

Automated Acoustic Analysis in Detection of Spontaneous Swallows in Parkinson's Disease

Marzieh Golabbakhsh · Ali Rajaei ·
Mahmoud Derakhshan · Saeed Sadri ·
Masoud Taheri · Peyman Adibi

Received: 9 December 2013 / Accepted: 31 May 2014 / Published online: 24 June 2014
© Springer Science+Business Media New York 2014

Abstract Acoustic monitoring of swallow frequency has become important as the frequency of spontaneous swallowing can be an index for dysphagia and related complications. In addition, it can be employed as an objective quantification of ingestive behavior. Commonly, swallowing complications are manually detected using videofluoroscopy recordings, which require expensive equipment and exposure to radiation. In this study, a noninvasive automated technique is proposed that uses breath and swallowing recordings obtained via a microphone located over the laryngopharynx. Nonlinear

diffusion filters were used in which a scale-space decomposition of recorded sound at different levels extract swallows from breath sounds and artifacts. This technique was compared to manual detection of swallows using acoustic signals on a sample of 34 subjects with Parkinson's disease. A speech language pathologist identified five subjects who showed aspiration during the videofluoroscopic swallowing study. The proposed automated method identified swallows with a sensitivity of 86.67 %, a specificity of 77.50 %, and an accuracy of 82.35 %. These results indicate the validity of automated acoustic recognition of swallowing as a fast and efficient approach to objectively estimate spontaneous swallow frequency.

M. Golabbakhsh (✉)
Medical Image and Signal Processing Research Center,
Isfahan University of Medical Sciences, Hezarjarib Street,
81745-319 Isfahan, Iran
e-mail: marzgol@gmail.com

A. Rajaei
Isfahan Neurology Research Center, Isfahan University of
Medical Sciences, Hezarjarib Street, 81745-319 Isfahan, Iran
e-mail: slp.alraj@yahoo.com

M. Derakhshan · S. Sadri
Department of Electrical and Computer Engineering, Isfahan
University of Technology, 84156-83111 Isfahan, Iran
e-mail: derakhshanmahmoud@ymail.com

S. Sadri
e-mail: sadri@cc.iut.ac.ir

M. Taheri
Isfahan Health Management of Social Security Organization,
Salman Farsi Street, Ghadir Building, Isfahan, Iran
e-mail: theiryalbert@yahoo.com

P. Adibi
Department of Internal Medicine, School of Medicine,
Isfahan University of Medical Sciences, Hezarjarib Street,
81745-319 Isfahan, Iran
e-mail: adibi@med.mui.ac.ir

Keywords Swallowing assessment · Acoustic ·
Spontaneous swallow · Swallow frequency · Deglutition ·
Deglutition disorders

Introduction

Swallowing is one of the most complicated and integrated mechanisms that the human body performs. The gold standard for swallowing evaluation is videofluoroscopy, which allows the visualization of recorded swallows for a more precise diagnosis. However, this procedure involves expensive equipment, can be carried out only in a clinical setting, and requires radiation exposure which limits its duration. Acoustic means have been proposed as a noninvasive method for swallowing analysis. Microphones and/or accelerometers are used to record breath and swallowing sounds, which are then analyzed using digital signal processing techniques. The research on using swallowing sounds to supplement the clinical evaluation of dysphagia has shown promising results [1–5].

A model to explain swallowing sound generation has been proposed by Yadollahi and Moussavi [6]. The model simulates the physiological causes of swallowing sounds and indicates the pharyngeal wall characteristics. In later studies, modifications to the mathematical modeling were presented and evaluated based on their application for diagnosing swallowing disorder (dysphagia) [7, 8].

Swallowing sounds can be manually distinguished from breath sounds by auditory means, which is time-consuming and subjective. As such, automated detection algorithms have been reported [1, 9–15]. In the report by Makeyev et al. [9], an algorithm based on multilayer feed forward neural networks was proposed for decomposing tracheal sounds into swallowing and respiratory sound segments. In other reports [10–15], tracheal sound signal features were used to classify a tracheal sound segment as either a swallowing sound or a breath sound.

Swallowing fulfills two important needs: oral ingestion of food and liquid and pharyngeal clearance [16]. However, in all of the above-mentioned studies, subjects were fed in order to swallow. Cray et al. [16] have shown that an isolated acoustic recording technique is a suitable approach for measuring spontaneous swallow frequency. Changes in spontaneous swallow frequency may be a susceptible indicator of swallowing dysfunctions [16]. Frequency of swallowing has also been studied for objective quantification of ingestive behavior [11, 17, 18]. In the latter works, sensors were placed in a wearable monitoring device and chewing and swallowing patterns of participants were registered. Swallowing sounds were recorded using a microphone located over the laryngopharynx. In Sazonov et al. [17], the sounds were detected manually; however, the same data were used in a later study to implement automatic detection. The manual scores of chews and swallows were used as the gold standard to train automatic machine-learning software.

As machine-learning algorithms need a training phase, they make sustained learning difficult. In general, machine-learning techniques are data oriented: they model the relationships enclosed in the training data set. If the training data set is not a legitimate representation of the problem domain, the resulting model may differ from the actual problem domain [19].

During the last decade, nonlinear diffusion filters (NLDF) became a solid and well-established tool in signal and image analysis. Using a priori knowledge, these models simplify images while preserving and enhancing semantically important information such as edges, lines, or flow-like structures [20].

The purpose of this study was to develop an automatic robust method for the detection of spontaneous swallows from the sound signal recorded over the laryngopharynx.

This method can be used for automatic detection of swallowing frequency in the future.

Method

Data

Data were collected from 34 subjects (19 males) who had Parkinson's disease (PD) (age = 59.85 ± 11.46 years). They were referred for videofluoroscopy swallow study (VFSS) assessment as part of their routine medical care. Patients' demographic data, including age, sex, time from commencement of disease, and the Hoehn and Yahr (HY) scale score showing PD stage, were collected via an informed interview. The study has been approved by the Biomedical Ethics Board of Isfahan University of Medical Sciences and all subjects or their legal guardian signed a consent form to participate in the experiments. Two tests were performed on each subject: sound recording and oropharyngeal videofluoroscopy. The two tests were performed independently, with the videofluoroscopy study used to validate which subjects suffered from aspiration in general.

Acoustic Recording Study

The acoustic recording was conducted in the fluoroscopy room, which was at 24°C and had indoor lighting. Ambient noise was minimized during the recordings. Swallowing sounds were recorded via microphone (C417 omnidirectional condenser lavalier microphone, AKG Acoustics, Austria) connected to a portable digital sound recorder (Edirol R-44, Japan). The microphone was fixed at the center of a 1.5-cm-diameter rubber piece. The rubber piece was attached to the skin surface using double-sided adhesive tape. The recorder amplified and digitized the sound signal at a sampling rate of 44.1 kHz. The microphone was placed over the laryngopharynx for 15 min while the subject was in a seated position and watched a nature movie. Spontaneous swallows that were registered were innate saliva swallows and no food was used in this test. The number of swallows that each subject performed varied within the 15-min period. All data recordings were carried out by a speech language pathologist (SLP) and a biomedical engineer.

Videofluoroscopy Swallowing Study (VFSS)

VFSS was used as it is considered the gold standard for evaluating swallowing disorders. The subjects' videofluoroscopy recordings were analyzed to identify aspirations to

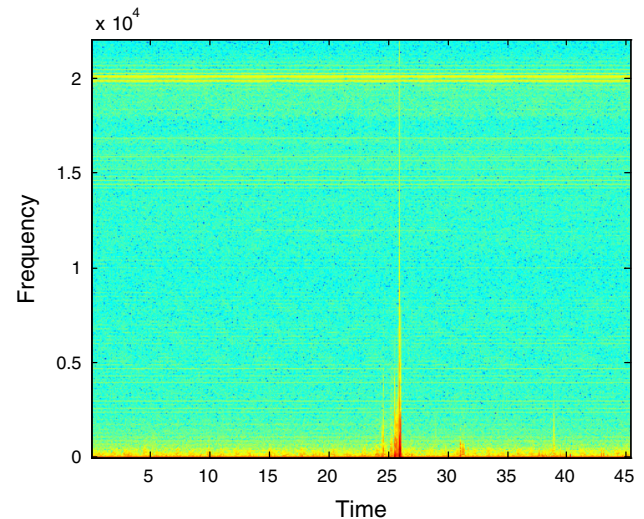
Table 1 VFS results for 34 patients

Patient No.	Age (years)	Gender	Hoehn-Yahrstage (HY)	Time from start of PD (years)	VFS result
1	58	F	1	9	Nonaspirated
2	61	M	1	–	Aspirated
3	71	M	1	2	Nonaspirated
4	65	F	2	12	Nonaspirated
5	72	F	1	–	Nonaspirated
6	70	M	1	18	Nonaspirated
7	46	F	1	8	Nonaspirated
8	58	F	3	4	Aspirated
9	60	F	2	6	Nonaspirated
10	61	M	2	–	Nonaspirated
11	76	F	4	8	Aspirated
12	30	F	2	–	Nonaspirated
13	63	M	1	2	Nonaspirated
14	70	M	3	2	Nonaspirated
15	58	M	1	1	Nonaspirated
16	35	M	2	9	Nonaspirated
17	79	F	2	5	Nonaspirated
18	60	F	1	1	Nonaspirated
19	61	M	1	2	Nonaspirated
20	61	M	1	3	Nonaspirated
21	57	M	2	6	Nonaspirated
22	71	M	1	2	Nonaspirated
23	79	F	4	2	Nonaspirated
24	68	F	4	–	Nonaspirated
25	56	M	1	–	Nonaspirated
26	60	M	2	–	Aspirated
27	36	F	2	12	Nonaspirated
28	58	M	2	1	Nonaspirated
29	58	M	2	3	Nonaspirated
30	51	M	2	2	Nonaspirated
31	50	M	2	4	Nonaspirated
32	62	F	2	4	Nonaspirated
33	65	M	2	–	Aspirated
34	49	F	2	3	Nonaspirated

Dash (–) means data are missing

compare with the acoustic signal's characteristics of the same subjects.

Considering VFSS results along with the recorded sound signals aids in the development of the relationship between the acoustic signal's characteristics and the physiology of swallowing. This leads to a better design of sound signal analysis steps for swallowing detection. For example, different models may be required for the aspirated versus nonaspirated group. However, our hypothesis was that swallowing detection could be done effectively with a single model.

**Fig. 1** Sound spectrum

During the VFSS swallow evaluations, at least four bolus types were administered: 5 ml of thin liquid from a spoon, thin liquid from a cup, 5 ml of a semisolid (pudding) from a spoon, and a cookie. Patients were directed to sit in a posture similar to when they eat their regular food. Subjects were imaged in the lateral projection. An SLP who was trained in administering and interpreting VFSS results evaluated and classified them as aspirated or nonaspirated. Results are shown in Table 1.

To find interrater reliability, 22 of the VFSS results were assessed again by a radiologist who is expert in evaluating VFSS results.

Swallowing Detection

For calculation purposes in MatLab, the sound signal was segmented into 20 intervals of 45 s. For each interval, a spectrogram of the signal was computed using a hamming window and taking the Fourier transform (FFT) (Fig. 1). A single judge labeled the sound intervals of each subject's signal as either a swallow segment or a nonswallow segment by listening to the audio signal and simultaneously reviewing its visual representation shown in the spectrogram. This identification was set as the reference point. To be consistent, for each subject, a 45-s interval of each recording that included at least one swallow was used as a swallow segment, and one 45-s interval of the same subject that did not include any swallows was considered a nonswallow segment.

The spectrogram image was then filtered by NLDF, which builds a smoother image in grayscale (Fig. 2). This filter makes the swallows stronger while decreasing the effect of breath sound and noise. Suppose that the gray-

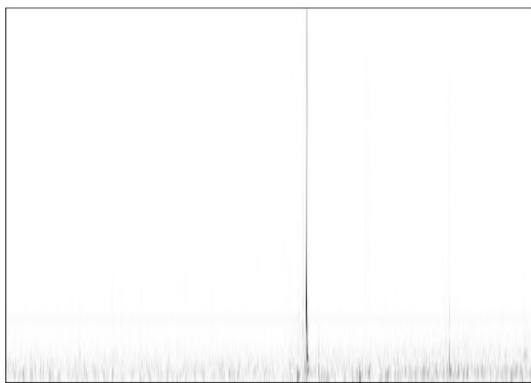


Fig. 2 Gray-scale image of spectrum

Table 2 Frequency ranges

f_{min} (Hz)	f_{max} (Hz)
50	300
150	450
50	600
300	600

scale image is presented by the matrix $A(T,F)$, where T is the vector containing the time samples and F is the vector of frequencies. Each element of the matrix has a value that determines its gray level (Fig. 2).

To find swallows, at each time sample t , we obtained

$$s = \sum_{f=f_{min}}^{f=f_{max}} A(t,f)$$

Swallows could be easily classified by setting a threshold for each vector of S . If part of the signal was noisy, it often led to a false-positive detection. Four different frequency ranges of the spectrum were selected as f_{min} and f_{max} to calculate S (Table 2). Figure 3 shows the analysis steps.

Subjects provided a total of 34 swallow segments with 48 swallows and 34 nonswallow segments. To calculate the accuracy, specificity, and sensitivity, true positives were defined as the segments marked as swallow, and true negatives as the nonswallow segments. Sensitivity, specificity, and accuracy were calculated for each frequency range separately.

Results

The frequency range of 150–450 Hz led to the highest sensitivity of detection for spontaneous swallows, 86.67 %,

and the highest specificity, 77.50 %, with an accuracy of 82.35 %. Five subjects showed signs of aspiration on VFSS evaluation. Table 1 gives the VFSS results for the 34 subjects, along with a description of the data. The Mann–Whitney test showed no significant differences between individuals with or without aspiration in terms of HY stage ($p = 0.168 > 0.05$), age ($p = 0.421 > 0.05$), and gender ($p = 0.843 > 0.05$).

Mean interrater measurement of reliability percentage for judgment of VFSS results was 96 % based on the interitem correlation (ICC) matrix.

Discussion

This study investigated the automatic detection of spontaneous swallows by acoustical means. In voluntary swallowing, perioral, lingual, and submental striated muscles activate in a coordinated fashion in the oral phase. However, in spontaneous swallowing, the oral phase is bypassed most of the time [21]. Therefore, the sound signal is shorter and more abstract compared to data of previous studies on automatic detection of swallowing events. In fact, VFSS is not a suitable tool for evaluation of spontaneous swallows, as the administration of food or fluid, or the instruction to swallow, does not maintain the natural swallow frequency. One of the major strengths of the current study was that it was designed to record innate spontaneous swallows from subjects in a relaxed setting.

The proposed method’s accuracy for detecting swallows was 82.35 %, which is comparable to that of a similar study [11]. The Sazonov et al. [11] approach consisted of a machine-learning method: support vector machines (SVM), which included training and threefold cross validation. The advantage of the proposed method was that there is no need for training and it does not depend directly on the recorded data. Consequently, this method was less costly and more time efficient. Our dataset consisted of people with normal VFSS results, as well as people classified as “aspirated” based on VFSS results (PD patients at different stages of the disease). Due to the small number of subjects, accuracy, sensitivity, and specificity were not calculated for different demographic groups. In future works, using larger datasets, data could be first divided into different groups based on age, gender, HY, time from the start of disease, and VFSS results (aspirated or nonaspirated). Accuracy, sensitivity, and specificity could then be measured independently for each group to determine if the

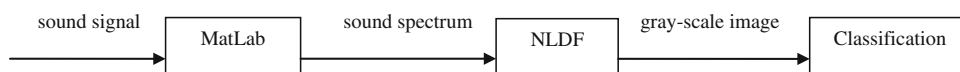


Fig. 3 Diagram of analysis steps

algorithm performs consistently across various cases. Segmenting the data into groups may obtain better results as a more specific threshold could be set in the classification stage, allowing the algorithm to be calibrated independently each time.

In the Sazonov et al. study [11], the sound recording was split into a series of overlapping epochs with an optimal duration. Optimal epoch duration was determined in a grid search procedure from a set of four values: 3, 1.5, 0.75, and 0.375 s. Processing each epoch led to a feature vector. The classifier took the feature vector corresponding to each epoch as input and gave a predicted class label as output. For manual labeling of epochs, if any part of the epoch belonged to a manually marked swallow, it was labeled as a swallow epoch; otherwise it was considered a nonswallow epoch. As predicted, class labels represented the accuracy of the classifier on an epoch level, and a conversion from the epochs to swallowing events was done using manual scores and automatic scores together [11]. In the current study, swallow segments and nonswallow segments were both 45-s intervals and predicted class labels were calculated directly. Therefore, this study designed a more robust procedure for obtaining measures of performance of the swallow detection test.

A limitation of the current study was the need to have a quiet environment. If recording included noise originating from talking, body movements, occasional intrinsic sounds (e.g., coughing), and background noise of different sources, a more complicated algorithm would be needed.

During acoustic recording, breath sounds as well as swallowing sounds were recorded. This could be considered an advantage of the proposed approach. Coordination of deglutition and breathing is crucial for normal pharyngeal function and maintenance of airway protection. Many structures and muscles perform dual functions in respiration and swallowing [22–24]. Hårdemark Cedborg et al. [22] reported on a new instrument, a bidirectional gas flow indicator, to detect airflow during swallowing. The indicator was used to study the coordination of deglutition with breathing in six normal people spontaneously swallowing while breathing air and hypercapnic gas [23]. However, with acoustical recording there is no need to use another tool for airflow, as flow has been shown to have a direct relationship with respiratory sound. Consequently, airflow can be estimated from the respiratory sounds [25, 26]. The sound signals recorded in this study clearly contained respiratory sounds which could be used for flow calculation and analysis. However, a more accurate method would be to use two microphones and simultaneously record the respiratory sounds at the suprasternal notch [25, 26] and the swallowing sounds at the laryngopharynx. These sets of data can be used to investigate the interaction between breathing and swallowing, changes in respiratory cycle

durations surrounding the swallow, and the relationship of breathing and swallowing coordination with the overall swallowing problems of different patient populations with dysphagia [24]. Furthermore, by investigating breathing–swallowing phase relationships and temporal characteristics of swallow apnea, the clinical relevance of integrated breathing–swallowing function on the health and nutrition of dysphagic patients or patients with pulmonary problems [24] could be ascertained. Also of relevance would be a study to see if the increased demand to breathe leads to more swallows during inspiration and/or expiration.

Spontaneous swallow frequency may be a sensitive index for dysphagia and related incidences [16]. It is reported in Crary et al. [27, 28] that spontaneous swallow frequency is a strong indicator of dysphagia in acute stroke. Frequency of swallowing can be studied also for objective quantification of ingestive behavior [11, 17, 18]. In implementing this device, patient care and safety could be improved and would potentially save both time and clinical workflow.

Two main categories of devices in which our method could be incorporated are continuous high-acuity and periodic low-acuity devices. Continuous high-acuity devices are used when patients are in a critical state and need continuous monitoring and the use of biomedical devices. The data are processed from this environment and sent to the electronic medical record (EMR) and validation is completed by the nurse. Periodic low-acuity devices are used when the patient's assessment information is collected periodically to evaluate the patient's condition. These data are also sent to the EMR [29]. The automatic evaluation of swallowing with our proposed method requires no time limit and needs no extra workflow. Therefore, it helps with the continuous monitoring of a patient's state in high-acuity areas. The tool could also be adjusted to automatically report regular assessments of patients in low-acuity areas.

Conflict of interest The authors have no conflicts of interest to disclose.

References

1. Pehlivan M, Yüceyar N, Ertekin C, Çelebi G, Ertaş M, Kalayci T, Aydogdu I. An electronic device measuring the frequency of spontaneous swallowing: digital phagometer. *Dysphagia*. 1996;11(4):259–64.
2. Santamato A, Panza F, Solfrizzi V, Russo A, Frisardi V, Megna M, Ranieri M, Fiore P. Acoustic analysis of swallowing sounds: a new technique for assessing dysphagia. *J Rehabil Med*. 2009;41(8):639–45.
3. Shirazi SS, Buchel C, Daun R, Lenton L, Moussavi Z. Detection of swallows with silent aspiration using swallowing and breath sound analysis. *Med Biol Eng Comput*. 2012;50(12):1261–8.

4. Steele CM, Sejdić E, Chau T. Noninvasive detection of thin-liquid aspiration using dual-axis swallowing accelerometry. *Dysphagia*. 2013;28(1):105–12.
5. Lazarek LJ, Moussavi ZM. Classification of normal and dysphagic swallows by acoustical means. *IEEE Trans Biomed Eng*. 2004;51(12):2103–12.
6. Yadollahi A, Moussavi ZM. A model for normal swallowing sounds generation based on wavelet analysis. 2008 CCECE, 21st IEEE Canadian Conference on Electrical and Computer Engineering, Niagara Falls, Ontario, Canada, 4–7 May 2008.
7. Shirazi SS, Moussavi ZM. Acoustical modeling of swallowing mechanism. *IEEE Trans Biomed Eng*. 2011;58(1):81–7.
8. Aboofazeli M, Moussavi Z. Analysis of swallowing sounds using hidden Markov models. *Med Biol Eng Comput*. 2008;46(4):307–14.
9. Makeyev O, Sazonov E, Schuckers S, Lopez-Meyer P, Baidyk T, Melanson E, Neuman M. Recognition of swallowing sounds using time-frequency decomposition and limited receptive area neural classifier. *Appl Innov Intell Syst*. 2008;XVI:33–46.
10. Aboofazeli M, Moussavi Z. Swallowing sound detection using hidden Markov modeling of recurrence plot features. *Chaos Solitons Fractals*. 2009;39(2):778–83.
11. Sazonov ES, Makeyev O, Schuckers S, Lopez-Meyer P, Melanson EL, Neuman MR. Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. *IEEE Trans Biomed Eng*. 2010;57(3):626–33.
12. Damouras S, Sejdic E, Steele CM, Chau T. An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry. *IEEE Trans Signal Process*. 2010;58(6):3352–9.
13. Aboofazeli M, Moussavi Z. Automated extraction of swallowing sounds using a wavelet-based filter. *EMBS'06 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, August 31, September 3, 2006, vol 1, pp 5607–10.
14. Aboofazeli M, Moussavi Z. Automated classification of swallowing and breath sounds. *EMBS'04 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Conference Proceedings: Linkages for Innovation in Biomedicine: 1–5 September 2004*, San Francisco, vol 6, pp 3816–19.
15. Amft O, Troster G. Methods for detection and classification of normal swallowing from muscle activation and sound. *Pervasive Health Conference and Workshops*, Innsbruck, Austria, November 29, December 1, 2006, pp 1–10.
16. Crary MA, Sura L, Carnaby G. Validation and demonstration of an isolated acoustic recording technique to estimate spontaneous swallow frequency. *Dysphagia*. 2013;28(1):86–94.
17. Sazonov E, Schuckers S, Lopez-Meyer P, Makeyev O, Sazonova N, Melanson EL, Neuman M. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. *Physiol Meas*. 2008;29(5):525.
18. Amft O, Troster G. On-body sensing solutions for automatic dietary monitoring. *IEEE Pervasive Comput*. 2009;8(2):62–70.
19. Mitchell TM. *Machine learning*. WCB. Boston: McGraw-Hill Science/Engineering/Math; 1997.
20. Weickert J. A review of nonlinear diffusion filtering. In: *Scale-Space Theory in Computer Vision, First International Conference, Scale-Space '97*, Utrecht, The Netherlands, July 2–4, 1997, Proceedings (Lecture Notes in Computer Science). Berlin: Springer-Verlag, 1997, No. 1252, pp 1–128.
21. Ertekin C. Voluntary versus spontaneous swallowing in man. *Dysphagia*. 2011;26(2):183–92.
22. Hårdemark Cedborg AI, Sundman E, Bodén K, Hedström HW, Kuylenstierna R, Ekberg O, Eriksson LI. Co-ordination of spontaneous swallowing with respiratory airflow and diaphragmatic and abdominal muscle activity in healthy adult humans. *Exp Physiol*. 2009;94(4):459–68.
23. Cherniack NS. Inspirational lessons from watching swallows. *Exp Physiol*. 2009;94(4):398–9.
24. Martin-Harris B. Coordination of respiration and swallowing. *GI Motility Online*, 16 May 2006. doi: [10.1038/gimo10](https://doi.org/10.1038/gimo10). Available at <http://www.nature.com/gimo/contents/pt1/full/gimo10.html>.
25. Golabbakhsh M. Tracheal breath sound relationship with respiratory flow: modeling the effect of age and airflow estimation. *Dissertation*, University of Manitoba, 2004.
26. Yadollahi A, Moussavi ZM. A robust method for estimating respiratory flow using tracheal sounds entropy. *IEEE Trans Biomed Eng*. 2006;53(4):662–8.
27. Crary M, Carnaby-Mann G, Sura L, Sia I. Spontaneous swallow frequency as a screening protocol for dysphagia in acute stroke. *Neurology*. 2013;80(S32):004.
28. Crary MA, Carnaby GD, Sia I, Khanna A, Waters MF. Spontaneous swallowing frequency has potential to identify dysphagia in acute stroke. *Stroke*. 2013;44(12):3452–7.
29. McAlpine B, Vankampen D. *Clinical engineering and information technology: Working together to implement device integration*. *Biomed Instrum Technol*. 2011;45(6):445–9.

Marzieh Golabbakhsh MSc

Ali Rajaei BSc

Mahmoud Derakhshan BSc

Saeed Sadri PhD

Masoud Taheri MD

Peyman Adibi MD