**ABSTRACT:**

Introduction: Advances in computer image analysis have enabled the use of new functional imaging methods in the diagnosis of laryngeal diseases. Particularly interesting techniques of dynamic laryngeal imaging involve High Speed Videendoscopy (HSV). This still-developed technique allows to overcome the limitations of laryngovideostroboscopy (LVS) and a more detailed analysis of the glottal function based on the image of the actual vibrations of the vocal folds. It also enables the determination of objective coefficients parameterizing phonatory vibrations of the vocal folds.

Aim: The aim of this pilot study was to evaluate the use of a high-speed videendoscopy set with laser illumination for the diagnosis of glottic pathology in ENT practice.

Material and methods: The study included 40 patients who underwent LVS followed by HSV. The modern HSV examination kit—Advanced Larynx Imager System (ALIS), used for the first time in a clinical setting in Poland, is characterized by significantly improved, compared to the previously used high-speed cameras, operational parameters—a light head, the possibility of continuous lighting operation without excessive heating of the head tip, registration of the image in full color scale. Thanks to such modernization, the safety and course of the examination do not differ from laryngoscopy conducted with commonly used recorders. The device owes some of these improvements to a laser illuminator which was used for the first time as the main light source in a high-speed camera. In the study, two cases were selected to present the results of HSV and the analysis of the generated kymograms—a woman with no glottic pathology and a man with a polyp of the right vocal fold. In the first case, the HSV examination compared with the LVS revealed a discrete glottis functional disorder in the form of a tendency to hyperphonation. The patient with an organic lesion had a clearly visible irregularity of vocal fold vibrations, which also allowed to trace mucosal wave disturbances related to its reflection from the pathological structure of the glottis and the formation of a return wave, both on the fold affected by the lesion and, to a lesser extent, contralaterally. The glottic dysfunctions observed in the studied patients were confirmed in the generated kymograms and the graphs of the glottal width waveform (GWW), as well as in the parameters calculated on their basis, assessing the frequency and amplitude of phonatory vibrations.

Conclusions: The use of high-speed videendoscopy allows for a much more accurate assessment of the phonatory function of the glottis than in laryngovideostroboscopy. The presented HSV system allows for obtaining high quality kinematic images of the larynx, color fidelity, and contrast. The use of this technology in laryngological practice enables precise structural and functional assessment of the glottis and detection of discrete phonation disorders that elude the techniques used so far.

**KEYWORDS:** High Speed Videendoscopy, kymography, laryngoscopy, phonation, video recording, vocal cords

**ABBREVIATIONS**

ALIS – Advanced Larynx Imager System
GAW – glottal area waveform
GWW – glottal width waveform
HSV – High Speed Videendoscopy
LVS – laryngovideostroboscopy

**INTRODUCTION**

The production of the human voice is a complex mechanism in which the sound waves generated in the glottis undergo changes in the upper respiratory tract, and more specifically in the supraglottic part of the vocal tract. The primary source of its production is vibration of the vocal folds. Assessment of their structure and kinematics remains a recognized technique in diagnostics and therapy [1–3]. Their phonatory function can be subjected to audio and video analysis which involve the assessment of, among others: regularity and amplitude of vibrations, mucosal wave, phase difference and glottal closure [4, 5]. Over the years, efforts have been made to obtain data that would be as free as possible from disturbances arising during the movement of a sound wave through the air spaces of the upper respiratory tract. Such a recording was obtained by direct visualization of the larynx, which resulted in the dynamic development of endoscopic imaging techniques.

At present, laryngovideostroboscopy (LVS) remains the most common method of recording vibratory movements of the vocal folds, but it comes with certain limitations due to the technical aspects of recording. In the case of this technique, by synchronizing the strobe light with the frequency of the patient’s voice, the examiner obtains...
Fig. 1. Advanced Larynx Imager System (ALIS) (Diagnova Technologies) high-speed video recording set, equipped with a light source using laser diodes (ALIS Lum-MF1) and a High-Speed camera (ALIS Cam HS-1) in schematic diagram (A) and photo (B).

Fig. 2. Comparison of images of the glottis for Subject 1, obtained by means of four different methods: (A) White Light Laryngovideoendoscopy; (B) Laryngovideostroboscopy (LVS); (C) HD ‘reconnaissance’ mode of high-speed camera; (D) High Speed Videoendoscopy (HSV).

For this reason, the interest of researchers who study kinematic imaging of the larynx is raised by the still developing High Speed Videoendoscopy (HSV). This method allows for obtaining a dynamic image of the vocal folds, which is used for further analysis. It is not, however, subject to the limitations of videostroboscopy, because the image is recorded at a speed of up to several thousand frames per second, which enables recording of the actual vibrations of the vocal folds in successive consecutive cycles [5–7]. This makes it impossible to assess irregular vocal fold vibrations during a stroboscopic examination, in which the obtained image of the glottis is blurred, like endoscopy with the use of a static light source [8, 9]. Moreover, the recorded sample of stable phonation must be long enough to obtain a useful image.

Such detailed analysis enables the detection of various disorders of vocal function, unnoticeable during examination performed by means of the techniques used so far – some deviations are revealed only after a careful analysis of several hundred glottal cycles [15].

The aim of the present pilot study was to assess the use of the high-speed camera in laryngovideoendoscopy by means of a novel set which employs laser lighting, prepared for clinical application.
Dynamic images of the larynx obtained from the HSV were compared to the recordings conducted by means of LVS.

**MATERIALS AND METHODS**

The study included 25 patients hospitalized at the Department of Otolaryngology, Head and Neck Oncology of the Medical University of Lodz, diagnosed due to a hypertrophic vocal fold lesion (20 polyps, 5 vocal fold cyst), and 15 patients hospitalized for another otorhinolaryngological disease with a normophonic voice and without vocal fold pathology.

In all patients, the larynx was visualized with LVS. Next, an examination was performed using a high-speed camera.

Approval for this study was granted by the Ethical Committee of the Medical University of Lodz (no. RNN/96/20/KE 08/04/2020) and all patients gave informed consent to participate in the study.

First, laryngeal endoscopy was performed using the rigid Olympus HD Laryngoscope 90 endoscope (WA96105A). For digital recording of the image we used Olympus Visera Elite OTV-S190 camera with a xenon lamp providing white light and Olympus CLL-S1 strobe lamp.

Then, standard parameters of the vocal fold vibrations were assessed: regularity and symmetry of vibrations, mucosal wave, and the shape of glottal closure. The recordings were then subjected to kymographic analysis using dedicated software (Diagnova Technologies). Each sample, after semi-automatic image stabilization, required long-term manual preparation by an ENT doctor, i.e., determination of the glottal area with corrections for individual frames of the recording as necessary. After such preparation, the software generated a kymographic cross-section, which was then subjected to digital brightness equalization — important due to fluctuations in the light intensity of the strobe. For the purposes of parametric analysis, the obtained image required the determination of the edges of the vocal folds, and due to the much lower contrast than in the HSV recording, it was necessary to manually correct the detection by the examining physician.

Next, the HSV image was recorded using the Advanced Larynx Imager System (ALIS) (Diagnova Technologies), equipped with a light source using laser diodes (ALIS Lum-MF1) and a High-Speed camera (ALIS Cam HS-1). Laryngeal endoscopy was performed with the use of a rigid, oval endoscope Fiegert – Endotech E12.4/7.2 with a light guide with a 4.8 mm thick fiber optic cable. The weight of the head - an endoscope with a camera and an optical fiber - was about 200 g. For the purposes of the study, images were recorded at a rate of 2400 frames per second.

Two image recording modes were applied, i.e., the HD so-called “reconnaissance mode”, used for the initial assessment of the larynx, pathology identification and image centering on the glottis, and then the High-Speed mode, in which only the central area of the sensor is captured, which allowed to eliminate geometric distortions. Recording resolution was respectively – 960 x 800 px and 480 x 400 px. The light source provided narrow-band illumination with a wavelength of 405, 520, and 638 nm, which was then
customary to quantify the cyclicality of vocal fold vibrations using the glottal area waveform (GAW) [16, 17]. This approach enables the imaging of changes in the phonatory vibrations of the glottis to an average degree. In the case of organic lesions of the larynx covering a particularly defined part of the vocal fold, it was decided to use a glottal width waveform (GWW) graph [5, 13]. After initial automatic image stabilization, the software generated a kymographic section for a given point along the glottis, in which the region containing its central axis was then selected.

In the next stage, based on the differences in image brightness, the software determined the location of the edges of vocal folds, and created a GWW graph that reflects instantaneous changes in glottal width at different time points [18].

Next, for a precise quantitative analysis, the graph was used to determine the parameters describing the regularity of vocal fold

transmitted through a fiber optic cable to the optics. Greater efficiency and easier transmission of the laser beam made it possible to minimize the loss of brightness and to significantly reduce heat emission. The intensity of light generated by the laser illuminator was higher in the registration of the High-Speed sequence than in the reconnaissance mode. This arrangement allowed for real-time viewing, while reducing the heating of the endoscope tip. Image sharpness was adjusted semi-automatically. Two lengths of High-Speed recordings were used – Quick HSV with a length of 200 frames and Normal HSV – 2000 frames, which at the recording speed of 2400 frames per second corresponded to 83 ms and 830 ms, respectively. The image obtained from the endoscope was processed automatically and included: noise removal from the camera sensor, white balance correction, histogram equalization, and gamma correction.

The obtained recordings were then subjected to kymographic analysis using the dedicated Diagnova Technologies software. It is

Fig. 4. Analysis of data generated on the basis of high-speed videoendoscopy for Subject 1: (A, B, C) HSV image of the glottis with the cross-section along the course of the vocal folds marked with a gray line, for which a videokymogram was obtained; the posterior, middle and anterior parts of the glottis respectively; (D, E, F) videokymograms generated at the appropriate locations: the posterior, middle, and anterior sections of the glottis respectively; (G) Glottal width waveform (GWW) plot obtained from the videokymogram for the middle part of the vocal folds (solid blue line) and the waveform of the signal used for the GWW approximation (dotted line).
vibrations in subsequent cycles, e.g., fundamental frequency \( F_0 \), frequency perturbation parameters – jitter group, and amplitude perturbation parameters – shimmer group [19, 20].

**RESULTS**

In all patients, the larynx was visualized using LVS and the strobe parameters were assessed. Next, a high-speed camera examination was performed, obtaining high-quality HSV images exceeding the quality of those obtained during LVS. High-speed recordings allowed for visualizing the vibrations of the vocal folds also in cases that were difficult to evaluate stroboscopically, e.g., the presence of a significant component of vestibular phonation.

The recordings provided grounds for generating kymograms, which enabled the determination of parameters characterizing phonatory vibrations.

For the presentation of the High-Speed Endoscopy system with a laser light source (ALIS Cam HS-1 – Diagnova Technologies), we selected laryngeal images of two patients: a 26-year-old non-smoking female with no voice disorders and no glottis pathology (Subject 1) and a 50-year-old non-smoking male reporting hoarseness for several months (Subject 2).

In Subject 1, laryngovideoendoscopy revealed no deviations in the structure and function of the larynx. The unchanged mucosa was visualized, with no morphological changes, the laryngeal structures showed full symmetry both in the respiratory phase and during phonation (Fig. 2A.).

Laryngovideostroboscopy also showed no significant deviations in the regularity of vibrations, mucosal wave, symmetry of vocal fold vibrations, and glottal closure. Only discharge was found in the front ¼ of the vocal folds (Fig. 2B.). The videostrobokymography included 4 glottal cycles and provided grounds for confirming the observations made during the analysis of the recording (Fig. 3.).

The HD reconnaissance mode image of the high-speed camera confirmed the previous observations, while deviations from the norm were found in the proper part of the HSV study. There was a tendency to an hourglass-shaped glottal closure and mucus accumulation around the anterior commissure and on the upper surface of the middle ¼ of the vocal folds at the site characteristic of vocal nodules. These features were indicative of a tendency towards hyperphonation (Fig. 2C., D.).

The kymographic sections obtained in the recording showed a slight glottal insufficiency in the posterior ¼ of the vocal folds, which was not visualized in the sections generated from LVS. This confirms the higher diagnostic sensitivity of HSV.

For the kymographic cross-section corresponding to the middle ¼ of the vocal folds, we generated a graph of GWW reflecting instantaneous changes in the glottal area, and the following parameters were determined: average fundamental frequency \( F_0 = 302.7 \) Hz, Jitter: 1.18%, and Shimmer: 1.90% (Fig. 4.).

Laryngovideostroboscopy performed in Subject 2, i.e., a 50-year-old non-smoker without comorbidities, reporting hoarseness for several months, revealed a hypertrophic lesion on the free edge of the right vocal fold in the form of a fibrous polyp on a wide base, no other deviations were found, and respiratory and phonatory mobility was preserved (Fig. 5A.).

Laryngovideostroboscopy revealed movement of the organic lesion to the upper surface of the glottis during phonation, almost symmetrical vibrations of the vocal folds in their posterior part and limitation of phonatory vibrations at the site of the polyp (Fig. 5B.). Almost complete glottal closure was observed. The performed strobokymographic analysis confirmed the previously found deviations, and also showed an extension of the closed phase (Fig. 6.).

The image obtained by means of high-speed HD reconnaissance mode remained consistent with previous observations. The proper part of HSV imaging made it possible to visualize the mucosal wave more accurately at the corresponding levels of the glottis: posterior ¼, medial and anterior ¼, of the glottis (Fig. 5C., D.).

The kymographic cross-section corresponding to the middle part of the glottis which manifested as the absence of a mucosal wave in the section of the right vocal fold corresponding to the location of the polyp. Kymographic cross-sections generated for the anterior and posterior parts of the glottis also revealed some irregularities in vocal fold vibrations, greater in the anterior ¼.

In the cross-section of the front part of the glottis length, ahead of the organic lesion, a disturbance of the mucosal wave was visualized, related to the effect of interference/reflection of the wave from the pathological structure and the evoked return wave. It was
more pronounced on the right side, to a lesser extent on the side contralateral to the polyp, with intensity varying during phonation. The observations from kymographic analysis were confirmed in quantitative computer analysis.

Based on the GWW – a graph that reflects instantaneous changes in the glottal area at different glottal length levels generated from the kymographic section corresponding to the polyp location, the average fundamental frequency $F_0 = 159.5$ Hz and the parameters of frequency and amplitude perturbation were determined, respectively; Jitter 1.07% and Shimmer 4.49%. The increased value of Shimmer, i.e., the relative change in the amplitude of phonautary vibrations, reflects the turbulent air flow through the glottis caused by the organic change in the middle ⅓ of the vocal fold (Fig. 7.).

**DISCUSSION**

The paper presents the first research in Poland carried out with the use of a complete, modern set of a high-speed camera with a laser illuminator, a light head, and optimized operating parameters. The use of previous devices was associated with significant problems related to the weight of the camera, reaching even up to 600g, which significantly hindered patient examination, necessary to obtain the appropriate amount of reliable material.

In addition, an obstacle reducing the clinical usefulness of the devices was that images were recorded in shades of gray (e.g., the Wolf [21] camera). Another problem was the need to significantly increase brightness when recording at high speed. Due to the limited possibilities of adjusting the brightness of classic light sources, in the first high-speed cameras the endoscope was inserted without previewing, and the low efficiency led to a significant heating of its tip, which meant the introduction of necessary breaks to cool the device, thus extending the time of examination [22]. In the described set, a modern camera allowed to enhance the image quality (resolution 480 x 400 px), with the total mass of the head reduced to about 200 g. The use of a laser illuminator allows for a smooth increase of the lighting power to full intensity, only for a short recording time with a speed of 2400 fr./s, which in turn does not cause a perceptible increase in the temperature of the endoscope to a greater degree than in classical videostroboscopy. The image is recorded in the full scale of colors with their faithful reproduction. The device allows to record the larynx without additional interruptions, apart from the necessary disinfection and application of local anesthesia.

It should be emphasized that examination with the use of the described high-speed camera is much less time-consuming compared to the older systems. According to the authors, with the Wolf HRES 5562 camera [23], 2s of high-speed recording of the larynx requires 8.88-minute-long reconstruction of the sequence by the evaluating physician, which, as a consequence, significantly prolongs the functional assessment of the larynx. For this reason, older types of cameras were used primarily in laboratory research. However, the High-Speed camera system (ALIS Cam HS-1) allows recording the sequences in two modes, the „Quick HSV” mode, lasting only 200 frames (~80 ms for 2400 fps), which allows viewing them within 10 seconds, and the „Normal HSVP” mode, which allows for long-term analysis; sequences cover 2000 frames and browsing takes about 100s. In the presented system, the preview of the laryngeal image is visible immediately (which was not available in older type of cameras), i.e., the examining physician can safely (under image control) insert the endoscope
faithful color reproduction allows for precise determination of the free edges of the vocal folds, i.e., the contour of the glottal gap, along with its changes during phonatory cycles. Moreover, the diagnostic advantage of using high-speed videoendoscopy over stroboscopy should be emphasized [26–28].

HSV makes it possible to detect aperiodic disturbances occurring constantly or periodically during phonation, map all phases of the recorded phonatory cycles exactly frame by frame and at a selected location of the glottis, as well as to evaluate discrete moments of transition between the glottal cycle phases [21, 29, 30]. The obtained kymographic sections contain up to several hundred consecutive phonatory cycles.

The obtained recordings have all the advantages of the best laryngovideoendoscopy techniques, i.e., high image quality and high resolution, allowing the assessment of laryngeal structures and possible functional changes [24, 25]. Appropriate contrast together with faithful color reproduction allows for precise determination of the free edges of the vocal folds, i.e., the contour of the glottal gap, along with its changes during phonatory cycles.
recorded based on a very short phonation (compared to several cycles in the LVS) and are also characterized by greater accuracy of imaging when it comes to the movements of the vocal fold edges.

In Subject 1, it was only thanks to the use of a high-speed camera that an hour-glass-shaped glottal closure was noticed, and the generated kymograms revealed periodic discrete irregularities in phonatory vibrations.

In Subject 2, the HSV recording and the kymograms generated on its basis revealed the absence of mucosal wave in the part of the vocal fold corresponding to the location of the polyp, i.e., in the middle ⅔ of the right vocal fold. Moreover, HSV enabled the detection of disturbances in the phonatory vibrations in the front ⅓ and the rear ⅓ of this fold, related to the presence of the organic lesion itself and the return wave which it caused, as well as phonatory disturbances of lower intensity on the contralateral side. The described disorders were confirmed in the quantitative assessment of phonatory vibrations in the form of amplitude perturbation disturbances – increased shimmer value. Recent literature data emphasizes the importance of obtaining parameters that evaluate vocal fold vibration disturbances (including frequency and amplitude perturbation derived parameters), not only from acoustic analysis but also from HSV images [31].

To sum up, the conducted pilot study confirms the possibility of using a high-speed camera not only for research purposes, but also in laryngological practice. The authors present the clinical application of the HSV system based on the use of a laser light source during endoscopy for high-speed recording. Such a solution has not been found in the available literature. Therefore, it seems that it has been used as the main light source for a high-speed videoendoscopic camera diagnostic kit for the first time.

Its high efficiency allowed for increasing the comfort of examination through continuous previewing, and also improved the safety of the patient.

The endoscope tip did not reach temperatures that would cause injury/burn to the examined person. An additional advantage is the ability to change the position of the optics from 90 degrees to 70 degrees during clinical tests, as well as to connect the system to selected fiberscope models.

The presented reports concern preliminary research, and the results should be interpreted with caution. Due to the small number of subjects included in the study, in the future the authors intend to conduct further studies covering larger population groups. It is planned to further optimize data acquisition from high-speed recording and analysis, and to add the function of automatic detection of the beginning and end of phonation, as well as short-term disturbances. In addition, the semi-automatic quantitative assessment will be expanded with the possibility of measuring all glottal cross-sections simultaneously and calculating other parameters characterizing phonatory vibrations.

CONCLUSIONS

1. The HSV technique, due to its higher temporal resolution, is a more accurate tool for assessing phonatory function than LVS. It enables a more precise assessment of the periodicity of vibrations of the vocal folds and its disturbances;

2. In the presented HSV system, it was possible to eliminate the typical technical limitations during the registration of the dynamic image of the larynx by means of high-speed endoscopy with the use of older generation devices. Thus, the presented set is ready for use in ENT practice;

3. The use of laser lighting allowed for obtaining kinematic images of high quality, color fidelity, and contrast. This technology allows for a precise structural and functional assessment of the larynx.

REFERENCES


