Smartphone-Based Real-time Assessment of Swallowing Ability From the Swallowing Sound

**DUSHYANTHA JAYATILAKE[1], (Member, IEEE), TOMOYUKI UENO[2], YOHEI TERAMOTO[2],
KEI NAKAI[3], KIKUE HIDAKA[3], SATOSHI AYUZAWA[4], KIYOSHI EGUCHI[3],
AKIRA MATSUMURA[3], AND KENJI SUZUKI[1], (Member, IEEE)**

(Selected Conference Paper)

[1]Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan
[2]Rehabilitation Department, University of Tsukuba Hospital, Tsukuba 305-8576, Japan
[3]Faculty of Medicine, University of Tsukuba, Tsukuba 305-8575, Japan
[4]Tsukuba University of Technology, Tsukuba 305-8521, Japan

CORRESPONDING AUTHOR: D. JAYATILAKE (dush@ai.iit.tsukuba.ac.jp)

This work was supported by the Ministry of Health, Labour and Welfare, Japan.

ABSTRACT Dysphagia can cause serious challenges to both physical and mental health. Aspiration due to dysphagia is a major health risk that could cause pneumonia and even death. The videofluoroscopic swallow study (VFSS), which is considered the gold standard for the diagnosis of dysphagia, is not widely available, expensive and causes exposure to radiation. The screening tests used for dysphagia need to be carried out by trained staff, and the evaluations are usually non-quantifiable. This paper investigates the development of the Swallowscope, a smartphone-based device and a feasible real-time swallowing sound-processing algorithm for the automatic screening, quantitative evaluation, and the visualisation of swallowing ability. The device can be used during activities of daily life with minimal intervention, making it potentially more capable of capturing aspirations and risky swallow patterns through the continuous monitoring. It also consists of a cloud-based system for the server-side analyzing and automatic sharing of the swallowing sound. The real-time algorithm we developed for the detection of dry and water swallows is based on a template matching approach. We analyzed the wavelet transformation-based spectral characteristics and the temporal characteristics of simultaneous synchronised VFSS and swallowing sound recordings of 25% barium mixed 3-ml water swallows of 70 subjects and the dry or saliva swallowing sound of 15 healthy subjects to establish the parameters of the template. With this algorithm, we achieved an overall detection accuracy of 79.3% (standard error: 4.2%) for the 92 water swallows; and a precision of 83.7% (range: 66.6%–100%) and a recall of 93.9% (range: 72.7%–100%) for the 71 episodes of dry swallows.

INDEX TERMS Dysphagia, Swallowing sound, screening, cervical auscultation, videofluoroscopy, mHealth, realtime monitoring, bedside monitoring.

I. INTRODUCTION

Deglutition or swallowing is a complex neuromuscular process that consists of both voluntary and reflexive actions of approximately 50 paired muscles [1]. The swallowing process is usually described under 3 phases: the oral phase, which is sometimes divided into 2 as the oral preparatory phase and the oral transit phase, the pharyngeal phase and the oesophageal phase. Dysphagia or difficulties in swallowing can result from a wide variety of functional or structural deficits in any of these phases, and the most common of these causes include strokes and the neuromuscular diseases such as the Parkinson's disease and Amyotrophic lateral sclerosis or ALS. Dysphagia could be both congenital and acquired, and more common with the elderly. Aspiration due to dysphagia, which is the misdirection of oropharyngeal or gastric contents into the larynx instead of the stomach, is a major cause of pneumonia, in particular with the immuno-compromised patients. In Japan, pneumonia is the 3rd leading

cause of deaths in 2011 with 98.9 deaths/100000 people [2], and with the ageing population, this rate is expected to increase even further.

The videofluoroscopic swallow study (VFSS), which is commonly known as the modified barium swallowing examination, is considered the gold standard for the diagnosis of dysphagia. VFSS captures sequential video-radiographic images of barium contrast-impregnated food and liquid as they are transported during the oral cavity, pharyngeal cavity, and esophagus in realtime. VFSS needs to be performed by trained staff, and is subjected to exposure to radiation. Therefore, it is not suitable to evaluate fatigue effect on swallowing, and as barium is an unnatural food bolus patients often do not like the taste of it. Furthermore, the assessment of VFSS is more qualitative and requires expert knowledge to interpret the VF images. Considering those various complexities, some researchers such as Leder *et al.* [3] have suggested the use of screening techniques prior to the use of VFSS.

Dry swallowing, Repetitive saliva swallowing test (RSST), Water swallow test, Coloured water test, Cervical auscultation of swallowing and Swallowing provocation test (swallowing reflex test) are some of the screening test that can be performed without the need of any special tools [4]. Among the screening tests, the RSST, which we have considered during this study, is a simple test developed by Oguchi et. al for the screening for functional dysphagia and has a sensitivity of 98% for the aspiration [5]. RSST counts the number of repetitive dry swallows within a period of 30 s, and if the count is less than 3, then there is a strong possibility for the subject to have dysphagia. In RSST, the number of swallows is counted by the movement of laryngeal elevation using visual observation or by palpating, which are difficult to perform without proper training and experience.

During assisted feeding in elderly care facilities especially operating with limited staff, even basic information such as ''swallowed'' is important for smooth and safe feeding without aspirating. Furthermore, since aspiration is not likely to happen with every swallow, especially in patients with mild dysphagia, continuous monitoring during their usual eating and drinking environment has a better success in assessing the patient's swallowing ability [6]. As swallowing sound can be captured continuously and during activities of daily life (ADL) with minimal intervention, it is an ideal approach to monitor swallowing activities. However, as Hamlet *et al.* [7] have reported, the clinicians who use swallowing sound make judgements qualitatively, and therefore it is difficult to establish quantitative diagnostic approaches. Therefore, this study is carried out as a collaboration of the medical and the engineering faculties to develop a device-based swallowing ability monitoring system for bedside or daily life usage that is reliable and repeatable.

The swallowing sound is produced due to the movement of the bolus through the swallowing phases as well as the movements of various anatomical structures at those phases. The oral phase consists of voluntary processes and the pharyngeal and the oesophageal phases are under involuntary neuromuscular control, where the events are triggered by sensing the bolus inside the pharynx. The phases of swallowing are coordinated with each other through central pattern generators and peripheral reflexes, however, the timing characteristics are also altered by the volume of the oral intake [8], [9]. Cervical auscultation is the technique of using acoustic information related to swallows and airways typically captured with a stethoscope to estimate the swallowing function. Although highly subjective, it is a popular approach with significant agreement with VFSS [10], [11]. Takahashi et al. are among the first researchers to discuss instrumentation for cervical auscultation [12]. They found that the site over the lateral border of the trachea immediately inferior to the cricoid cartilage is the optimal site for detection of swallowing sound; and the center of the cricoid cartilage and the midpoint between the site over the center of the cricoid cartilage and the site immediately superior to the jugular notch also offer good signal to noise characteristics. Morinière et.al have compared the audio wave profile during swallowing with the corresponding VFSS images to understand origin of the sound components in the swallowing sound and there duration characteristics [13]. By analysing duration of the swallowing sound, peak intensity of the acoustic signal recorded, the peak frequency of the acoustic signal, and the breath before and after the swallow, Santamato *et al.* [14] have reported that the mean duration of swallowing sounds for a liquid bolus of 10 ml water differs significantly between healthy subjects and dysphagia patients. In healthy volunteers they observed swallowing sound duration between 276.0 ms and 714.0 ms for 10 ml water with an average of 438.1 ms. They also reported that the mean duration of swallowing sounds for a liquid bolus of 10 ml water is significantly different for healthy subjects and patients with dysphagia due to neurological diseases. By considering anteriorposterior and superiorinferior movement of the hyolaryngeal structure during swallowing, Steele *et al.* [15] and Sejdic *et al.* [16] have proposed a dual-axis accelerometer system to classify healthy and dysphagic subjects, and achieved a classification accuracy over 90%. There approach is heavy on processing and designed as a medical screening test, which yield a binary pass-fail result and not quantitative results that can be evaluated subjectively by experienced medical practitioners. Using the post-swallow breath sound Shirazi *et al.* [17] achieved 91% for the sensitivity and 85% for the specificity for detecting aspiration. However, there classification is limited to aspiration only. Lee *et al.* [18] investigating a different approach, the Doppler radar for non-contact monitoring of breathing patterns, successfully detected normal breathing, sleep apnoea, and number of other problematic breathing patterns. However, occlusion is a problem in there system, and suggested that a light cervical auscultation-based system, which is contact based but much more free compared to traditional chest straps or spirometers, could provide a useful alternative.

Since the human threshold of audibility decreases with the decreasing frequency, it is difficult to qualitatively evaluate

below 1 kHz. As a result, it is more useful to have instrument-based methods to automatically and quantitatively analyse swallow sound characteristics. The Swallowscope, which is illustrated in the Fig. 1, is the wearable device that we introduced earlier for the convenient monitoring of swallowing activities at bedside or during activities of daily life [19]. It is based on *m*Health and uses the swallowing sound to evaluate swallowing ability in realtime. It mainly consists of a neck-worn system that is made up of a contact microphone and an LED-based visualisation unit, an audio amplifier and an Android-based smartphone for capturing, processing and visualising. The Swallowscope provides quantitative measures: duration of each swallow, average duration of a swallow and the total number of swallows during monitoring periods as well as the qualitative measure of safe or not. A medical examiner could use the screen of the smartphone for detailed description of swallowing activities, and a caregiver or the family members could use the visualizations at the neck-wear for a simpler realtime assessment of each swallow such as whether the swallowing happened normally or whether there has been signs of aspiration. By using a smartphone for processing the data, we wanted to make the technology widely available and more affordable. Furthermore, by automatically uploading/backing-up the swallowing related raw-data to a remote server, it is possible to re-evaluate the whole swallowing process as well as perform more detailed offline analysis at a latter time by using more demanding algorithms.
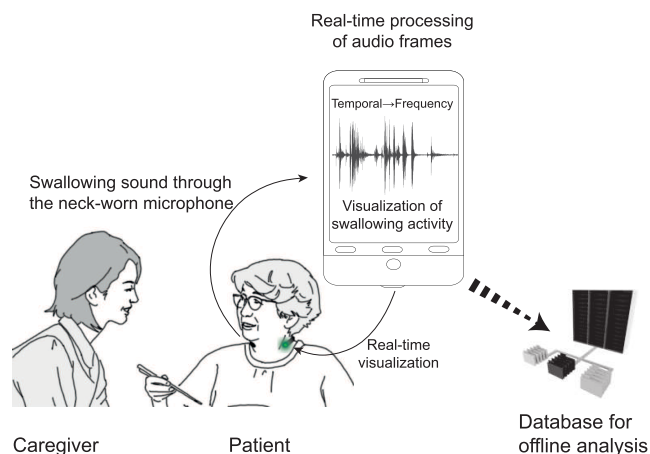

FIGURE 1. The Swallowscope for the continuous monitoring of swallowing activities at the bedside and during activities of daily life.

Earlier we reported the specific features of sound produced during swallowing [20]. This paper describes in details the development of a feasible realtime algorithm and evaluates the Swallowscope by using it to detect 3 ml water swallows and in RSST-based screening.

II. MATERIALS AND METHODOLOGY
The development of the proposed algorithm consisted of 3 phases: identifying key-characteristics of the swallowing sound, determining appropriate ranges for those

key-characteristics, and development of the final algorithm. We analysed swallowing sound of healthy subjects as well as the swallowing sound synchronised with lateral videofluoroscopic images of dysphagic subjects to develop the realtime swallow recognition algorithm.

A. SUBJECTS
We recruited 2 groups of subjects: a first group of 15 healthy volunteers and a second group of 70 subjects reported for videofluoroscopic examinations at the University of Tsukuba Hospital for the evaluation for dysphagia. This 70 consisted of all the subjects reported for the evaluation for dysphagia from April 2013 to September 2014 without any exclusions. The medical conditions of the subjects reported for the evaluation included head & neck cancer, cerebral hemorrhage, stroke, ALS, Guillain-Barre syndrome, myasthenia gravis, progressive supranuclear palsy and spinocerebellar degeneration. They were all evaluated by speech therapists (speech-language pathologist) before participating in the study, and depending on their evaluation on the severity, the subjects were administered with boluses of different consistencies: thin liquids, thick liquids and solids. We obtained informed consent from the subjects or from their guardians when appropriate. The study is approved by the institutional review board of the University of Tsukuba Hospital, H24-050IRB.

B. EXPERIMENTAL SETUP
VFSS-based experiments were conducted at the VF facility of the University of Tsukuba Hospital using an imaging equipment made by the Toshiba Corporation, Japan. As shown in Fig. 2, subjects were seated on a video imaging chair with the inclination set to 90 degrees. A one-yen coin, which has a diameter of 20 mm and visible in the VF images was affixed on the chin to determine the size of the bolus. Liquid boluses were administered by a speech therapist through a syringe and an SpO_2 oxygen saturation monitor was used to determine breathing efficacy and possible apnoea conditions. We captured the swallowing sound by using a skin-contact type throat microphone (Voice Touch Co. Ltd., Taiwan), which has
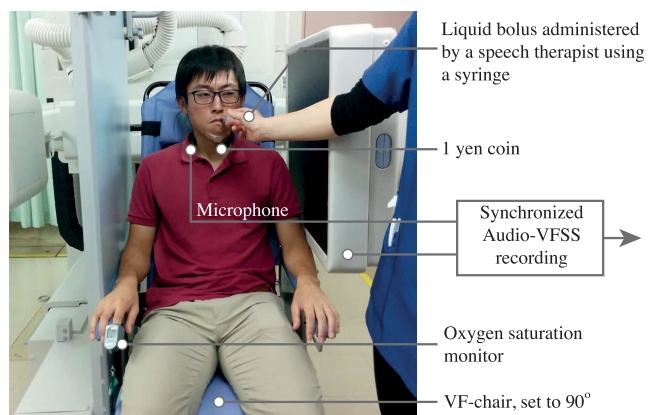

FIGURE 2. Setup for simultaneous VFSS and swallow sound capturing.

a sensitivity of 40 db – 45 db and a frequency range of 200 Hz – 3 kHz, and placing it around the neck close to the C5 level of the cervical region of the spinal cord as shown in Fig. 3. Swallowing sound and the VF images were recorded using a Canon ivis HF R31 video camera (Canon Inc., Japan). Videos were recorded using MPEG-4 AVC/H.264 and the audio was captured at 44.1 kHz and later down-sampled to 11025 Hz before analysis, in order to be compatible with our decision to use 11025 Hz for audio capturing with the smartphone.

FIGURE 3. Microphone placement at the C5 level. Mild pressure from the spring characteristics of the neckband creates good contact between the skin and the microphone. The VF image shows a lateral view.

C. PRELIMINARY ANALYSIS OF SWALLOWING SOUND

By carefully synchronizing lateral VF images with the audio we analysed the event based temporal and frequency characteristics of the audio waveform. In order to synchronize VF images with audio, we triggered an audio signal within the VF environment and time-wise aligned the VF images with audio waveform, eliminating any delays associated with the VF and audio data capturing systems. Although, it is difficult to establish universal timing parameters for a swallow, for the purpose for automatic recognition it is possible to define common ranges for healthy swallows. Fig. 4 shows a comparison of audio waveform with VF images (the data, which was taken from a dysphagia patient is fairly close to
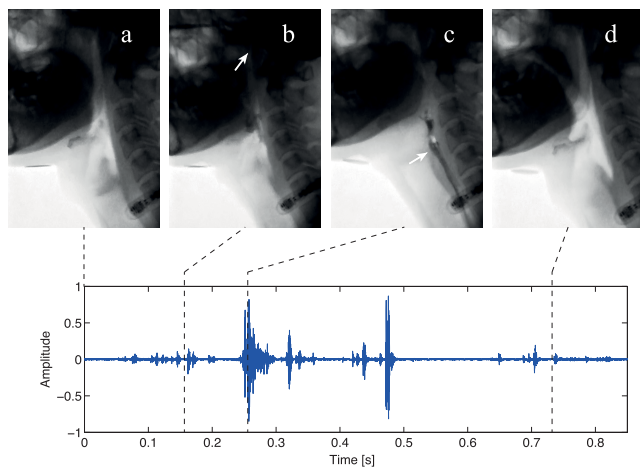

FIGURE 4. Comparison of the swallowing sound with the corresponding videofluoroscopic images. It is possible to couple the swallowing sound with a chronological series of anatomical events. (a) Resting (b) Larynx and soft palate ascend while the bolus entering the pharynx. Nasal passage is closed (c) Movement of the bolus through the UES. (d) Larynx return.

a healthy-swallow audio pattern). Using the software AudioDirector 4 (CyberLink Corp., Taiwan), a speech therapist marked the onset of swallowing events through frame-by-frame visual inspection of the lateral VF images obtained at a rate of 30 fps. As it can be seen from Fig. 4, the sound components produced during swallowing are coupled to a chronological series of anatomical activities: (a) resting state, (b) movement of the soft palate and the larynx while bolus moving into the pharynx, (c) movement of the bolus into the oesophagus through the upper esophageal sphincter (UES), and (d) returning of the larynx back to the resting position. Acoustic event related to (c) was the most prominent and present in 100% of the healthy samples we examined (the 8 healthy subjects were from the second group who were cleared of dysphagia after VFSS). However, events (b) and/or (d) were not observed in some of the samples. When labelling swallowing events on VFSS by the speech therapist, it was done by visually inspecting 30 fps videos frame-by-frame, and it was difficult to determine the exact occurrence of some events such as start of Hyoid movement or start and end of UES transition. As a result there could be ±1 frame error in VFSS labels.

Depending on the underlying cause for dysphagia changes, we noticed significant variations from the swallow timing parameters proposed by Morinière *et al.* in [13]. Furthermore, even for the same subject, depending on the readiness for the command swallow and breath status, etc. we noticed further changes in these timing parameters. We also noticed that cough or some vocals such as /a/ /a/ /a/ can generate similar temporal acoustic patterns. Therefore, more parameters were required to improve robustness of the swallow-detecting algorithm.

Figure. 5 illustrates audio waveforms and their spectrograms of 4 example cases; a normal swallow, cough, heavy breathing and throat clearing. A spectrogram, similar to a spectrum, describes the sound signal in terms of energy spread over its frequency components, but unlike a spectrum which provides only the stationary information, a spectrogram also displays the changes that occur over time. Hence it is used commonly to represent frequency and temporal characteristics of audio signals simultaneously [21]. Power of each frequency region, normalised to one in this case, is represented by colours described by the colour-map at the top of the Fig. 5. For the spectrograms, we obtained audio as 16-bit data at a rate of 11025 Hz and performed Short-Time Fourier Transformation on windows of 512 samples, with 50% overlapping. As it can be seen from the figures, cough and throat clearing showed stronger frequency components around 500 Hz. Furthermore, we found that the zero crossing rate to be the greatest for cough followed by swallow; and speech/vocals showed the least number of zero crossings.

D. SELECTION OF PARAMETERS FOR THE SWALLOW ACTIVITY RECOGNITION ALGORITHM

In this study, we tuned the realtime algorithm to achieve 2 purposes: to detect liquid swallows and to detect dry swallows
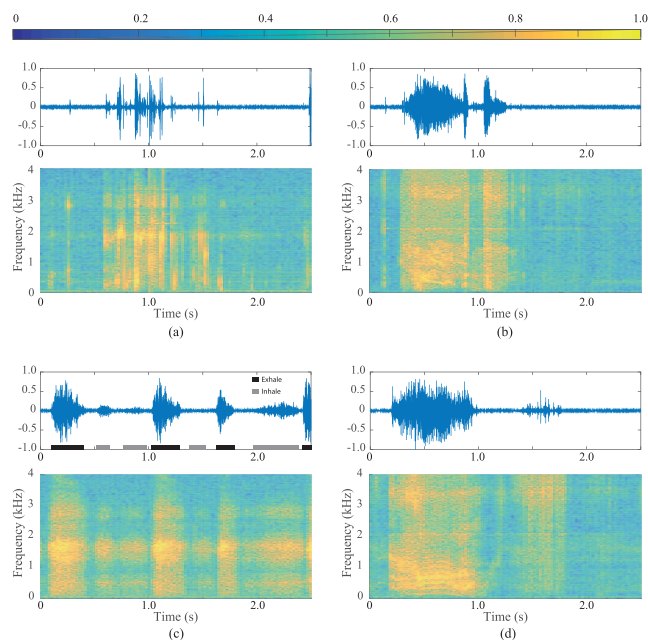
FIGURE 5. Audio and spectrogram characteristics of 4 different swallowing related events. (a) Swallow. (b) Cough. (c) Heavy breathing/gasping. (d) Clearing throat.
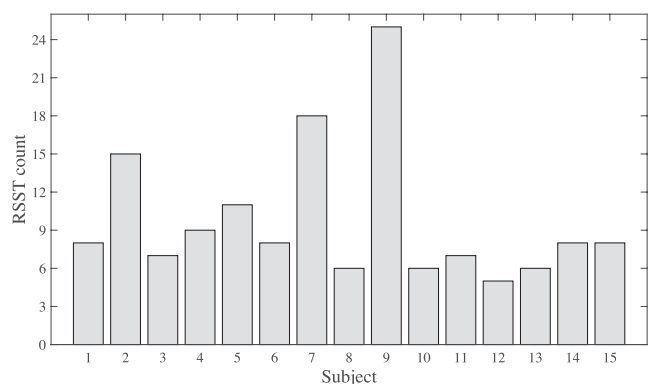

FIGURE 6. Variation of number of swallows during RSST of 15 healthy subjects. Usually the swallow count during 30 s period is less than 10.

during RSST. In order to find necessary parameters for the swallow activity detection algorithm while satisfying both liquid swallows and RSST, we first performed RSST on the 15 healthy subjects of the first group (Age 22 to 39). Since the subjects were healthy, tests were performed without VF, and in order to analyse data for feature extraction, we asked a speech therapist and an experienced swallowing sound coder, who were familiar with the VFSS and the swallowing sound, to label the starting and end points of each swallow during RSST through the auditory inspection of the swallowing sound. For the inspection of audio, we used the open source software Audacity 2.0.5 (Audacity Team). As it can be seen in Fig. 6, we observed significant variations between 5 and 25 (median = 8) for the number of swallows in RSST. Although 2 subjects managed 18 and 25 swallows each, based on the self evaluations of the subjects we determined it is very hard to perform more than 10 dry swallows within a period

of 30 s, which suggests on average 3 s rests between the swallows. Fig. 7 summarises the variation of average swallow lengths of the 15 subjects. For the 15 subjects, the overall average length of a swallow was 0.51 s (S.D = 0.16). From the swallow length data we noticed large but non-significant variations in lengths of swallow segments from person to person (p-value < 0.01), and smaller variations for the same person. Swallows during RSST consists of both swallowing of saliva as well as dry swallows or just the swallowing action without saliva, which is sometimes referred to as failed swallows. The extra effort sometimes required by some subjects to generate saliva made the oral preparation phase longer. These were among the contributing factors for occasional long and short swallow durations.
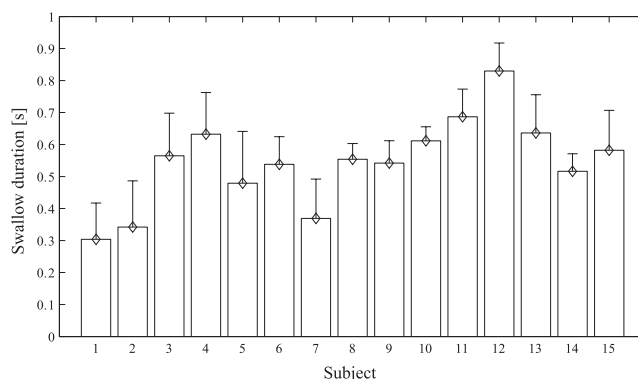

FIGURE 7. Variation of sound-based average length of a swallow during RSST of 15 healthy subjects.

E. REALTIME PROCESSING OF SWALLOWING SOUND

The proposed swallowing sound processing consists of 3 phases: preliminary detection of a swallow sound pattern, elimination of cough and vocalisations and the final detection. In the smartphone-based algorithm, which has been designed for the android platform, audio is captured as 16-bit pcm (Pulse-code modulation) at a rate of 11025 Hz and processed after acquiring into a 512-size buffer (46.4 ms).

At the beginning, the algorithm stays in a standby mode until it receives a buffer with non-silent audio. Here, an acquired audio data buffer is considered silent if the absolute maximum amplitude value is less than a predefined threshold. The threshold, which is typically set at 0.1 [range: 0.0 – 1.0], can be adjusted by the user during runtime to accommodate any noise present in the monitoring environment. Once a non-silent buffer is received, the algorithm changes to an acquiring mode and continues collecting buffers until it come across 4 continuous silent buffers (186 ms). As it can be seen from Fig. 4, the audio produced at different anatomical levels are generally separated by silent periods that are typically shorter than 3-buffers. Therefore the buffer count of 4 is selected to balance between the swallow detection lag and the false-negative rate. After that, the received audio segment, which is bounded by 4 or more buffers of silence, is rechecked to eliminate cough and vocalisations. For the cough and loud speech elimination, a continuous wavelet

transformation [22] is performed with the Complex Gaussian mother wavelet (gaus5) and the maximum absolute amplitude of the scale 19 is compared against a predefined threshold. Afterwards, the average zero-crossings count per buffer is evaluated to remove speech, where the detection threshold is typically set to 90 with the facility for the user to adjust during runtime. Finally, the audio segment is tested for: (a) total length, which should be between 5-buffers (232 ms) and 18-buffers (836 ms), and (b) the number of sections or the detected anatomical regions, which should be between 2 and 4 for water swallows and 1 and 4 for RSST. Here, we define sections as audio areas in the segment that are separated by shorter than 3 silent buffers. As illustrated in Fig. 8, if all the above conditions are satisfied, the audio segment is considered a proper swallow.

F. THE SWALLOWSCOPE

Fig. 9 shows the smartphone-based Swallowscope. The swallowing sound is captured from the neck-worn wired contact microphone and the realtime-processing outcome is displayed on the screen. The microphone is placed approximately around the c5 level; however, since the
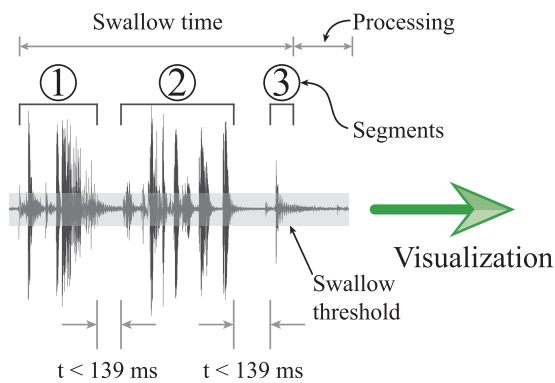

FIGURE 8. Realtime detection of swallowing sound. Segments are associated with different portions of the swallowing process.
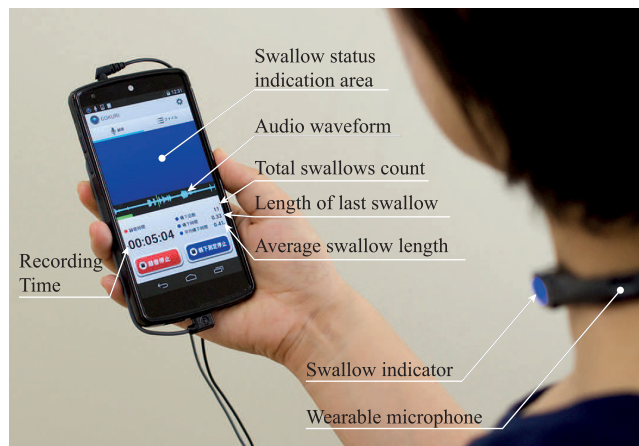

FIGURE 9. Overview of the Swallowscope, showing the indicator side of the neck-worn microphone and the display of the smartphone application.

features used for swallow recognition were less sensitive to the microphone location [20], it was usually worn by the subjects, themselves. The display statistics consists of the number of swallows recognised, the total time taken for the last recognised sallow and the average time taken for a swallow. The recognition outcome is shown both at the indication area of the smartphone and with the indicating full-colour-LED of the neck-worn microphone with 3 different colours: blue for no swallow or stand-by, green for a proper swallow and red for cough.

III. EVALUATION

A. AUTOMATIC RECOGNITION OF DRY SWALLOWS OF RSST

First we evaluated the automatic detection accuracy of the Swallowscope for RSST. We recruited the 8 subjects who were determined as healthy from the water swallow test, and the tests were performed with simultaneous lateral VF, which was used as the ground-truth to confirm swallows. VF images were evaluated by a speech therapist to label swallows and the automatic recognition was done by using an exact offline version of the realtime algorithm.

All together we counted 71 swallows and the Swallowscope recognised 69 of them with 2 false negative and 14 false positive. To understand the reliability of our detection we calculated the precision (fraction of retrieved instances that are relevant) and recall (fraction of relevant instances that are retrieved) of automatic recognition estimated according to the ground-truth obtained from VF evidence. Overall, the automatic swallow recognition algorithm achieved a precision of 83.7% (range: 66.6%–100%) and a recall of 93.9% (range: 72.7%–100%).

Fig. 10 shows the results of one of the RSST, reproduced with the exact offline version of the realtime algorithm.
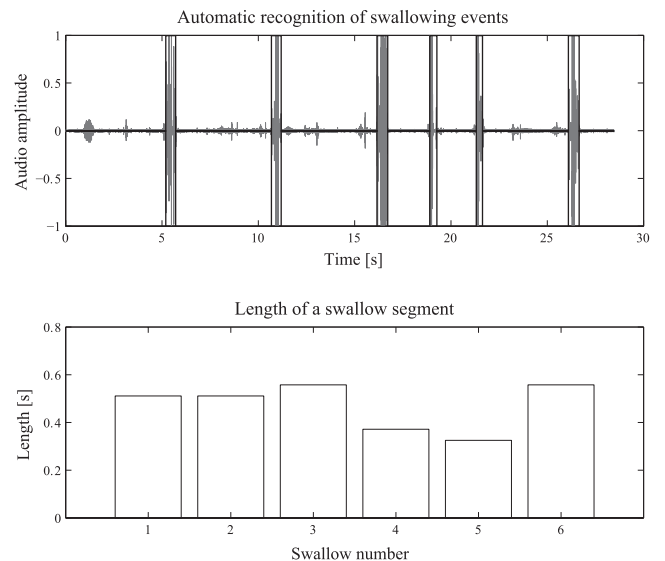

FIGURE 10. An example of realtime recognition of swallowing activities during RSST.

The top figure shows the swallow waveform and the bottom figure shows the variation of duration for each swallow. In this instance, the Swallowscope recognized all the swallowing events accurately. Our methodology enables the extraction of quantitative trend information, and the Fig. 11 compares the timing characteristics of swallows during RSST of 3 healthy subjects and 3 subjects with reduced swallowing ability. In general the swallows of the dysphagic subjects were longer and the number of swallows were much smaller.
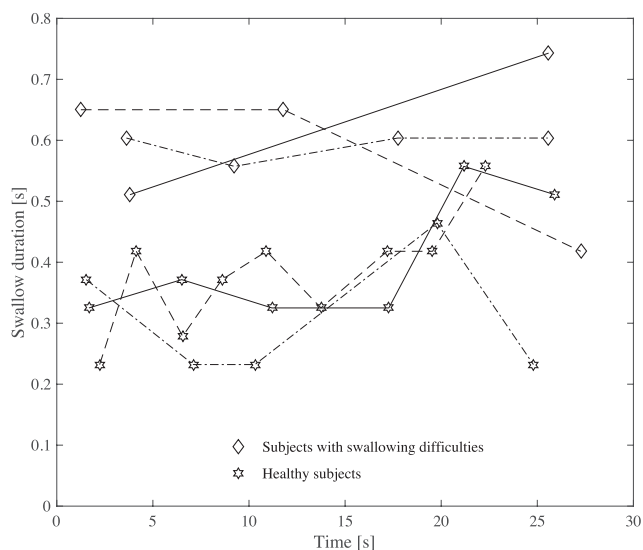

FIGURE 11. Comparison of RSST timing characteristics of 3 healthy subjects and 3 subjects with swallowing difficulties.

B. AUTOMATIC RECOGNITION OF WATER SWALLOWS OF DYSPHAGIA PATIENTS

We selected the simultaneous VF and swallow sounds of 31 subjects with dysphagia (23 males, mean age = 68.8). We only included the subjects who were tested for at least 2 counts of 3 ml 25% Barium mixed water swallows. Possible candidate were initially identified based on a prior evaluation by speech therapists on the likelihood to aspirate and the ability to test with 3 ml water swallows. For each command swallow, with the aid of lateral VFSS, a speech therapist labelled the start of soft palate elevation, start of Hyoid bone movement, start of the Hyoid bone lateral motion, start of UES transition, end of UES transition, start of Hyoid bone descend and Larynx returning to rest. Usually the start of Hyoid bone movement happens with the bolus entering pharynx, right after completing oral phase, and the end of UES transition indicates the completion of pharyngeal phase. We then used an offline reproduction of the realtime algorithm to detect swallows. As VF was active only during command swallows (to minimise exposure to radiation) we did not count false-positives as there was no evidence to confirm whether there has not been any saliva or oral residue swallows during those instances. Fig. 12 shows a summary of automatic detection of swallowing activities compared against ground truth determined using

• Hy: Start of Hyoid movement (VF)
• Swallow auto: Automatically detected swallowing event
• UES_(s): Start of UES transfer (VF)
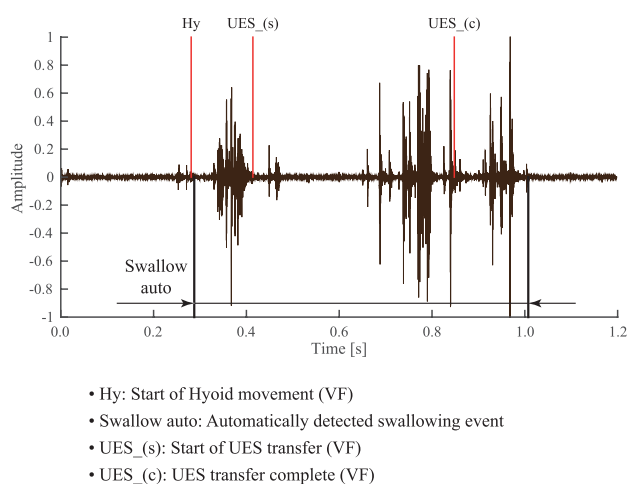• UES_(c): UES transfer complete (VF)

FIGURE 12. Automatic detection of swallowing activities compared against synchronised VF evidence.

VF for 2 of the swallows. Our algorithm automatically detected all or some of the swallowing events of all the 31 subjects, and the overall detection accuracy for the 92 swallowing episodes was 79.3% (S.E: 4.2%).

IV. DISCUSSION

As could be seen in Fig 10, we could analyse the progress of swallowing activities: whether the subject is slowing down by using the gap between swallows, or whether the time for a swallow is increasing from the swallow length information. These parameters could indicate whether the swallowing function of the subject is getting unusually tired. We consider the shortening in the swallows 4 and 5 could be attributed to mouth drying and not producing enough saliva.

Due to the quantitative nature of devise-based evaluation, it is possible to use the Swallowscope to obtain more information about the swallowing process. Figure 11 demonstrates this by comparing timing characteristics of swallows during RSST of 3 healthy subjects and 3 subjects with reduced swallowing ability. Compared to the subjects with reduced swallowing ability, the healthy subjects managed more swallows that were fairly well distributed over the 30 s interval. Furthermore, the swallows too were much shorter, indicating much stronger and quicker swallows.

In the current stage where we focused more on the RSST and command swallows of 3 ml of water, we wanted to increase the recall: the true positive rate, to make sure the Swallowscope's categorisation of not risky could be accepted with high confidence. Algorithm provided in this paper has 2 improvements over our previously proposed algorithm in [19]: (a) the number of sections in RSST and water use 2 different values because in RSST some sound components are not always audible due to dry nature of the swallows and (b) we introduced zero crossing detection for the better elimination of speech components. These enhancements have resulted in good detection rates for both water and dry swallows.

A. CLINICAL BENEFITS OF THE SWALLOWSCOPE

Although, cervical auscultation has been used by the clinical practitioners for diagnosing, it is subjective and as a result has been difficult to develop any standardised tests that are based on cervical auscultation. However, as many researchers have observed, rich vein of information contained in swallowing sound is capable of making useful decisions on swallowing ability and the swallowing function. The Swallowscope attempts to exploit these advantages to develop a tool that can be used in ADL.

Standard screening techniques used to assess swallowing ability have 2 main limitations: the examiner who conducts the screening test needs to be an expert to identify various outcomes of the test including correctly recognising swallows, and the outcome of the tests are subjected to qualitative interpretation of the examiner. Use of the Swallowscope on the other hand does not require any special knowledge nor training and the results it produce are quantitative. As a result, they are reproducible and comparable.

Ability to recognise swallows is useful for the development of quantitative methodologies to evaluate swallowing ability. The Swallowscope can assist long-duration monitoring such as during ADL or during all through night, which would otherwise require tedious efforts to localise swallowing activities. Furthermore, a recent study has found statistically significant correlations between the spontaneous swallowing frequency measured over a period of 30 minutes and dysphagia [23]. Despite the significance of this finding, it would be difficult to implement practically without a system to do it automatically, and the Swallowscope can be used easily to realise this. The Swallowscope also has the potential to enhance the RSST, not only by removing subjectivity and ambiguity, but also adding more parameters such as the slowing-down rate.

Due to the additives like barium or thickening agents used in the liquid, paste or solid food used for VFSS or video endoscopy-based testing, which is commonly referred to as VE, are often described by the patients as unpleasant. Furthermore, being surrounded by equipment and medical staff make them nervous, and likely result in different from normal swallowing patterns. Compared to VFSS or VE, Swallowscope-based realtime assessing can be conducted at familiar surroundings while having regular meals, so the assessment would reflect better on the actual circumstances.

One of the frameworks we are looking to investigate with the Swallowscope is the mealtime assistance of the elderly in nursing homes. The world population is ageing rapidly and the desires of governments to minimise medical expenses is shifting the focus from hospitalisations to home-based care, increasing the number of the elderly who does not have quick access to hospital-based care. The analysis of swallowing activities with the Swallowscope is realtime, and therefore the caregivers during assisted feeding can make sure that the patient has swallowed the food before feeding again. As it is visualised on the neck-worn unit it self, caregivers do not have to look away to check some monitors to confirm swallowing. That enables to have a natural face-to-face conversation while feeding, putting the patient at ease. Just as coughing indicates possible aspiration, ability to cough properly is important to make sure that the person can cough any aspirated food particles out. Healthcare workers would find that information useful when feeding someone who is in a risk of aspirating, and we could use the cough detection to determine whether their coughs are strong enough to serve this purpose. Firmness of foods, liquid density, and the eating posture etc. could have significant effects on the ability to swallow safely and the response of the elderly to such factors could change frequently. Therefore, the realtime assessment of the Swallowscope could also be a strong indicator on the appropriateness of the texture and the consistency of the food. The sharing of information through a cloud service could enable the development of electronics medical records or EMR system for the home-based patients and nursing-care homes. That way many different parties caring for the elderly would find it easy to share various types of critical information such as caring for meals or making decisions about their health. Furthermore, the home caregivers could easily obtain expert advice from the healthcare professionals. As a result the Swallowscope could provide the necessary support to assist caregivers during mealtime assistance, develop meal plans to minimize the risks of aspiration and bridge the related healthcare professionals together to provide better support and service.

V. CONCLUSIONS

Swallowscope is the smartphone-based wearable device we developed to continuously monitor swallowing activities and assess the swallowing ability without causing discomfort to the wearer. It can analyse swallowing sounds in realtime and generate quantitative results: the number of swallows and the swallowing duration, which can assist bedside screening, and share them through a cloud-based system. We achieved very good performances in terms of both the precision and recall. Our current realtime algorithm is capable of recognizing precise temporal characteristics of swallowing activities. It can accurately determine the starting point and the duration of each swallow. We further demonstrated how the quantitative measures can help to estimate swallowing function by focusing on various performance parameters such as the variation of swallow duration and the variation of interval between the swallows during RSST. These features could lead the way establish more precise parameters to evaluate the swallowing ability, produce more reliable screening methods and develop more customised mealtime assistance protocols.

REFERENCES

[1] B. Jones, Ed., *Normal and Abnormal Swallowing: Imaging in Diagnosis and Therapy*. New York, NY, USA: Springer-Verlag, 2003, ch. 2.

[2] (2011). ''Vital statistics in Japan—The latest trends,'' Dept. Statist. Inf., Ministry of Health, Labour and Welfare, Tokyo, Japan, Tech. Rep., 2011. [Online]. Available: http://www.mhlw.go.jp/english/database/report.html

[3] S. B. Leder, D. M. Suiter, H. L. Warner, L. M. Acton, and B. A. Swainson, "Success of recommending oral diets in acute stroke patients based on passing a 90-cc water swallow challenge protocol," *Topics Stroke Rehabil.*, vol. 19, no. 1, pp. 40–44, 2012.

[4] S. Horiguchi and Y. Suzuki, "Screening tests in evaluating swallowing function," *Jpn. Med. Assoc. J.*, vol. 54, no. 1, pp. 31–34, 2011.

[5] K. Oguchi, E. Saitoh, M. Mizuno, M. Baba, M. Okui, and M. Suzuki, "The repetitive saliva swallowing test (RSST) as a screening test of functional dysphagia (1) normal values of RSST," (in Japanese), *Jpn. J. Rehabil. Med.*, vol. 37, no. 6, pp. 375–382, 2000.

[6] B. Martin-Harris and B. Jones, "The videofluorographic swallowing study," *Phys. Med. Rehabil. Clin. North Amer.*, vol. 19, no. 4, pp. 769–785, 2008.

[7] S. Hamlet, D. G. Penney, and J. Formolo, "Stethoscope acoustics and cervical auscultation of swallowing," *Dysphagia*, vol. 9, no. 1, pp. 63–68, 1994.

[8] W. J. Dodds, E. T. Stewart, and J. A. Logemann, "Physiology and radiology of the normal oral and pharyngeal phases of swallowing," *Amer. J. Roentgenol.*, vol. 154, no. 5, pp. 953–963, May 1990.

[9] I. M. Lang, "Brain stem control of the phases of swallowing," *Dysphagia*, vol. 24, no. 3, pp. 333–348, 2009.

[10] P. M. Zenner, D. S. Losinski, and R. H. Mills, "Using cervical auscultation in the clinical dysphagia examination in long-term care," *Dysphagia*, vol. 10, no. 1, pp. 27–31, 1995.

[11] C. Borr, M. Hielscher-Fastabend, and A. Lücking, "Reliability and validity of cervical auscultation," *Dysphagia*, vol. 22, no. 3, pp. 225–234, 2007.

[12] K. Takahashi, M. E. Groher, and K. I. Michi, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, no. 1, pp. 54–62, 1994.

[13] S. Morinière, M. Boiron, D. Alison, P. Makris, and P. Beutter, "Origin of the sound components during pharyngeal swallowing in normal subjects," *Dysphagia*, vol. 23, no. 3, pp. 267–273, 2008.

[14] A. Santamato *et al.*, "Acoustic analysis of swallowing sounds: A new technique for assessing dysphagia," *J. Rehabil. Med.*, vol. 41, no. 8, pp. 639–645, 2009.

[15] C. M. Steele, E. Sejdić, and T. Chau, "Noninvasive detection of thin-liquid aspiration using dual-axis swallowing accelerometry," *Dysphagia*, vol. 28, no. 1, pp. 105–112, 2013.

[16] E. Sejdić, C. M. Steele, and T. Chau, "Classification of penetration–aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1859–1866, Jul. 2013.

[17] S. S. Shirazi, A. H. Birjandi, and Z. Moussavi, "Noninvasive and automatic diagnosis of patients at high risk of swallowing aspiration," *Med. Biol. Eng. Comput.*, vol. 52, no. 5, pp. 1–7, May 2014.

[18] Y. S. Lee, P. N. Pathirana, C. L. Steinfort, and T. Caelli, "Monitoring and analysis of respiratory patterns using microwave doppler radar," *IEEE J. Transl. Eng. Health Med.*, vol. 2, Nov. 2014, Art. ID 1800912.

[19] D. Jayatilake *et al.*, "Swallowscope: A smartphone based device for the assessment of swallowing ability," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inf. (BHI)*, Jun. 2014, pp. 697–700.

[20] M. Nagae and K. Suzuki, "A neck mounted interface for sensing the swallowing activity based on swallowing sound," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2011, pp. 5224–5227.

[21] V. W. Zue and R. A. Cole, "Experiments on spectrogram reading," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4. Apr. 1979, pp. 116–119.

[22] R. X. Gao and R. Yan, "Continuous wavelet transform," in *Wavelets*. New York, NY, USA: Springer-Verlag, 2011, pp. 33–48.

[23] M. A. Crary, G. D. Carnaby, and I. Sia, "Spontaneous swallow frequency compared with clinical screening in the identification of dysphagia in acute stroke," *J. Stroke Cerebrovascular Diseases*, vol. 23, no. 8, pp. 2047–2053, 2014.

DUSHYANTHA JAYATILAKE (M'08) received the B.Sc. (Hons.) degree in engineering from the University of Peradeniya, Sri Lanka, in 2002, and the Ph.D. degree in engineering from the University of Tsukuba, Japan, in 2012. In 2011, he was a Visiting Student with the Biomedical Robotics and Biomicrosystems Laboratory, University Campus Bio-Medico, Rome. From 2003 to 2008, he was a Lecturer with the Department of Production Engineering, University of Peradeniya, and currently a Researcher with the Faculty of Engineering, Information and Systems, University of Tsukuba. His areas of interest include assistive robotics, rehabilitation robotics, mechatronics, and biomedical engineering.

TOMOYUKI UENO received the M.D. degree from the University of Tsukuba, Japan, in 2001. In 2008, he became a Board Certified Neurologist of the Japanese Society of Neurology, and a Board Certified Physiatrist of the Japanese Association of Rehabilitation Medicine in 2013. From 2006 to 2011, he was an Assistant Professor with the Department of Neurology, Ibaraki Prefectual University, and currently an Assistant Professor with the Department of Rehabilitation, University of Tsukuba Hospital. His areas of interest include rehabilitation for stroke patients and dysphagia patients.

YOHEI TERAMOTO received the B.S. degree in civil engineering from Chubu University, Aichi, Japan, in 2000, license as a Speech Therapist from the Tokaiiryou-Fukushi Vocational School, Aichi, in 2005, and the master's degree in medical sciences from the University of Tsukuba, in 2015. He was a Construction Consultant from 2000 to 2004. Sinec 2007, he has been with the University of Tsukuba Hospital. His areas of interest include neuroscience and rehabilitation robotics.

KEI NAKAI received the M.D. degree from the School of Medicine, Gumma University, Japan, in 1994, and the Ph.D. degree in medicine from the University of Tsukuba, Japan, in 2005. From 1994 to 2000, he was a Resident in Neurosurgery with the University of Tsukuba Hospital. Since 2007, he has been an Assistant Professor with the Department of Neurosurgery, Faculty of Medicine, University of Tsukuba. He has received the Japanese Board Certificate of Neurological Surgery in 2001, and the Japanese Board Certificate of Stroke and Cerebrovascular Diseases in 2005, and has been a General Clinical Oncologist of the Japanese Board of Cancer Therapy since 2008. His areas of interest include neurosurgery, brain tumor, boron neutron capture therapy, and rehabilitation robotics. He is a member of the International Congress on Neutron Capture Therapy, The Japanese Neurosurgical Society, The Japan Stroke Society, The Japanese Association of Rehabilitation Medicine, the Japanese Society of Spinal Surgery, and The Japan Society for Neuro-Oncology.

KIKUE HIDAKA received the bachelor's degree in sociology from Meiji Gakuin University, Japan, in 1995, and the master's in medical science and the Ph.D. degree in medicine from the University of Tsukuba, Japan, in 2001 and 2005, respectively. She is currently a Professor with the Faculty of Medicine, University of Tsukuba. Her areas in interest include coma patients, nursing, rehabilitation, and dysphagia. She is a Licensed Nurse and a Speech Therapist, and also a member of the Japan Geriatrics Society, the Japan Primary Care Association, the Japanese Society of Nursing Art and Science, the Japanese Society of Nursing Research, and the Japan Coma Society.

SATOSHI AYUZAWA received the M.D. and Ph.D. degrees from the University of Tsukuba, in 1986 and 1998, respectively. In 1992, he became a Board Certified Neurosurgeon of the Japan Neurosurgical Society, and a Board Certified Member of the Japanese Society of Clinical Neurophysiology in 2008. Since 2010, he has been the President of the Society for Mind-Body Science. From 2005 to 2012, he was an Assistant Professor with the Department of Neurosurgery, University of Tsukuba, and currently an Associate Professor with the Department of Health, Tsukuba University of Technology. His areas of interest include functional neurosurgery, complementary and alternative medicine, and integrative medicine.

KIYOSHI EGUCHI received the B.Medicine degree and the Ph.D. degree in medical science from the University of Tsukuba, Japan, in 1982 and 1993, respectively. From 1993 to 2000, he was an Assistant Professor with the Department of Rehabilitation Medicine, Saitama Medical School, Japan. He was an Associate Professor until his death with the Faculty of Medicine and the Director of the University Hospital's Department of Rehabilitation, University of Tsukuba. Since 1989, he has been a Board Certified Specialist of the Japanese Association of Rehabilitation Medicine.

AKIRA MATSUMURA received the M.D. degree from the University of Tsukuba, in 1980, and the Doctor of Medicine degree from Georg-August University, Germany. He is a member of the International Society for Neutron Capture Therapy, the International Society for Magnetic Resonance in Medicine, the Japanese Society for Magnetic Resonance in Medicine, the Japanese Society for Pediatric Neurosurgery, the Japanese Society of Spinal Surgery, and the Japan Society of Brain Tumor Pathology. He is a Professor of Medicine and a Vice President with the University of Tsukuba. He is also the Director of the University of Tsukuba Hospital. His areas of interest include neurosurgery, neurooncology, spinal surgery, stroke, neutron capture therapy, charged particle therapy, and neuroradiology.

KENJI SUZUKI (M'03) received the B.S. degree in physics and the M.E. and Ph.D. degrees in pure and applied physics from Waseda University, Tokyo, Japan, in 1997, 2000, and 2003, respectively. He is currently an Associate Professor with the Center for Cybernics Research, and also the Principal Investigator of the Artificial Intelligence Laboratory with the University of Tsukuba, Japan.

He was a Research Associate with the Department of Applied Physics, Waseda University, from 2002 to 2005. He was also a Visiting Researcher with the Laboratory of Physiology of Perception and Action, College de France, France, in 2009, and the Laboratory of Musical Information, University of Genoa, Italy, in 1997. His research interests include assistive and rehabilitation robotics, cognitive robotics, biosignal processing, and computational behavior science.