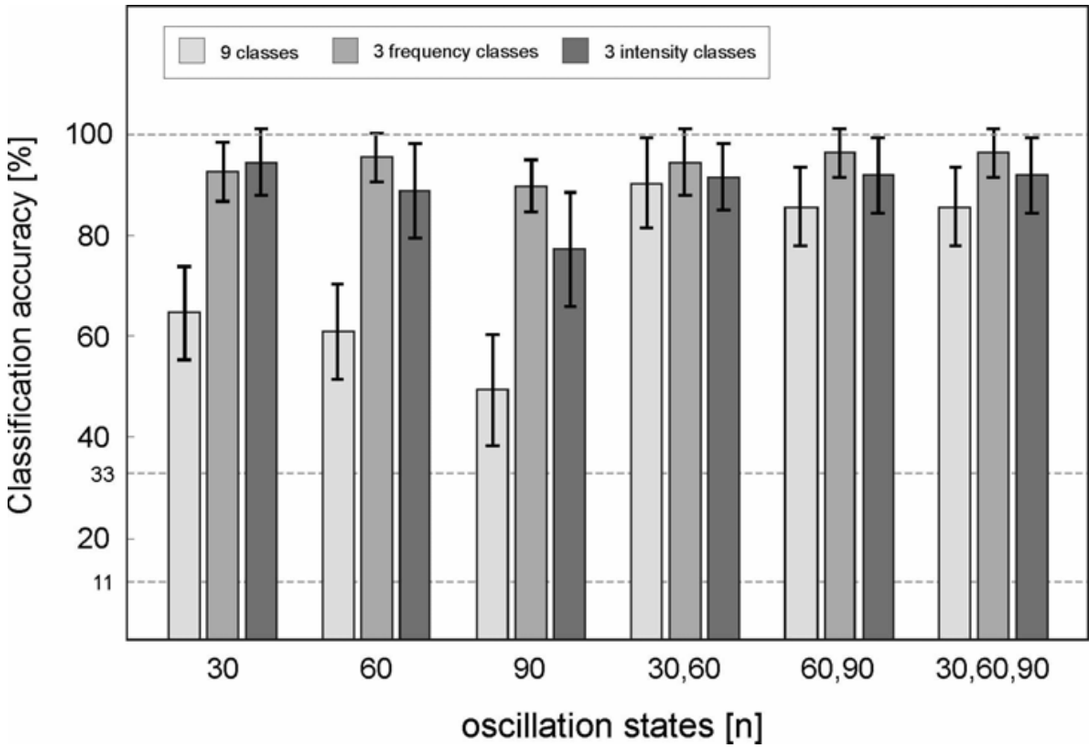


Fig. 7. Mean SVM classification accuracies and standard deviations achieved by applying part of the PVG features vector [**ω**] representing different oscillation states *n={30,60,90}*. Highest classification accuracy is obtained by a combination of the different oscillation states.

In a final step it was investigated which PVG components contribute most to the classification accuracy. For this purpose the feature vector [**ω**] (eq. (30)) was divided into parameter groups representing the average vibration type $[\boldsymbol{\omega}_1] := (\overline{\mathbf{O}_{y,i}^{\alpha n}}\ \overline{\mathbf{C}_{y,i}^{\alpha n}})$, the average deflection characteristics $[\boldsymbol{\omega}_2] := (\overline{\mathbf{PO}_{y,i}^{\alpha n}}\ \overline{\mathbf{PC}_{y,i}^{\alpha n}})$, the average lateral vibration symmetry $[\boldsymbol{\omega}_3] := (N_{O,C}^{n})$, and the average temporal stability of vocal fold vibrations $[\boldsymbol{\omega}_4] := (\boldsymbol{\sigma}(\mathbf{O}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PO}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{C}_{y,i}^{\alpha n}), \boldsymbol{\sigma}(\mathbf{PC}_{y,i}^{\alpha n}))$. Figure 8 shows the classification accuracies obtained by the different parts of the feature vector [**ω**].

The isolated consideration of the average vibration type [$\boldsymbol{\omega}_1$] results into the highest classification accuracy of 52.8%±6.8% for the nine class problem and a mean accuracy of 85.1%±10.58% for the frequency and intensity class problems. By comparing the results in Fig. 6 and Fig. 8, it can be seen, that information about the mean vibration type (Fig. 8) already gives better classification results than information about the conventional parameters as speed quotient, open quotient, glottal closure insufficiency (Fig. 6). Information about vocal fold deflection amplitudes [$\boldsymbol{\omega}_2$], left/right discrepancies [$\boldsymbol{\omega}_3$] and

vibration instabilities [ω_4] do not reach the same level of classification accuracy. However, combining all PVG features increases considerably the classification accuracy of up to 96.2%±4.7% for the frequency class problem.

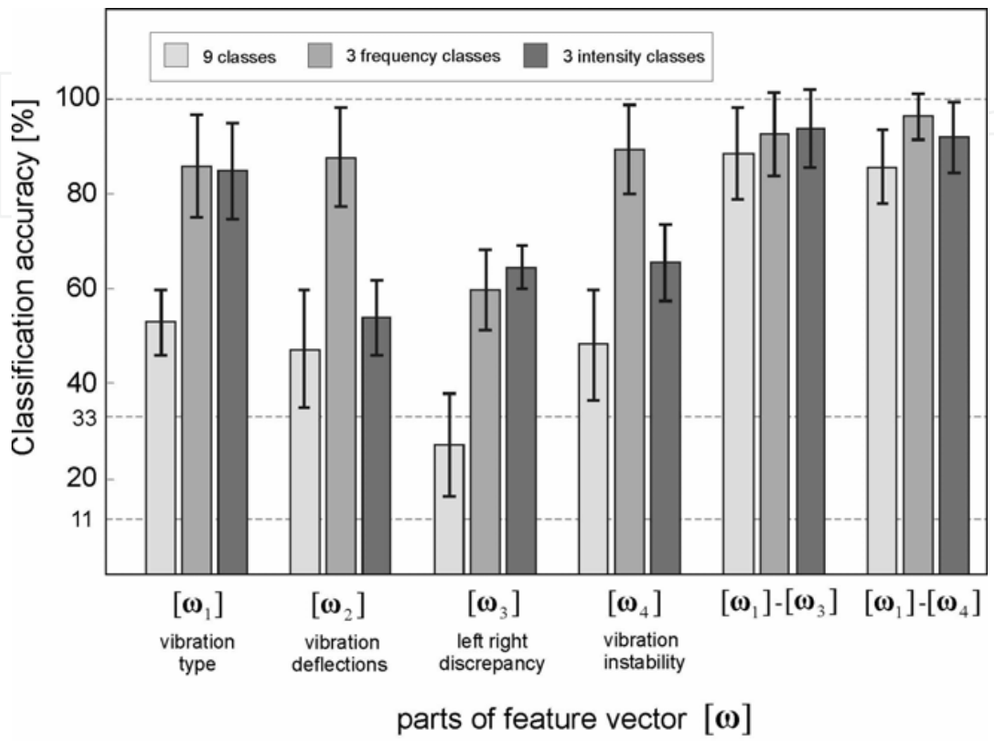


Fig. 8. Results show the comparison between the different features within the PVG parameters. The PVG parameters are split into groups representing the spatio-temporal vibration type, information about vibration amplitudes and symmetry as well as vibration instabilities. Fusing all information ω_1-ω_4 to a common feature vector results highest classification performance (i.e. frequency classes). The performance of the different classification results shows that the more precisely the vocal fold dynamics is described using a combination of several PVG features the better the dynamical changes of vocal fold dynamics can be captured.

3.3 SVM classification of the acoustic signal

To give an overview of the acoustic measures, Tab. 4 shows the means and standard deviations for all 10 computed acoustic voice quality parameters used for classification.
Table 5 summarizes the classification results for acoustic parameters. The best classification performance (93.45%) was achieved for the frequency class problem (CF1-CF3). The accuracy for the three class intensity problem (85.64%) was just slightly higher than accuracy for the combined nine class problem (83.73%). In contrast to the classification results obtained using the PVG parameters the acoustic parameters reached lower classification accuracies. Nevertheless, for the nine class problem still a classification accuracy of more than 80% could be achieved. It proves that even for a single subject frequency and intensity changes of the voice signal influence voice quality outcome measures.

	CS1	CS2	CS3	CS4	CS5	CS6	CS7	CS8	CS9
Jitter (%)	0,30 ±0,06	0,13 ±0,02	0,12 ±0,03	0,21 ±0,04	0,21 ±0,08	0,10 ±0,03	0,21 ±0,05	0,11 ±0,02	0,12 ±0,05
Shimmer (%)	2,17 ±0,44	1,07 ±0,17	0,98 ±0,35	1,69 ±0,36	1,48 ±0,25	0,88 ±0,45	1,68 ±0,42	0,85 ±0,22	0,84 ±0,11
HNR (%)	23,4 ±1,6	30,2 ±1,0	33,2 ±1,7	27,6 ±1,9	28,8 ±1,3	32,4 ±3,0	28,6 ±2,2	34,6 ±1,7	28,4 ±1,6
SNR (%)	23,4 ±1,6	30,2 ±1,0	33,2 ±1,7	27,6 ±1,9	28,8 ±1,3	32,4 ±3,0	28,7 ±2,2	34,6 ±1,7	28,4 ±1,6
NNE (%)	-3,0 ±1,5	-13,6 ±1,9	-17,1 ±2,7	-8,6 ±4,0	-11,1 ±2,9	-21,2 ±1,1	-9,3 ±2,6	-13,1 ±2,5	-21,5 ±2,3
STD F0 (Hz)	1,4 ±0,5	1,0 ±0,3	1,4 ±0,5	1,4 ±0,4	1,5 ±0,7	1,5 ±0,4	2,8 ±1,5	2,3 ±0,6	1,6 ±0,3
STD Period (ms)	0,06 ±0,02	0,04 ±0,01	0,04 ±0,01	0,04 ±0,01	0,04 ±0,02	0,03 ±0,01	0,03 ±0,01	0,02 ±0,01	0,02 ±0,01
Mean Amp (%)	86 ±4,8	92 ±2,1	91 ±3,1	86 ±4,1	90 ±3,1	90 ±3,1	85 ±5,5	88 ±4,3	93 ±2,3
STD Amp. (%)	5,9 ±1,7	3,5 ±1,3	4,4 ±1,5	6,1 ±1,5	5,4 ±2,1	4,8 ±1,8	6,2 ±2,1	5,1 ±1,5	2,7 ±0,9
F0 Tremor (Hz)	4,0 ±2,6	2,6 ±1,2	2,8 ±1,3	3,3 ±1,3	2,7 ±1,3	2,1 ±0,8	2,8 ±1,5	2,5 ±1,8	1,8 ±0,7
Amp. Tremor (Hz)	2,5 ±1,3	2,1 ±1,2	2,4 ±1,5	2,6 ±1,0	3,0 ±1,3	2,2 ±1,1	2,6 ±1,4	2,4 ±1,2	4,9 ±3,8

Table 4. Mean values and standard deviations of the 10 acoustic measured parameters (Dr.Speech 3.2) grouped for the nine paradigms. The vertical grey shadings correspond to the frequency classes.

SVM accuracy for acoustic parameters			
	Intensity	Frequency	Frequency/Intensity
Accuracy (%)	85.64	93.45	83.73
STD (%)	6.14	8.25	8.60

Table 5. Overall accuracy of the acoustic SVM classification results.

4. Discussion

The endoscopic imaging of vocal fold vibrations is an essential part of clinical examination of voice disorders. Digital high-speed videolaryngoscopy is the state-of-the-art technology for investigation of asymmetric and irregular vocal fold vibrations (Doellinger, 2009). Similar to stroboscopy, high-speed videos are frequently evaluated by visual inspection relying on the experience of the investigator. There is still no objective or standardized procedure for describing the entire vibration patterns of vocal folds. Besides the description of vocal fold vibrations, the acoustic analysis of the voice signal gives valuable information for describing the severity of voice disorders. However, in most of the applied methods the acoustic properties and the laryngeal vibrations are separately examined. Thus, there is still little knowledge about the direct relation between the acoustic voice signal and the vibration pattern of vocal folds.

In this work, we presented a novel approach, called Phonovibrography, allowing an objective analysis of the visible vocal fold dynamics. Here, quantitative features are derived from PVG images which describe precisely the entire characteristics of vocal fold dynamics. For validation purpose Phonovibrography was applied to 108 high-speed sequences recorded from a single healthy female subject with normal voice. The female subject was instructed to produce 9 different phonatory tasks, i.e. phonation at different frequency and intensity combinations. A sequence length of one second time (> 150 glottal cycles) was chosen. The simultaneously recorded acoustic signals were analyzed using established voice quality measures (www.drspeech.com). Thus, besides evaluating the PVG analysis approach the effect of different phonation conditions on both the laryngeal vibrations and the acoustic voice signal could be studied.

Choosing just a single subject for validating the accuracy of the proposed PVG approach is mandatory as only within a healthy subject the phonatory tasks related changes of vocal fold vibration patterns can be interpreted in a correct way. For a single subject the extensive data acquisition comprising the recording of 108 repeated phonatory tasks is very time-consuming and potentially incriminating for the subject. Thus, collecting such a full data set from several subjects is difficult to achieve. As far as we know this examination presents the worldwide most detailed analysis of vocal fold vibrations within a single subject. Besides evaluating the performance of novel analysis approaches, the data set can further be used to investigate very precisely the fundamental principles of voice production in normal voice.

In the present study we applied methods from the field of machine learning towards recognition of different phonatory tasks within vocal fold dynamics as well as within the simultaneously recorded acoustic signals. Even though endoscopic and voice data represent different physical properties describing voice production (tissue vibrations vs. acoustic sound pressure), both modalities could be used to individually classify the nine different phonatory tasks within normal voice of one female.

4.1 Classification of vocal fold vibrations

The results given in Fig. 6 clearly show that a very high SVM classification accuracy (up to 96%) could be obtained using the new introduced PVG features. Even the classification of the nine class problem showed a very high performance of 85.5% which is in the same range as the results obtained using the acoustic measures, Tab. 5. It can be concluded from the results that the investigated frequency and intensity variations can be quantitatively traced back to alterations of the laryngeal dynamics. Furthermore, changes of vocal folds dynamics induce alterations of the acoustic signal as shown in Tabs. 4 and 5. To our knowledge, this is the first time that vocal fold vibrations could be quantitatively described so precisely during different phonation tasks and that the different phonatory task could automatically be classified at the vocal fold level.

The results obtained by the PVG parameters were further compared to symmetry/glottal parameters (Eqs. (13) and (22): [s], [g]) which are frequently used to describe vocal fold vibrations. Fig. 6 shows, that using the conventionally used glottal and symmetry parameters the performance of the classification is highly reduced. Using the feature vector [s] only a classification accuracy of approx. 15% for the nine class problem could be obtained. The glottal features [g] show a better performance with approx. 40% but are still far worse than the classification accuracy (94%) obtained using PVG parameters ω. The low classification results obtained by the glottal parameters show, that the reduction of the complex 2D vocal fold vibration pattern to a few parameters based on 1D glottal area

waveform signal is not sufficient for analyzing the laryngeal vibrations completely. Likewise, putting the focus only onto specific features as vocal fold symmetry (amplitude, phase, frequency) – which is frequently evaluated within the subjective assessment of stroboscopic or high speed movies - is not sufficient to fully describe vocal fold vibrations.
Having a closer look at PVG features at different oscillation states $n=\{30,60,90\}$, similar results were found for $n=30$ and $n=60$ state (Fig. 7). While the three class problems could still be classified with a high accuracy, for the nice class problem a classification accuracy of only approx. 60% was obtained. For $n=90$ the classification results show a similar behavior with a slightly reduced performance. However, when fusing all information obtained from the three oscillation states, the highest classification results were obtained. The increase of the performance documents that a precise analysis of vocal fold dynamics demands to describe the entire vibration pattern very comprehensively as it is done by PVG parameters which describe the temporal and spatial propagation of vocal fold vibrations.
Splitting up PVG parameters in different features ($[\omega_1]$: vibration type, $[\omega_2]$: deflection information, ω_3: symmetry, and ω_4: instabilities) further proves the benefit of including all extracted parameters together. Considering the parameter features separately (Fig. 8) the classification accuracy is reduced. Nevertheless, despite the feature reduction the classification accuracy using PVG parameters ω_1 - which comprises only information about the mean spatio-temporal vibration propagation of vocal folds - still shows a better performance than glottal [g] and symmetry [s] parameters together. Combining all features together results into highest classification accuracy of up to 96%. This again suggests the necessity of considering a combination of all features types as deflections, discrepancy, and instability.

4.2 Comparison of acoustics and vocal fold vibration classification

The highly consistent results obtained from acoustic and motion data show that within a subject vocal fold vibrations as well as the acoustic voice signal obtained from different trials can only be compared if they are recorded at similar intensity levels and similar fundamental frequencies. Recordings at significantly different intensity levels or frequencies will definitely cause different perturbations measures (e.g. Jitter, Shimmer, HNR, SNR, NNE) as well as changes within the laryngeal vibrations (Rovirosa et al., 2008). The results suggest that in clinical practice the repeated examination of a subject's voice needs to be performed at a comparable phonatory condition. Otherwise, the clinical value of measurements as objective and representative voice quality measures is highly limited.
In this work it could be shown that PVG analysis is a sufficiently sensitive approach to successfully identify even subtle changes in vocal fold vibratory characteristics induced by different phonatory tasks. As the sensitivity of the PVG approach could successfully be demonstrated, it can be used in ongoing studies to investigate vocal fold vibrations in presence of voice disorders. For studying pathologically induced alterations of vocal fold dynamics within a subject it must be considered that the examinations should be done under similar phonation conditions to exclude examination dependent influences.

5. Conclusion

Digital high-speed videolaryngoscopy is the state-of-the-art technology for investigating normal and pathological vocal fold vibrations. However, without adequate image analysis there is hardly an additional benefit comparing to the currently used stroboscopy technique

in sense of evidenced based medicine. The Phonovibrogram (PVG) has the potential to overcome the subjective or semi-automatic assessment of high-speed videos (Kunduk et al., 2010). Within this study it was proven that PVG image analysis has the necessary sensitivity to capture even minor alterations within vocal fold vibrations induced just by frequency and intensity variations. It was further shown that alterations of vocal fold vibrations are also detectable within acoustic perturbation measures. The high accordance between the results further proves that changes within the acoustic signal can directly be traced back to alterations of vocal fold vibrations. In respect to future clinical application, PVG analysis may be a useful tool to standardize the description of healthy and abnormal vocal fold vibrations. Objective Phonovibrography can directly be applied after examination and the obtained PVG images can easily be documented and stored on a hard-disc-drive using a lossless image data format which is essential for evidenced based medicine. An objective endoscopic image analysis tool, such as PVG, describing the vocal fold dynamics, could not only enhance voice assessment techniques but also help to objectively determine the outcome following an intervention in voice disorders (Voigt et al., 2010).

6. Acknowledgements

This work was supported by Louisiana State University, Faculty Research Grant 2008/2009, Deutsche Forschungsgemeinschaft (DFG) grant no. FOR894/1 and no. LO1413/2-1.

7. References

Beyer H.G. Schwefel H.P. (2002). Evolution strategies - a comprehensive introduction. *Natural Computing*, vol. 1, pp. 3-52.

Braunschweig T., Flaschke J., Schelhorn-Neise P., Doellinger M. (2008). High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonia. *Med Phys Eng*, vol. 30,no. 1, pp. 59-66.

Doellinger M., Hoppe U., Hettlich F., Lohscheller J., Schuberth S., Eysholdt U. (2002). Vibration parameter extraction from endoscopic image series of the vocal folds. *IEEE T Biomed Eng*, vol. 49, no. 8, pp. 773–781.

Doellinger M., Braunschweig T, Lohscheller J, Eysholdt U., Hoppe U. (2003). Normal voice production: computation of driving parameters from endoscopic digital high speed images. *Methods Inf Med*, vol. 42, no.3, pp. 271-276.

Doellinger M. (2009). The next Step in voice assessment: High-Speed digital endoscopy and objective evaluation. *Current Bioinformatics*, vol. 60, no. 2, pp. 101-111.

Doellinger M., Lohscheller J., McWhorter A., Kunduk M. (2009). Variability of Normal Vocal Fold Dynamics for Different Vocal Loading in One Healthy Subject Investigated by Phonovibrograms. *J Voice*, vol. 23, no. 2, pp. 175-181.

Deliyski D.D., Petrushev P.P., Bonilha H.S., Gerlach T., Martin-Harris B., Hillman R.E. (2008). Clinical Implementation of Laryngeal High-Speed Videoendoscopy: Challenges and Evolution. *Folia Phoniatr Logop*, vol. 60, no. 1, pp. 33-44.

Duchesne S., Caroli A., Geroldi C., Barillot C., Frisoni G.B., Collins D.L. (2008). MRI-based automated computer classification of probabile AD versus normal controls. *IEEE Trans Med Imaging*, vol. 27, no. 4, pp. 509-520.

Eysholdt U. & Lohscheller J. (2008). Phonovibrogram: vocal fold dynamics integrated within a single image. *HNO*, vol. 56, no. 12, pp. 1207-1212.

Hild K.E., Erdogmus D., Torkkola K., Principe J.C. (2006). Feature extraction using information-theoretic learning. *IEEE Trans Pattern Anal Mach Intell*, vol. 28, no. 9, pp. 1385-1392.

Hsu C.W., Chang C.C., Lin C.J. (2003). A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University.

Jiang J.J., Tangc S., Dalal M., Wu C.H, Hanson D.G. (1998). Integrated analyzer and classifier of glottographic signals. *IEEE Trans Rehabil Eng*, vol. 6, no. 2, pp. 227-234.

Karnell M.P. (1991). Laryngeal perturbation analysis: minimum length of analysis window. *J Speech Hear Res*, vol. 34, no. 4, pp. 544-548.

Kohavi R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *IJCAI*, pp. 1137-1145.

Kumar A. & Zhang D. (2006). Personal recognition using hand shape and texture. *IEEE Trans Image Process*, vol. 15, no. 8, pp. 2454-2461.

Kunduk M., Doellinger M., McWhorter A., Lohscheller J. (2010). Assessment of the variability of vocal fold dynamics with and between recordings with high-speed imaging and by Phonovibrogram. *Laryngoscope*, vol. 120, no. 5, 981-987.

Lin H. (2008). Identification of spinal deformity classification with total curvature analysis based on coded structured light. *IEEE Trans Biomed Eng*, vol. 55, no. 1, pp. 376-382.

Lohscheller J., Toy H., Rosanowski F., Eysholdt U., Doellinger M. (2007). Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med Image Anal*, vol. 11, no. 4, pp. 400-413.

Lohscheller J., Eysholdt U., Toy H., Doellinger M. (2008a). 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*, vol. 27, no. 3, pp. 300-309.

Lohscheller J., Doellinger M., McWhorter A., Kunduk M. (2008b). Quantitative analysis of vocal loading effects on vocal fold dynamics using Phonovibrograms. *Ann Otol Rhinol Laryngol*, vol. 117, no. 7, pp. 484-493.

Lohscheller J. & Eysholdt U. (2008). Phonovibrogram Visualization of Entire Vocal Fold Dynamics. *Laryngoscope*, vol. 118, no. 4, pp. 753-758.

Murphy P.J. (1999). Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *J Acoust Soc Am*, vol. 105, no. 5, pp. 2866-2881.

Neubauer J., Mergell P., Eysholdt U., Herzel H. (2001). Spatio-temporal analysis of irregular vocal fold oscillations: biphonation due to desynchronization of spatial modes. *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3179-3192.

Qiu Q., Schutte H.K., Gu L., Yu Q. (2003). An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr Logop*, vol. 55, no. 3, pp. 128–136.

Rovirosa A., Ascaso C., Abellana R., Martínez-Celdrán E., Ortega A., Velasco M., Bonet M., Herrero T., Arenas M., Biete A. (2008). Acoustic voice analysis in different phonetic contexts after larynx radiotherapy for T1 vocal cord carcinoma. *Clin Transl Oncol*, vol. 10, no. 3, pp. 168-174.

Ruben R.J. (2000). Redefining the survival of the fittest: Communication disorders in the 21st century. *Laryngoscope*, vol. 110, no. 6, pp. 241-245.

Schwarz R., Doellinger M., Wurzbacher T., Eysholdt U., Lohscheller J. (2008). Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *J Acoust Soc Am*, vol. 123, no. 5, pp. 2717-2732.

Selvan S. & Ramakrishnan S. (2007). SVD-based modeling for image texture classification using wavelet transformation. *IEEE Trans Image Process*, vol. 16, no. 11, pp. 2688-2696.

Titze, I.R. (2006). *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech, Iowa City, IA 52242, USA, ISBN 978-0-87414-156-6

Tokuda I., Horácek J., Svec J.G., Herzel H. (2007). Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *J Acoust Soc Am*, vol. 122, no. 1, pp. 519-531.

Vapnik V.N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., ISBN-10: 0387987800, New York, NY, USA.

Voigt D., Doellinger M., Braunschweig T., Yang A., Eysholdt U., Lohscheller J. (2010). Classification of functional voice disorders based on Phonovibrograms. *Artif Intell Med*, vol. 49, no. 1, 51-59.

Westphal L & Childers D. (1983). Representation of glottal shape data for signal processing. *IEEE Trans Acoust Speech*, vol. 31, pp. 766–769.

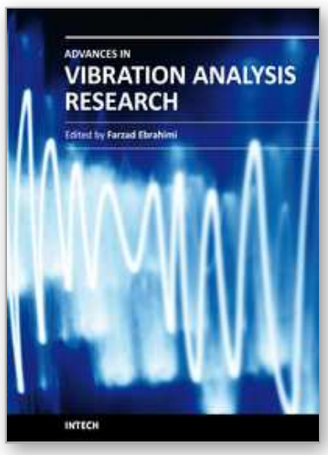
Wurzbacher T, Schwarz R., Doellinger M, Hoppe U., Eysholdt U., Lohscheller J. (2006). Model-based classification of non-stationary vocal fold vibrations. *J Acoust Soc Am*, vol. 120, no. 2, pp. 1012–1027.

Wurzbacher T., Doellinger M., Schwarz R., Hoppe U., Eysholdt U., Lohscheller J. (2008). Spatiotemporal classification of vocal fold dynamics by a multi mass model comprising time-dependent parameters. *J Acoust Soc Am*, vol. 123, no. 4, pp. 2324-2334.

Yan Y., Ahmad K., Kunduk M., Bless D. Analysis of vocal-fold vibrations from high-speed laryngeal images using a hilbert transform-based methodology. *J Voice*, vol. 19, no. 2, pp. 161–175.A.

Yang A., Lohscheller J., Berry D.A., Becker S., Eysholdt U., Voigt D., Döllinger M. (2010). Biomechanical Modeling of Human Vocal Fold Dyanmics by a 3D-Multi-Mass-Model. *J Acoust Soc Am*, vol.127, no. 2, pp. 1014-1031.

Zhang Y &. Jiang J.J. (2008). Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*, vol. 22, no. 1, pp. 1-9.



Advances in Vibration Analysis Research
Edited by Dr. Farzad Ebrahimi

ISBN 978-953-307-209-8
Hard cover, 456 pages
Publisher InTech
Published online 04, April, 2011
Published in print edition April, 2011

Vibrations are extremely important in all areas of human activities, for all sciences, technologies and industrial applications. Sometimes these Vibrations are useful but other times they are undesirable. In any case, understanding and analysis of vibrations are crucial. This book reports on the state of the art research and development findings on this very broad matter through 22 original and innovative research studies exhibiting various investigation directions. The present book is a result of contributions of experts from international scientific community working in different aspects of vibration analysis. The text is addressed not only to researchers, but also to professional engineers, students and other experts in a variety of disciplines, both academic and industrial seeking to gain a better understanding of what has been done in the field recently, and what kind of open problems are in this area.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen