

Characteristics of the swallowing sounds recorded in the ear, nose and on trachea

Samaneh Sarraf-Shirazi · Jonathan-F Baril ·
Zahra Moussavi

Received: 8 March 2012 / Accepted: 22 June 2012 / Published online: 18 July 2012
© International Federation for Medical and Biological Engineering 2012

Abstract The various malfunctions and difficulties of the swallowing mechanism necessitate various diagnostic techniques to address those problems. Swallowing sounds recorded from the trachea have been suggested as a noninvasive method of swallowing assessment. However, acquiring signals from the trachea can be difficult for those with loose skin. The objective of this pilot study was to explore the viability of using the ear and nose as alternative recording locations for recording swallowing sounds. We recorded the swallowing and breathing sounds of five healthy young individuals from the ear, nose and trachea, simultaneously. We computed time–frequency features and compared them for the different locations of recording. The features included the peak and the maximum frequencies of the power spectrum density, average power at different frequency bands and the wavelet coefficients. The average power calculated over the 4 octave bands between 150 and 2,400 Hz showed a consistent

trend with less than 20 dB difference for the breath sounds of all the recording locations. Thus, analyzing breath sounds recorded from the ear and nose for the purpose of aspiration detection would give similar results to those from tracheal recordings; thus, ear and nose recording may be a viable alternative when tracheal recording is not possible.

Keywords Acoustical analysis · Breath sound analysis · Swallowing sound analysis · Nose and ear recording

1 Introduction

The act of swallowing is one of the most complicated mechanisms of the human body that involves an intricately controlled and coordinated series of events. Any slight mismatch in the timing of the events may cause aspiration (the entry of bolus into the airway). Individuals with neuromotor impairments often have difficulty in swallowing (dysphagia) which includes any swallowing abnormality including aspiration. There are two methods currently used for the swallowing assessment: the fiberoptic endoscopic evaluation of swallowing (FEES) and the videofluoroscopic swallowing study (VFSS); the latter is known as the gold standard exam test for diagnosing swallowing disorders [9]. However, both techniques are invasive, costly and not convenient. In recent years, swallowing sound analysis has been suggested as a noninvasive, low cost and accurate alternative for diagnosis of dysphagia in general [6, 7, 14].

The early acoustical analyses of the swallowing mechanism were focused on the timing of the swallowing events [5, 8, 18]. Later, it was used for deriving swallowing sound main characteristics [16], and also for automatic segmentation of swallowing sounds in relation to its physiological events [1, 2, 11]. Our team's research has shown that the swallowing

S. Sarraf-Shirazi · J.-F. Baril
Department of Electrical and Computer Engineering,
University of Manitoba, Room E3-518 Eng. Bldg.,
75A Chancellor's Circle, Winnipeg, MB R3T 5V6, Canada
e-mail: umsarra@ee.umanitoba.ca

J.-F. Baril
e-mail: umbaril4@cc.umanitoba.ca

Z. Moussavi
Canada Research Chair in Biomedical Engineering,
Department of Electrical and Computer Engineering,
University of Manitoba, Room E3-513 Eng. Bldg.,
75A Chancellor's Circle, Winnipeg, MB R3T 5V6, Canada

Z. Moussavi (✉)
Research Affiliate, Riverview Health Centre, Room 458,
Admin bldg. 1 Morley Ave, Winnipeg, MB R3L 2P4, Canada
e-mail: mousavi@ee.umanitoba.ca;
zahra_moussavi@umanitoba.ca

sounds can distinguish between the dysphagic and control individuals [14]. In addition, we have shown that the tracheal breath sounds after a swallow can be used to detect silent aspiration by the acoustical analysis of the breath sounds [15].

One major limitation with acoustical diagnostic techniques for dysphagia is the difficulty of recording a good quality sound signal if the patient has loose skin over the trachea. Therefore, in this study we explore the viability of using the ear and nose as alternative recording locations for recording breath and swallowing sounds for the purpose of dysphagia and aspiration detection. Although the main motivation for this study comes from the application of the technique in older population, in this pilot study we tested the concept of ear and nose recordings in comparison to tracheal recording only in a few young people. We believe under normal conditions, the relationship between the sounds collected at ear, nose and trachea would not change significantly by age.

Swallowing sound consists of two distinct phases: the initial discrete sound (IDS) and the bolus transmission sound (BTS). The IDS usually has two distinct “clicks” occurring at the beginning of the pharyngeal phase of swallowing, and are associated with the opening of the upper esophageal sphincter [6]. On the other hand, the gurgle sounds that are generated (heard) during the BTS, are associated with the peristaltic contraction waves of the esophagus [6].

The swallowing sounds are commonly recorded over the trachea [5, 18]. Few studies have been devoted to finding the ideal location of sensor placement on the neck area to acquire adequate signals for analysis [17]. However, it is not always possible to record swallowing acoustics from the neck area; thus, the discovery of alternative recording areas is vital for patient diagnosis. Fortunately, the neck is not the only anatomical area where swallowing sounds can be detected [4]. The ear and nose are two under-researched anatomical areas that show great promise for use in the signal-collection of swallowing sounds. Very appropriately these organs in combination with the structures concerned in the swallowing mechanism form an entire field of study in medicine, otorhinolaryngology (the study of the ear, nose and throat). The objective of this pilot study was to make an initial exploratory foray into investigating the possibility of using the ear or nose as regions from which to record swallowing sounds for the purpose of identify aspiration and potential dysphagia.

2 Method

2.1 Experimental data

Five young healthy subjects (20.7 ± 2.3 years, 2 females) participated in this study and gave written consent. The study was approved by the Biomedical Ethics Board of the

University of Manitoba. The subjects were prepared in the following manner after being seated in an acoustically isolated room:

- (a) The neck of each subject was restrained using an Ambu[®] Perfit ACE extrication collar to limit any potential noise contributions due to neck movement.
- (b) The first of three Sony ECM-77B electret microphones (40 Hz–20 kHz bandwidth) was applied over the suprasternal notch using double-sided tape.
- (c) After piercing a hole through the center of a foam earplug (PharmaSystems Quiet Foam uHear Ear Plugs) the second microphone was inserted into the earplug, similar to what was employed in previous studies aimed at detecting respiratory sounds from the ear [12, 13]. The earplug and microphone were then inserted into the subject’s ear and adjusted until a satisfactory signal quality was achieved.
- (d) The third microphone was prepared and inserted into the subject’s nostril such that the microphone remained securely in place during swallowing, and did not fall out. For each subject the nose microphone was prepared by enveloping it in both plastic wrap followed by a fresh 3.5×9.5 cm sheet of 2-ply nose tissue to isolate the microphone from mucus and nasal fluids. The plastic wrap was placed such that it did not occlude the microphone head and the nose tissue was placed such that that a bubble of air remained between the nose tissue and microphone head. The left nostril was used as a default for the test, however, if the signal’s quality was found to be lacking, the alternate nostril was attempted.

The signals were amplified and filtered (5 Hz–5 kHz) using Biopac DA100C, and digitized by NI-DAQ (NI cRIO-9215) at 10,240 Hz sampling rate. After recording, the signals were filtered through a MATLAB Butterworth band pass filter (100–3,000 Hz) to eliminate high-frequency ambient noise, and low-frequency interferences such as heart sounds and muscle artifacts.

Figure 1 shows a diagram of the experimental setup described above. Each subject was handed a disposable drinking cup of water, and asked to use a plastic tablespoon to consume the water with spoon at their own pace but allow only one swallow within one breath cycle. Five to seven swallows were recorded. The bolus size of the water was limited to 15 ml (i.e., one full standard US tablespoon).

2.2 Signal analysis

The swallowing segmentation into IDS and BTS was done by aural and visual examination using the tracheal recording as a reference similar to those in [14]. Figure 2 shows the swallowing and the breath sound signals

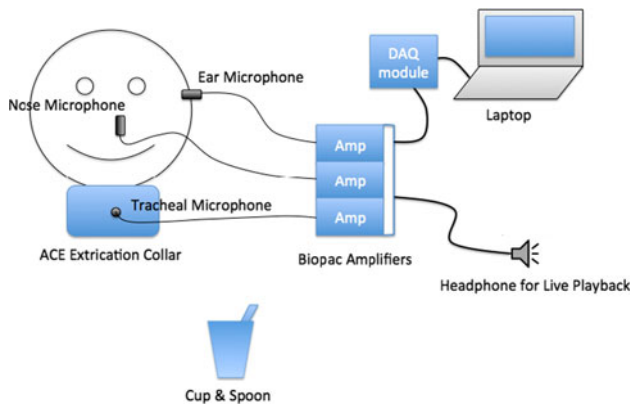


Fig. 1 Setup for swallowing experiment

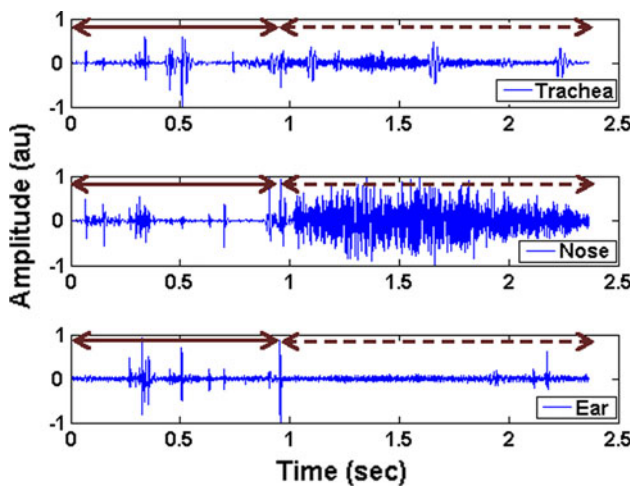


Fig. 2 A typical normalized swallowing and breath sounds signal as marked by the *solid arrow* followed by breath sounds as indicated by the *dashed arrow*; the signals are shown in time domain, and recorded simultaneously from trachea, nose, and ear. *Au* arbitrary unit

recorded at the three locations, trachea, nose and ear in time domain.

The recorded sounds were segmented into three sections: IDS, BTS and the post-swallow breath, which was an expiratory phase for all subjects in this study. Each signal segment was normalized to its variance (energy). Then, we calculated the power spectrum density (PSD) of each of these sections using Welch’s method [10] with 50 % overlapped Hanning windows of 50 ms in length. Figure 3 shows the PSD of the IDS segment of each recording. The three signals were normalized to their variance. The tracheal sound has the lowest magnitude due to the normalization. The tracheal graph and the ear PSDs would interchange, if the signals were not normalized.

We extracted the following features from the PSDs: (a) the peak frequency (f_{peak}) as the frequency at which the peak magnitude occurs; (b) the frequency at which the signal had lost 90 % of its power, called f_{max} ; (c) the

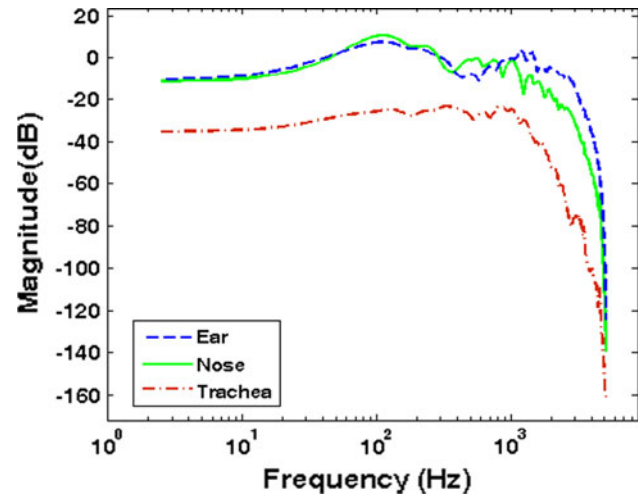


Fig. 3 Typical spectra of the IDS segment of a swallowing sound recorded at the ear, nose and trachea of one subject. Each segment was normalized to its total energy before spectral estimation

average power of the PSD over the octave bands: 150–300, 300–600, 600–1,200, and 1,200–2,400 Hz as were used in a previous study seeking to detect respiratory sounds at the external ear [13].

Lastly, we calculated the approximation wavelet coefficients at the second and third levels of decomposition using Symlet basis function of order 8. The energy of those wavelet coefficients were shown to distinguish between the two groups of dysphagic and control data [14]. Therefore, we were interested in investigating the quality of the nose and ear signals in comparison to tracheal sound with respect to the same characteristic features.

3 Results and discussion

3.1 Qualitative observations of swallow signal quality

Overall, the signals recorded at nose, ear and trachea, all had high-quality with respect to background noise. Compared to signals recorded over trachea, the signal-to-noise (SNR) of the nose breath sounds was higher and SNR of those in ear was lower. The quality of the swallowing sounds for both the ear and nose were highly comparable to that recorded over the trachea. Though, these differences, while noticeable, were slight. An interesting observation from the time domain signals (Fig. 2) was that the final discrete sound (FDS) could be clearly heard in the ear recording. FDS is a short duration click sound at the end of swallow and opening of the airway. It has been speculated to be due to the airway opening. However, based on our experience, FDS is not always present. Comparing all the three signals in the time domain, we found that although the signal recorded in the ear is not as strong as the tracheal

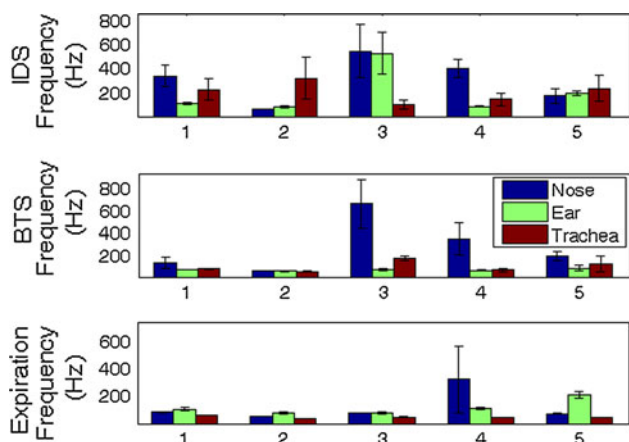


Fig. 4 f_{peak} of the IDS, BTS and the expiration segments for all locations averaged among subject's data

or nasal ones, its FDS segment (if present) could be picked up by the ear microphone, which confirms our speculation about the origin of the FDS segment.

3.2 Analysis of the peak (f_{peak}) and maximum frequency (f_{max})

Figure 4 shows the values calculated as f_{peak} for the water swallows of all the subjects' three recorded signals, and for each section (IDS and BTS) of the swallowing signal. The results showed lack of consistency between subjects and between features as to an exact peak and maximum frequency. A larger more in-depth study may reveal more about the effects of recording location on the peak and maximum frequencies; however, due to the limited sample size in this study, we refrain from drawing conclusions based on the apparent inconsistency of the data.

3.3 Analysis of average PSD magnitude over octave frequency bands

Figure 5 shows the average power calculated in the 4 octave frequency bands, averaged among the subjects, for different recording site. As can be seen, the signals of the three recording sites have a consistent pattern in terms of power over different frequency bands. The ear appears to have the lowest downward sloping trend (3.8 dB/octave frequency step decay for the breath sound), whereas the trachea has the greatest (7.7 dB/octave frequency step decay for the breath sound). The average power values of the ear falling at a slower rate than the nose and the trachea average power at a faster rate than the nose.

The average power calculated for the tracheal recording falls off at higher frequencies at a greater degree than the nose and ear is consistent with the fact that skin acts as a low-pass filter, with a varying degree of strength dependent

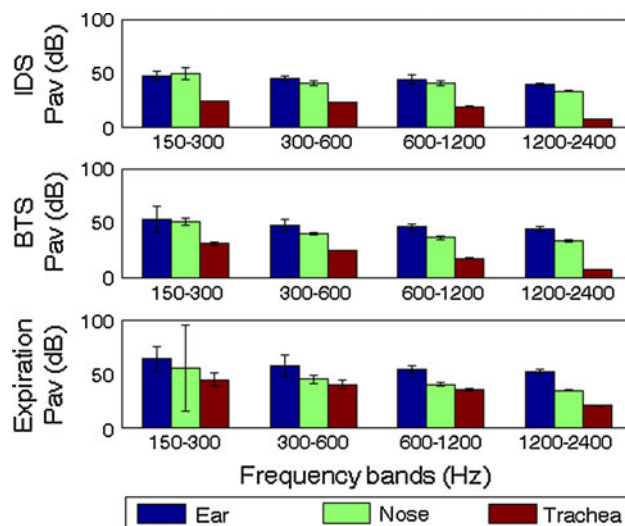


Fig. 5 The average power of the IDS, BTS and the expiration segments calculated over the octave frequency bands. The values are averaged among subject's data

on the skin thickness for frequencies from about 500 to 8,000 Hz [3]. As the ear and nose signals are not recorded through the skin, they do not suffer this effect.

It should also be noted that the ear signal appeared to have a higher noise floor than those recorded at nostril and trachea. This might have contributed to a lower signal drop-off in the higher frequencies (as ambient random noise is constant over all frequencies). Our ear recording results agree with those published in [13]. In that study of breathing sounds, recorded at the external ear, a gradual loss of 10–20 dB in signal strength between the 150–300 Hz and 1,200–2,400 octave bands was noted, which may indicate that the noise floor is of less concern than initially thought due to the noisy nature of the signals recorded in the ear.

As we are interested in the low-frequency components of the breath sounds (below 300 Hz) for aspiration detection, it is important that the average power of the signals does not change in the low frequencies; the higher frequencies are of less importance for aspiration detection. Since the PSDs of the signals of the three recording sites remain similar to each other (with less than 20 dB variation) and consistent in the low frequencies, it may be concluded that the ear and nose may hold promise for use in detecting aspiration.

3.4 Analysis of wavelet coefficients

Figure 6 presents the calculated third order wavelet coefficients for the IDS of water swallows averaged for each subject. It can be seen that there is no consistent pattern between the recording locations for the swallows in either the second or the third order decomposition. The wavelet

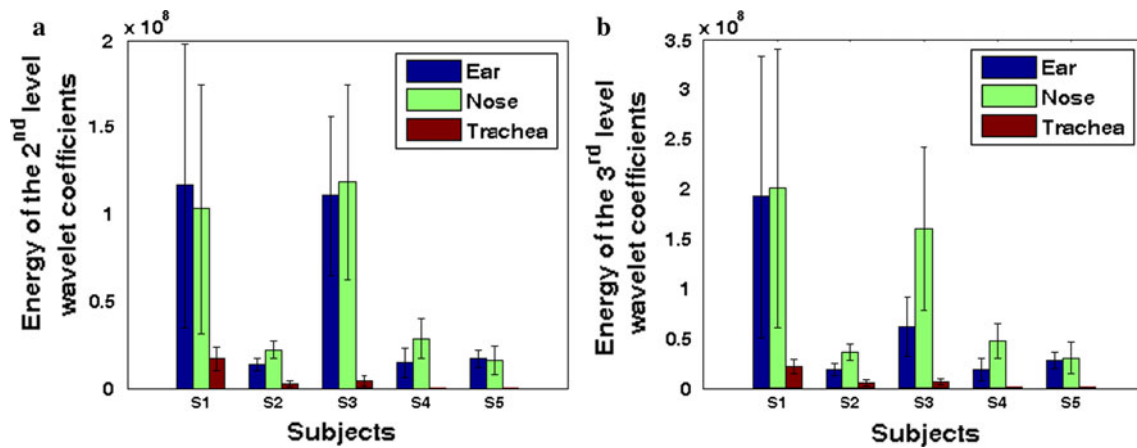


Fig. 6 The mean and standard error of the energy of wavelet coefficients averaged for the IDS segments of all subjects **a** second order decomposition, **b** third order decomposition. The value of the standard error shows the variation within each subject

coefficients and thus the fundamental waveforms are incongruent between the ear, nose and trachea. Thus, recordings from these locations should not be arbitrarily interchanged. This implies that during an acoustical swallowing assessment, if the goal is to diagnose dysphagia in general, all recordings must be taken from either the trachea, or the nose or the ear but not from a mixture of the recording sites.

3.5 Study limitation

There are certain limitations to the results discussed as well as issues discovered during our study that we suggest be considered in subsequent experiments. We found that noise recorded in the ear was strongly dependent on the placement of the microphone in the subject's ear. This is likely due to the variances in ear canal shape and the limitations of using a cylindrical earplug and rigid microphone. It is also important to note that a more thorough study should also have considered normalizations involved with inter-subject variance in the physical characteristics of the ear and nose to account for differences in ear canal shape, nose length, and various other factors. These factors were not considered in this pilot study. We also wish to stress that due to the small sample size used and the nature of this being a pilot study we chose to observe strong visual trends in the data as opposed to calculating precise numerical values whose accuracy and statistical significance could not be guaranteed.

4 Conclusion

In accordance with the objective of this study, we found that recording swallowing and breathing sounds at the ear or nose may be used as alternative recording site as to trachea

depending on the goal of acoustical swallowing assessment. If the goal is identifying people with dysphagia in general, the recording site cannot be used interchangeably between the subjects. On the other hand, if the goal is only to detect the swallows with aspiration within a dysphagic patient, the ear and nose sites may be used as an alternative recording site to trachea in case of a patient having loose skin over the neck. In summary, direct comparisons of swallowing sounds recorded at different sites are not recommended. However, recording swallowing sounds at the ear or nose in cases where tracheal recordings cannot be used is certainly viable for low-frequency breath sound analysis for use of aspiration detection for a dysphagic patient.

References

1. Aboofazeli M, Moussavi Z (2005) Analysis and classification of swallowing sounds using reconstructed phase space features. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP), pp 421–424
2. Aboofazli M, Moussavi Z (2009) Swallowing sound detection using hidden Markov modeling of recurrence plot features. *J Chaos Solitons Fractals* 39(2):778–783
3. Deddens AE, Wilson EP, Lesser THJ, Fredrickson JM (1990) Totally implantable hearing aids: the effects of skin thickness on microphone function. *Am J Otolaryngol* 11(1):1–4
4. Firmin H, Reilly S, Fourcin A (1997) Non-invasive monitoring of reflexive swallowing. *Speech Hear Lang* 10:171–184
5. Hamlet S, Penney DG, Formolo J (1994) Stethoscope acoustics and cervical auscultation of swallowing. *Dysphagia* 9(1):63–68
6. Lazareck LJ, Moussavi Z (2004) Classification of normal and dysphagic swallows by acoustical means. *IEEE Trans Biomed Eng* 51(12):2103–2112
7. Lee J, Blain S, Casas M, Kenny D, Berall G, Chau T (2006) A Radial basis classifier for the automatic detection of aspiration in children with dysphagia. *J Neuroeng Rehabil* 3(14):1–17
8. Leslie P, Drinnan MJ, Finn P, Ford GA, Wilson JA (2004) Reliability and validity of cervical auscultation: a controlled comparison using videofluoroscopy. *Dysphagia* 19(4):231–240

9. Logeman JA (1998) Evaluation and treatment of swallowing disorder. Pro-Ed, Austin
10. Manolakis DG, Proakis JG (2007) Digital signal processing. Prentice Hall, Englewood Cliffs
11. Moussavi Z (2005) Assessment of swallowing sounds' stages with hidden markov model. In: 27th Proceedings of IEEE engineering in medicine and biology (EMBS), pp 6989–6992
12. Pressler GA, Mansfield JP, Pasterkamp H, Wodicka GR (2002) Detection of respiratory sounds within the ear canal. In: Proceedings of the second joint EMBS/BMES conference, vol 2, pp 1529–1530
13. Pressler GA, Mansfield JP, Pasterkamp H, Wodicka GR (2004) Detection of respiratory sounds at the external ear. IEEE Trans Biomed Eng 51(12):2089–2096
14. Sarraf Shirazi S, Moussavi Z (2011) Acoustical modeling of swallowing mechanism. IEEE Trans Biomed Eng 58(1):81–87
15. Sarraf Shirazi S, Moussavi Z (2012) Acoustical detection of silent aspiration. Med Biol Eng Comput (conditionally accepted with revision)
16. Selley WG, Ellis RE, Flack FC, Bayliss CR, Pearce VR (1994) The synchronization of respiration and swallow sounds with videofluoroscopy during swallowing. Dysphagia 9(3):162–167
17. Takahashi K, Groher ME, Michi K (1994) Methodology for detecting swallowing sounds. Dysphagia 9(1):54–62
18. Takahashi K, Groher ME, Michi K (2004) Symmetry and reproducibility of swallowing sounds. Dysphagia 9:168–173