# QUANTITATIVE CHARACTERIZATION OF FUNCTIONAL VOICE DISORDERS USING MOTION ANALYSIS OF HIGHSPEED VIDEO AND MODELING

*Thomas Wittenberg* *Patrick Mergell* *Monika Tigges* *Ulrich Eysholdt*

Department of Phoniatrics & Pedaudiology
ENT-Clinic at the University Erlangen-Nürnberg
Bohlenplatz 21, 91054 Erlangen, Germany
Wittenberg@phoni.med.uni.erlangen.de

## ABSTRACT

A semiautomatic motion analysis software is used to extract elongation-time diagrams (trajectories) of vocal fold vibrations from digital highspeed video sequences. By combining digital image processing with biomechanical modeling we extract characteristic parameters such as phonation onset time and pitch. A modified two-mass model of the vocal folds is employed in order to fit the main features of simulated time series to those of the extracted trajectories. Due to the variation of the model parameters, general conclusions can be made about laryngeal dysfunctions such as functional dysphonia. We show the first results of semi-automatic motion analysis in combination with model simulations as a step towards a computer aided diagnosis of voice disorders.

| | **Hyper-functional** | **Hypo-functional** |
|---|---|---|
| mean pitch | high | normal |
| phonation onset | pathological hard | normal, soft |
| sound | hoarse, creaky, pressed, soundless | soft soundless, breathy |
| amplitudes | restrained with intensity | widened with intensity |
| closure irregularity | relatively long period | relatively short amplitudes |

Table 1. Characteristics of hyper- and hypofunctional Dysphonias [8]

## 1. MOTIVATION

One important section of phoniatrics deals with the diagnosis and classification of functional disphonia. Morphological or organic voice disorders, such as cysts, polyps, carcinoma or granuloma can be detected and classified with the naked eye and the supplement of an endoscope or a laryngeal mirror. In contrast, functional dysphonia can only be diagnosed with the support of a highly refined imaging device, since no morphological change is observed. The dysfunction is hidden in the aperiodic, asymmetric and transitory oscillation patterns. For the recording of such a short-time scale motion without the violation of Shannons sampling theorem, a highspeed video camera has to be used [1, 3]. Fig.(1) shows exemplarily one period of vocal fold vibrations of a hyperfunctional dysphonia recorded with digital highspeed video camera. In our department of phoniatrics, a digital highspeed video camera has been applied in clinical routine examinations of vocal cord disorders in addition to the conventional videostroboscopy system for the past three years [2, 9].

One main goal of our research is to get a deeper insight into mechanisms hidden in functional dysphonia. Tab.(1) shows a comparison of the general features of of the two major categories of functional voice disorders, the hyper- and hypofunctional dysphonias [8]. In this context the prefixes *hyper* and *hypo* refer to the global tonus of all muscles involved in phonation, whereas *hyper/hypo* means, the muscle activity is too high, too low respectively. *Hypo*functional dysphonias are usually found in male patients whereas hyperfunctional dysphonias are specific for female patients. Even for the trained phoniatric it is very difficult to distinguish stridently between the hyper- and hypofunctions of the laryngeal muscles, since one extreme muscle tonus is usually compensated by the contrary tonus in an other muscle.

In the past, for the diagnosis of functional dysphonias, the phonation onset as one characteristic feature has been described qualitatively and subjectively with the terms *hard, normal* and *soft*. Typical for a soft phonation onset is an incomplete prephonatic closure after the adduction. Moreover, it is characterised by a relatively long onset time. In contrast, normal and hard, phonation onsets, are related to short onset times and a complete closure prior to the first oscillation maximum. The prephonatic closures are usually longer than in the case of soft onsets [7]. Fig.(2) shows three different kymograms of phonation onsets, which have been subjectively classified as hard, normal and soft, respectively.

It is our aim is to create a quantitative classification base to complement the subjective visual and auditive diagnosis of the phoniatric.

## 2. MOTION ANALYSIS

A semi-automatic motion analysis algorithm has been designed and implemented to extract the vocal cord trajectories from the highspeed image sequences [9].

This work has been funded by the 'Deutsche Forschungsgemeinschaft' (DFG)

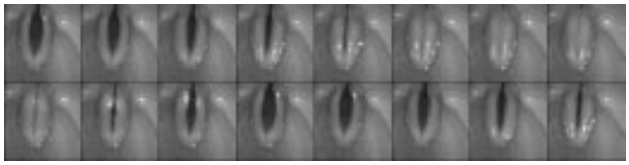

Figure 1. Example a of digital highspeed video of a Male subject, 32 years, with hyperfunctional dyphonia. One oscillation period of a vocal cord vibration. The recording speed was 1922 frames/s, with a resolution of 128x64 pixels x 8 bit grayscale.

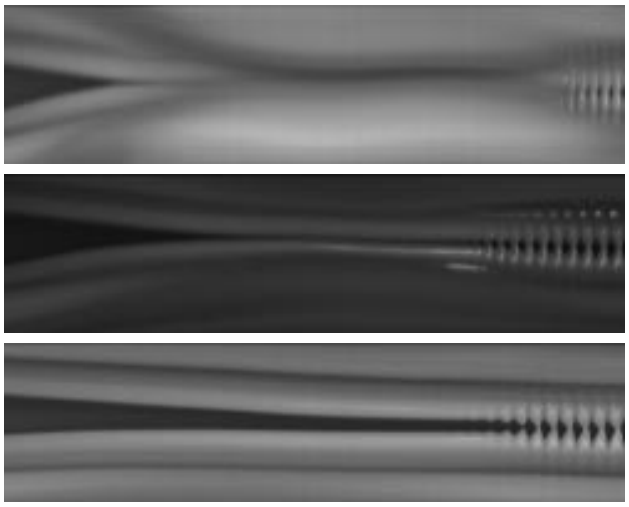

Figure 2. Different phonation onsets : hard phonation onset (top), normal phonation onset (center), soft phonation onset (bottom). From each image frame of the highspeed sequence (128x64 pixels x 8 bit grayscale, 1922 frames/s), one single image line (128 pixel) centered in anterior-posterior direction, and perpendicular to the principal glottal axis has been extracted. All image lines are arranged in chronolgical order, the time running from left to right. A time span of 0.25 seconds is shown.

In Fig.(1), one cycle of a vocal cord vibration is depicted. In the center of each single frame the two vocal folds can been seen. The dark area between the vocal folds is called glottis.
Using apriori knowledge about the anatomic design of the larynx and its physiological behaviour, the motion analysis problem can be broken down into two separate tasks:

- The detection and segmentation of the glottal area in each single frame of the sequence, and
- The calculation of selected trajectory motion points on the edge between the segmented area and the bordering vocal cords.

The first task is solved by applying a region growing algorithm to each single frame and thus separating the glottal area from the rest of the image. The seed for the region growing process can be calculated from the coordinates of the grayscale-minimum of each frame under the hypothesis, that the glottal area is always corresponding to the darkest region of the image and therefore contains the minimum grayscale value. To verify the spatial location of a potential seed, information about the location and expansion of the glottal area from the previous frames and a dynamic mean grayscale value are used. If the coordinates of the potential seed are outside the previous glottal area plus a tolerance region, or the value of the minimum grayscale is above the mean grayscale value of the past $n = 20$ frames, this seed will be rejected and a glottal closure is assumed.
The second task deals with the detection and calculation of significant tracking points on the edge of the vocal folds, which can be used to represent their motion during phonation and speech. In the past it has been useful to analyze three pairs of points on the vocal folds which correspond to the dorsal, central and ventral section of the larynx [6]. In our algorithm, these points can be detected from the glottal area calculated before. The first step consists of calculating the principle axis of the glottal area. This axis defines the temporal angular orientation and spatial expansion of the glottis. The principal axis is then divided by three orthogonal lines into four segments of equal length. Starting from the intersection points, these lines are traced until the border of the glottal area is reached. These endpoints mark the edge between the glottal area and the vocal folds as the temporal location of the individual tracking points. If these points are persued over all successive frames of an image sequence, an oscillating time series is generated. Analogous to the electro-glottogram (EGG) of the larynx, we call these trajectories **H**ighspeed-**G**lotto**G**rams (HGG).

Fig.(3) shows the resulting trajectories from the motion analysis of three different phonation onset modes of the kymograms in Fig.(2).

## 3. AUTOMATIC PARAMETER EXTRACTION: PHONATION ONSET TIME AND PITCH

In this section we present the automatic extraction of the phonation onset time $\tau$ from the HGG-data as one important parameter for the classification of functional dysphonias.
From the motion trajectories, as depicted in Fig.(3), the fundamental frequency $f_0$ can be obtained from its corresponding power spectrum. By using a bandpass filter ($f_0 \pm 10\%\ f_0$) the signal noise and motion analysis artifacts in the curve can be suppressed. From the filtered trajectory, the series of amplitude maxima is extracted using a simple peak-picking-algorithm and ignoring all peaks which are not spaced by the cycle duration. The resulting sequence of maxima describes the envelope curve of the phonation onset. The phonation onset time can be defined as the duration of amplitude growth from 32.2% to 67.8% of the saturation amplitude. These two thresholds have been chosen to be able to relate the experimentally obtained phonation onset time with a set of model parameters constituting a certain laryngeal configuration [4].

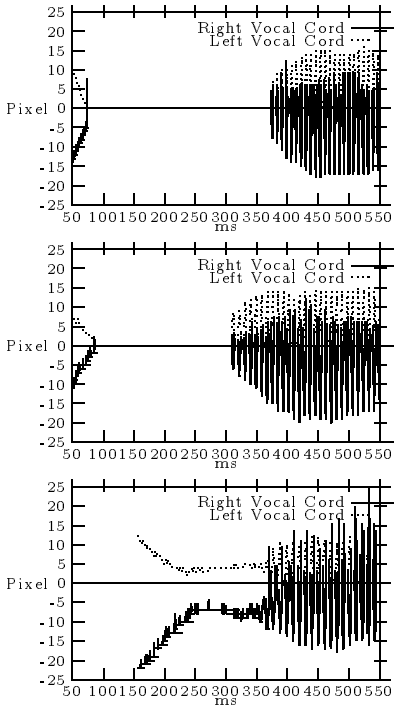


Figure 3. Highspeed glottograms of a subjective hard (top), normal (center) and soft (bottom) phonation onset. Relatively long prephonatic closure of the vocal folds during the hard onset (top), no prephonatic closure during the soft onset (bottom).

## 4. CONNECTION TO BIOMECHANICAL MODELS OF THE VOCAL FOLDS

The phonation onset can be considered in the framework of dynamical systems as a Hopf bifurcation, i.e. as a transition from damped to sustained oscillations due to parameter changes (adduction, subglottal pressure, tissue damping, etc. ). The analysis of the Hopf bifurcation has been performed using a simplified two-mass model. In this way, the functional dependencies between phonation onset time and the set of effective parameters, sketching the main features of the larynx, have been determined in the two-mass-model. In this context, it was possible to find a parametrization of the phonation onset envelope curve and thus perform excellent fits to the experimental data [4, 5]. In Fig.(4) we present three examples for the fits to the hard, normal and soft envelope curves, corresponding to the trajectories in Fig.(3). In this way we are able to compare directly experimental HGG-data with model simulations. Therefore, the nontrivial phonation mechanisms related to functional dysphonia become more and more transparent.

At first sight, we state that the quantitative analysis accords with the relation between the quality of the voice onset and the onset time established in section 1. Only the unique allocation of the analyzed phonation onset time corresponding to the subjective assessment 'normal' regarding the phonation onset time scale is difficult.

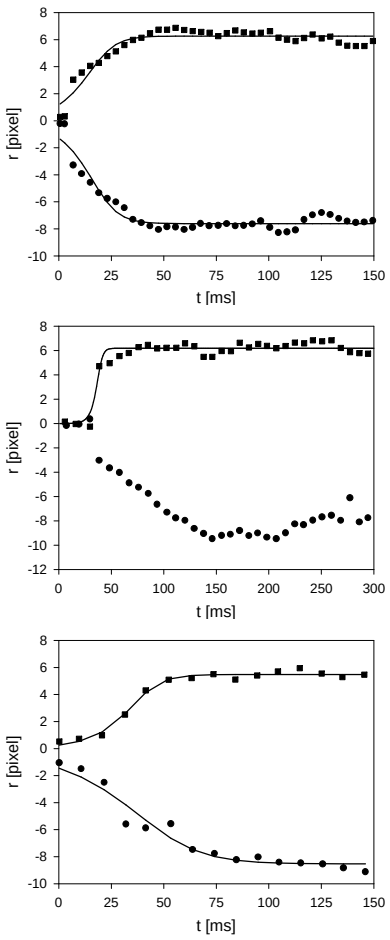


Figure 4. Envelope curves of the phonation onsets with different subjective assessments corresponding to the HGG's in Fig.(3): (top) *hard* : $\tau_L = (11.4 \pm 1.6)ms$. $f_0 = 231$ Hz. (center) *normal*: $\tau_L = (4.6 \pm 1.2)ms$. $f_0 = 113$ Hz. (bottom) *soft*: $\tau_L = (12.8 \pm 1.4)ms$, $\tau_R = (27.0 \pm 4.4)ms$. $f_0 = 101$ Hz.

## 5. RESULTS

In this section we present first results of quantitatively analyzed functional dysphonias, evaluated with the methods described above.

The vocal fold oscillations of 52 patients (16 male, 36 female) with functional dysphonias have been examined and recorded with our digital highspeed camera. All subjects where asked to produce the vowel /ɛ:/ at phonatory ease. From the resulting trajectories, the phonation onset time $\tau$ and the pitch $f_0$ have been calculated separately for the right and the left vocal cord. Fig.(5) shows a scatter diagram of the onset times versus the fundamental frequencies of the phonation of the right (top) and left (bottom) vocal fold oscillations. The data, based on a subjective visual and auditory medical diagnosis, can be grouped into three categories: 6 male patients with hyperfunctional dysphonia, 10 male patients with *hypo*functional dysphonia, and 36 female patients with hyperfunctional dysphonia.

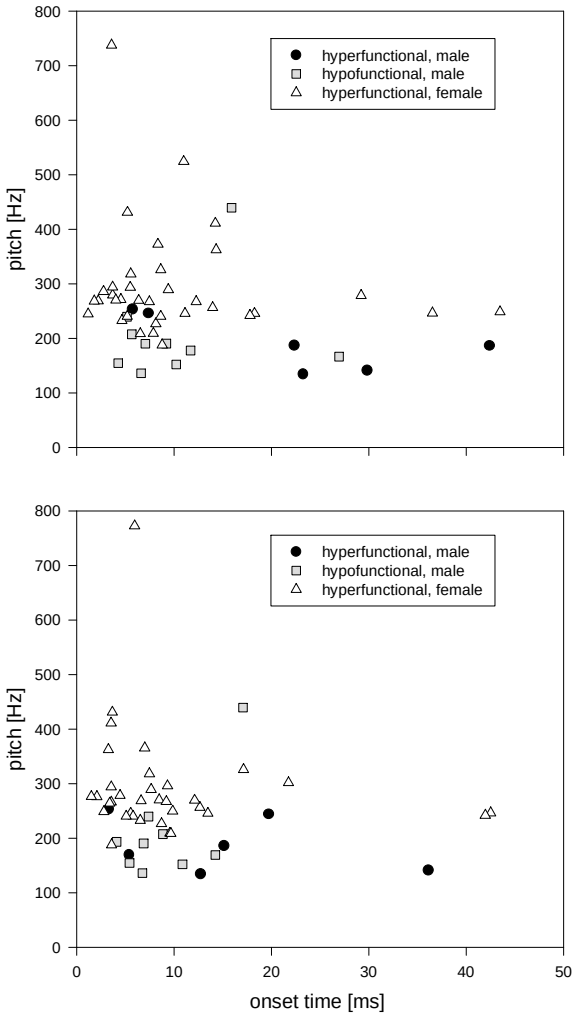


Figure 5. Scatter diagram showing the relationship between the phonation onset time and the phonation frequency.

As to be expected, the pitch of the female patients (240–320 Hz) is higher than the pitch of male group (100–200 Hz). The following statements can be made:

- The mean pitches (females: 280 Hz, males: 200 Hz) for both sexes are higher than the normal speech frequency ( females: 200 Hz, males: 125 Hz)
- The mean frequency of hyperfunctional males is higher (205 Hz) than the mean pitch of *hypo*functional males (190 Hz).
- The lower boundary for the onset time of the hyperfunctional females is at 1 ms with a clustering around 6 ms
- The lower boundary for the onset time of the *hypo*functional males is 4 ms with a clustering around 10 ms

The phontation onset time has an error range of 50–100% while the pitch error is in the order of 5%. At the present time, several types of error sources can be destinguished, originating from all stages of the data aquisition and reduction. These systematical errors are due to the nonstationarity of the laryngeal configuration during phonation, the finite temporal and spatial resolution (about 0.1 $mm$, 0.5 $ms$) of the highspeed camera, problems resulting from the endoscopic examination, present limitations of the motion analysis and finally the interpolation of the envelope curve from the maximum amplitudes.

## 6. CONCLUSION

The combined application of highspeed camera technique, digital motion analysis and model simulations is a very successful method to elude the complex mechanisms of phonation. It has been shown that the quantitative analysis yields similar results as the standard subjective assessments. Despite of the systematical errors, it is expected that, based on the clusters in the scatter plots, a distinction of functional dysphonia is possible. Prerequisite for a more detailed analysis of the clusters is the systematical evaluation of a larger data pool and the inclusion of subjective normal voices. We also expect a further distiction of functional dysphonias, if we compare the patients examination data before and after dedicated voice therapy and clinical treatment.

## REFERENCES

[1] DG Childers:*Laryngeal Pathology Detection.* Critical Reviews in Biomedical Engineering, (1977)

[2] U Eysholdt, M Tigges, T Wittenberg and U Pröschel: *Direct Evaluation of highspeed recordings of vocal fold vibrations.* Folia Phoniat., (1996)

[3] S Kiritani, H Hirose and H Imagawa: *High-speed digital image analysis of vocal fold vibration in diplophonia.* Speech communication 13, (1993)

[4] P Mergell, H Herzel, T Wittenberg, M Tigges, U Eysholdt : *Phonation Onset: High Speed Glottography and Modeling.* submitted to JASA.

[5] I Steinecke and H Herzel: *Bifurcations in an asymmetric vocal fold model.* J. Acoust. Soc. Am. 97, (1995)

[6] M Tanabe, K Kitajima, W Gould and A Lambiase: *Analysis of High-Speed Motion Pictures of the Vocal Folds.* Folia Phoniat. 27, (1975)

[7] M Tigges, T Wittenberg, U Pröschel, F Rosanowski and U Eysholdt: *Hochgeschwindigkeitsglottographie des Einschwingvorgangs bei verschiedenen Stimmeinsatzmoden*, Sprache – Stimme – Gehör 20, (1996)

[8] J Wendler and W Seidner: *Lehrbuch der Phoniatrie* VEB Georg Thime Leipzig (1987)

[9] T Wittenberg, M Moser, M Tigges and U Eysholdt: *Recording, processing and analysis of digital highspeed sequences in glottography.* Machine Vision and Applications, (1995)