

A novel high-speed digital imaging system for assessing vocal fold vibration

Kenichi KANEKO¹, Takeshi WATANABE¹, Masato INOUE², Haruo TAKAHASHI¹

¹ Department of Otolaryngology-Head and Neck Surgery, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8501, Japan

² Department of Electrical Engineering and Bioscience, School of Advanced Science and Engineering, Waseda University, Tokyo, Japan

Observing vocal fold vibration is essential when investigating the causes of hoarseness of voice and determining the appropriate treatment. However, in Japan, it is difficult at present to obtain commercial high-speed digital imaging (HSDI) systems for this purpose because of a lack of marketing routes. Accordingly, we devised an HSDI system using a high-speed camera (monochrome HiSpec1). We chose this camera as it has been an established instrument in industrial and laboratory use; it also provided excellent light sensitivity, and used a 400-watt xenon light source for the acquisition of bright and clear images. The HSDI system enabled video capture of up to 4.4 s with our usual setting of 422 × 384 pixels at 3000 frames per second (fps). It can also capture videos in various settings such as 396 × 256 pixels at 4000 fps. In addition, we equipped the system with devices for concurrent audio waveform acquisition. Our novel laryngeal HSDI system could be a useful clinical tool for vocal fold examination in patients with voice problems when access to commercially available devices is limited.

ACTA MEDICA NAGASAKIENSIA 59: 37–40, 2014

Key words: high-speed digital imaging, videostroboscopy, vocal fold vibration, hoarseness of voice, voice disorders

Introduction

Observing vocal fold vibration is essential when investigating the causes of hoarseness of voice and determining the appropriate treatment. Although videostroboscopy is widely used to create the illusion of continuous, slow-motion vibration of the vocal folds, its clinical use is limited because it relies on periodic vocal fold vibration and a stable phonation frequency to activate the strobe light^{1,2}. Although high-speed digital imaging (HSDI) can visualize true vocal fold vibration regardless of its aperiodicity, HSDI systems for observing larynx commercially sold overseas are costly and difficult to obtain in Japan at present because of a lack of marketing routes.

We previously devised a HSDI system using a low-price consumer digital camera (EXILIM PRO EX-F1; Casio

Computer Co., Ltd., Japan)^{3,4}, which enabled us to observe vocal fold vibration at a resolution of 336 × 96 pixels and a rate of 1200 frames per second (fps). It was a standalone, easy-to-handle system that enabled video capture of up to 7 min 18 s per gigabyte (GB) of its internal memory (maximum 32 GB). We applied this system to clinical practice and examined vocal fold vibration in approximately 300 patients from October 2008 to August 2013. This revealed that the system was very useful for patients with low fundamental frequency (F0) and large lesions, such as Reinke's edema and relatively large cysts. However, the temporal and spatial resolution of this device was often insufficient for patients with high F0 and small lesions, such as vocal fold nodules. In addition, this system was unable to acquire concurrent audio waveforms of the patient's voice.

We have now assembled a new system with a high-speed

Address correspondence: Kenichi Kaneko, MD, PhD, Department of Otolaryngology-Head and Neck Surgery, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8501, Japan

Tel: +81-95-819-7349, Fax: +81-95-819-7352, Email: kkaneko@nagasaki-u.ac.jp

Received March 1, 2014; Accepted May 14, 2014

digital camera (HiSpec1; Fastec Imaging Corporation, California) that was originally designed for industrial and laboratory use, and equipped it with a microphone for concurrent audio waveform acquisition. Here we illustrate the specifications and clinical usefulness of our system.

Technique

Our HSDI system uses the following components (Figure 1): (1) the HiSpec1 high-speed digital camera; (2) a 50-mm lens and lens adapters; (3) a 70° rigid laryngeal endoscope (STF-1; Nagashima Medical Instruments Co., Ltd., Japan); (4) a 400-watt xenon constant light source (Titan 400; Sunoptic Technologies®, Florida); (5) a foot switch; (6) a hand-made trigger box; (7) a cardioid condenser lavalier microphone

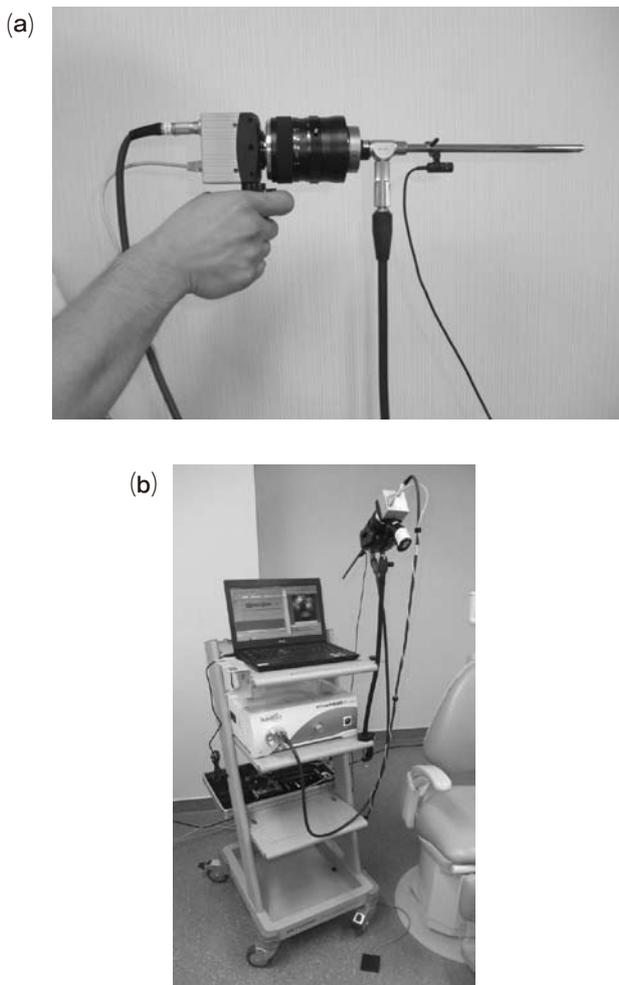


Figure 1. Photographs of our new HSDI system. (a) A high-speed digital camera, a 50-mm lens, lens adapters, a rigid endoscope, and a microphone. (b) An overall view.

(AT831b; Audio-Technica U.S., Inc., Ohio); (8) a mixer with microphone preamp and USB/Audio interface (Xenyx 302USB; Behringer GmbH, Germany); and (9) a personal computer. The institutional review board at Nagasaki University Hospital approved this system.

HiSpec1 is a digital camera with excellent light sensitivity (3200 ISO monochrome, 1600 ISO color), outstanding image quality (506 fps at a resolution of 1280×1024 and an adjustable frame rate up to 112,000 fps), and compact size [63 mm \times 63 mm \times 65 mm (height \times width \times depth), 280 g]. The camera has a 2-GB internal storage memory for video capture of up to 4.4 s with our usual setting of 422×384 pixels at 3000 fps. It can also capture videos in various settings such as 396×256 pixels at 4000 fps. We preferred a monochrome camera because it could provide higher contrast images than a color one.

The recording, playback, and preservation of images were controlled by the HiSpec Control Software (Fastec Imaging Corporation, California) when the camera was connected to the personal computer. The audio recording and playback were controlled by free software (Audacity®; Audacity Development Team, <http://audacity.sourceforge.net/>) using the same computer.

During video capture, the examiner records images in a circular recording mode, where the oldest image is overwritten by the newest until a trigger signal is input. The examiner also starts audio waveform acquisition through two channels; the left channel is for the patient's voice and the right is for a trigger signal.

The examiner transorally places the rigid endoscope into the larynx and the patient is instructed to phonate, at which point the examiner observes their vocal folds through the personal computer's monitor. When the examiner activates the foot switch, a trigger signal stops the image recording, and any images recorded for 4.4 s (422×384 pixels, 3000 fps) before the trigger are preserved. The trigger signal is recorded in a stereo audio file (44.1 kHz, 16 bit, WAVE file format) with waveforms from the microphone; using the trigger signal, we can synchronize the audio waveforms with the recorded images later. Figure 2 shows a diagram of the signals in our HSDI system.

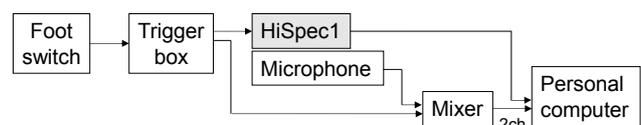


Figure 2. A diagram of signals in our HSDI system.

Figure 3 provides example images of vocal fold vibration in an 83-year-old male examined with this system. The patient presented with a severely breathy voice because of left vocal fold paralysis following an operation for a thoracic aortic aneurism. In this case, the vocal fold vibration showed a left–right asymmetry with respect to the amplitude of the vibration (increased amplitude on the left side) and a large glottic opening. The mucosal waves were also abnormally increased on the left side. During videostroboscopy, the vocal fold vibration could not be tracked because the pitch was not accurately detected.

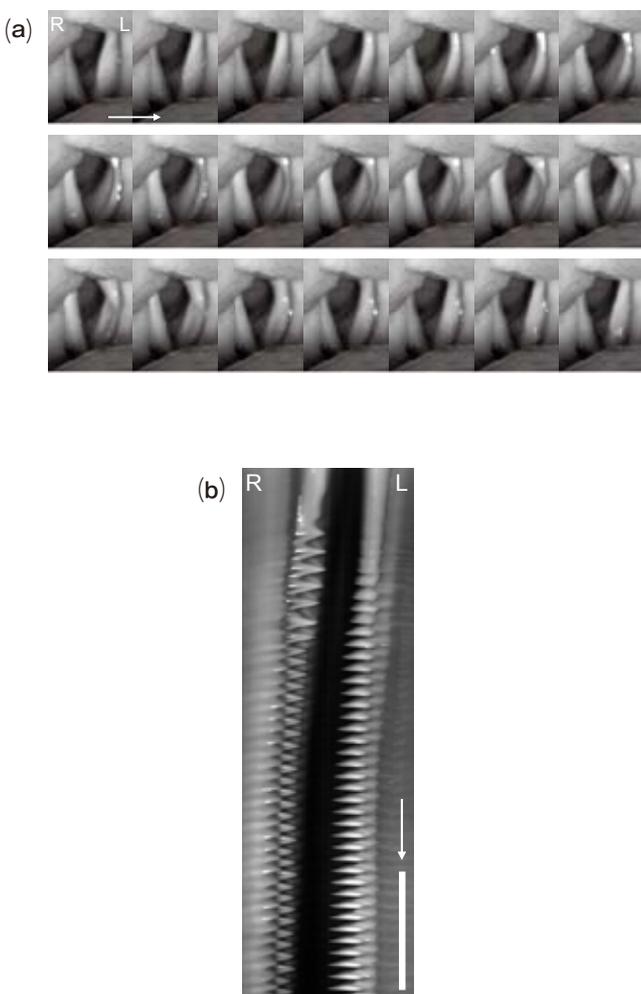


Figure 3. Vocal fold vibration in a patient with left vocal fold paralysis. (a) A sequence of images containing almost one glottal cycle shown in 21 images (7.0 ms). (b) Kymogram at vocal onset generated using our original software. White arrow shows time axis running downward. White bar represents 50 ms.

Discussion

Laryngeal HSDI systems that are commercially sold overseas, such as Model 9710 (KayPENTAX, New Jersey) and HreS Endocam 5562 (Richard Wolf GmbH, Germany), are specifically manufactured for the exclusive observation of vocal fold vibration. At present, in Japan, it is difficult to obtain these systems because of a lack of marketing routes. Therefore, we identified a need to design our own HSDI systems. Through trial and error, we have succeeded in designing suitable combinations of devices and settings for the clinical observation of vocal fold vibration.

When considering the requirements for our HSDI system, we desired a high-speed digital camera that was capable of capturing at least 3000 images per second, had sufficient spatial resolution and light sensitivity, and weighed as little as possible. The monochrome HiSpec1 camera met all these conditions and could be easily purchased through an agent in Japan, at relatively low cost when compared with systems sold overseas. Apart from the usual setting of 422×384 pixels at 3000 fps, it can capture videos in various settings such as 396×256 pixels at 4000 fps. It is on par with other systems in the resolution and the frame rate, i.e., 512×256 pixels at 4000 fps in Model 9710 and 256×256 pixels at 4000 fps in HreS Endocam 5562.

In general, HSDI needs a camera with good light sensitivity and a bright light source because of rapid shutter speeds; if not, the recorded images can be dark and unclear. Regarding our system, the monochrome HiSpec1 camera has excellent light sensitivity and the 400-watt xenon light source helps to capture bright and clear images. Monochrome cameras typically have higher sensor sensitivity and substantial spatial resolution, and smaller data-file sizes when compared to their color counterparts. Accordingly, we consider the monochrome camera to be favorable for high-speed laryngeal imaging, especially for edge detection. Obviously, it is limited in its ability to perform color analysis, such as the erythema of the vocal fold.

Because vocal fold vibration can be readily modified by phonation, it is desirable for a laryngeal HSDI system to acquire concurrent audio waveforms of a patient's voice. This allows assessment of the voice and vocal fold vibration simultaneously. Our new HSDI system employed concurrent audio waveform acquisition and synchronized the audio waveforms and image recordings later using the trigger signal as a marker.

Laryngeal HSDI enables us to assess the actual cycle-to-cycle variations of aperiodic vocal fold vibration, such as those in patients with moderate to severe disturbances of

voice quality and even those at voice onset and offset. It offers benefits over standard videostroboscopy in the analysis of aperiodic vocal fold motion, and may become an important adjunct to videostroboscopy in the evaluation of voice disorders^{1,2}. Our novel laryngeal HSDI system could be a useful clinical tool when access to commercially available devices is limited.

Conflict of interest

No competing financial interests exist.

References

1. Patel R, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Ann Otol Rhinol Laryngol* 117: 413-424, 2008
2. Kendall KA. High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. *Arch Otolaryngol Head Neck Surg* 135: 274-281, 2009
3. Kaneko K, Sakaguchi K, Tanaka F, et al. Vibration of the vocal folds observed using a high-speed movie system with consumer digital video camera (article in Japanese). *Practica Oto-Rhino-Laryngologica* 102: 379-384, 2009
4. Kaneko K, Sakaguchi K, Inoue M, Takahashi H. Low-cost high-speed imaging system for observing vocal fold vibration in voice disorders. *ORL J Otorhinolaryngol Relat Spec* 74: 208-210, 2012