Analysis of swallow sounds of healthy controls for different volumes of water

Divya Giridhar*, Achuth Rao M.V.†, Prasanna Suresh Hegde‡ and Prasanta Kumar Ghosh‡

*Department of Electronics and Communication, R.V. College of Engineering, Bangalore, India
†Department of Electrical Engineering, Indian Institute of Science, Bangalore, India
‡ Department of Head and Neck Surgery, Health Care Global Enterprises Limited, Bangalore, India

Email: divyagiridhar123@gmail.com, achuthr@iisc.ac.in, prasanna.h@hcgoncology.com, prasantg@iisc.ac.in

Abstract—This paper describes the effect of volume of water swallowed by healthy controls on the acoustic sound signals captured by means of cervical auscultation. Cervical auscultation uses a listening device for assessment of swallow sounds. Recordings are carried out using a paediatric stethoscope, coupled with a microphone, fastened to the cervical region of the subject inferior to the cricoid cartilage. In this study, forty one young healthy adults of 20-30 years of age are asked to swallow 2ml, 5ml and 10ml of water each separately six times. Five main components are identified and manually annotated by a clinical expert from each recorded swallow sound based on the swallow function. Twelve parameters are computed for every swallow sound. Peak intensity of the second swallow segment is found to be the best parameter to differentiate different volumes since it changes significantly across three volumes of water considered in this study.

Index Terms—swallow, acoustic analysis, cervical auscultation, swallow sound components

I. INTRODUCTION

Swallowing or deglutition is a process which involves the passage of food from the mouth to the oesophagus through the pharynx. It is a complex action that requires coordination of more than thirty muscles along with the cortical areas of the brain. The pharynx acts as the common entrance to the passage of food and oxygen which makes the swallowing process to be coordinated with respiration.

Swallowing plays a significant role in the digestion of food which is a source of nourishment for humans. Apart from this, swallowing also protects the airway by coordinating with the muscles of the respiratory tract which otherwise may cause pulmonary aspiration [1]. Any difficulty in swallowing is referred to as dysphagia. Dysphagia can be caused due to neurological disorders like stroke, Parkinson’s disease or amyotrophic lateral sclerosis, head and neck cancer and irregularities in the muscles of the pharynx and oesophagus. This can lead to medical conditions such as loss of weight, fatigue, choking, pulmonary aspiration and malnutrition [2].

There are several modes for assessing the swallow function. The first step, usually performed by speech language pathologists, is called screening. This is done to assess the presence of dysphagia and, hence, serves as a base for further evaluations. The common clinical swallow assessment methods include instrumental methods like videofluoroscopy, fibreoptic endoscopy, ultrasonography, surface electromyography and cervical auscultation [3]–[5].

Cervical auscultation is a non-invasive and feasible method of listening to swallow sounds using a stethoscope or a microphone [3], [4]. The cervical auscultation and other instrumental methods often agree to a certain extent [3]. Various studies in the past have carried out acoustical analysis of swallow sounds. Kamiyanagi et al. have evaluated the duration of acoustic signal and peak intensity of swallow sounds in patients who have undergone maxillectomy [6]. Duration of acoustic signal and peak intensity are used to analyse the signal characteristics. The effect of expiration on swallow sounds is discussed and compared with the videofluoroscopic results [5]. Tsuyoshi Honda et al. have characterized the swallow sounds into three components. They have observed the effect of bolus volume, consistency and position of the cervical auscultation device on the acoustic characteristics of swallow signal [7]. Sylvain et al. observed that gender does not play a major role in the parameters used for swallow sound analysis [8]. The authors have used an omnidirectional microphone for recording swallow sounds of different bolus consistencies and divided the swallow signal into five components. The duration of the three swallow components along with the intervals between them was computed and compared for both genders. Mitsuru et al. have used Fast Fourier Transform to analyse the effect of different bolus consistencies on the acoustic swallow signals [9]. A classification of normal and dysphagic swallows based on a discriminant algorithm is implemented which minimizes the need to conduct videofluoroscopic studies [10]. The effect of age on swallow duration and peak intensity is also studied. The duration of the swallow is found to increase with age and the peak intensity is found to decrease with age [11]. Santamato et al. have used the duration, peak intensity and peak frequency to classify healthy controls and dysphagic patients [2]. A smartphone based device, called swallowscope has been developed for the automatic recognition of dry and water swallows [12]. The optimal site for the detection of swallow sounds in the neck region has shown to be the position inferior to the cricoid cartilage encircling the trachea [13]. The electret microphone is considered to be a better acoustic detector than an accelerometer due to its resistance to ambient noise levels and better signal-to-noise ratio [14].

Automatic classification of healthy subjects and dysphagic patients greatly reduces the need for sophisticated medical instruments. It simplifies the protocols conducted during the
clinical assessment of dysphagia. A volume specific model can be used for automatic classification of subjects having dysphagia. For this reason, the basic analysis of the parameters influenced by the volume of substance consumed is essential. Most of the studies have conducted experiments on a small dataset. It is, however, essential to increase the number of subjects in order to have a deeper understanding of the swallow function and discover any common trends across multiple subjects.

In the present study, cervical auscultation is used for the recording of swallow sounds for three different volumes of water consumed. Various parameters pertaining to acoustic analysis of swallow signals are extracted and compared across different volumes. The best parameter to differentiate different volumes of water is the peak intensity ratio of the second swallow segment as it changes significantly across three combinations of water consumed in this study.

II. DATA COLLECTION

A digital cervical auscultation system is set up in a relatively noise-free and well-lit environment. Good source of light is needed to ensure the optimal placement of the device in the cervical region and for an accurate volume measurement of the water. The cervical auscultation system consists of a cervical auscultation device comprising a paediatric stethoscope with an acoustic tube connected to a microphone. A Life-Line Paediatric-Al Stethoscope and a Sorella’z portable 3.5mm microphone with a frequency range of 30-15000Hz, sensitivity of -52dB (with a tolerance of 5dB) and an impedance of 2.2k ohm is used for this purpose. The swallow waveforms obtained from this stethoscope are compared with Eko Core and Littmann stethoscope. Since the latter stethoscopes are tuned mainly for lung and heart signals, the former is used in this study. The device is fastened to the neck of the subject under test by means of a velcro attached to the stethoscope. The other end of the device is plugged into a laptop equipped with a sound recording and analysing software. The device is placed over the lateral border of the trachea as it has been shown to be the optimal position for recording [13]. The recording setup consisting of the digital cervical auscultation system is shown in Fig 1. The placement of the device in the cervical region is shown in Fig 2a. Fig 2b shows the components of the device that consists of the stethoscope, microphone and acoustic tube.

The subject is seated in a comfortable position with the auscultation device tied to the neck in a manner that ensures comfortable breathing and swallowing. It is ensured that the diaphragm mode is in contact with the skin. Water of volume 2ml is measured by means of a syringe and poured into a paper cup. The subject is provided with general instructions which involve sipping the entire water from the cup, keeping it in the mouth and swallowing it in a single gulp only when prompted by the experimenter. A mobile application is used to record the beginning and end timestamps of the swallow sounds along with the audio recording software. These timestamps are used for splitting the recordings containing multiple swallows into audio files each containing one swallow sound. Invalid swallows are discarded. These are typically caused by coughing, drooling, nasal leakage or uttering undesirable sounds while swallowing. Once swallowed, the same quantity of water is poured into the paper cup and the process is repeated six times following which the audio recording software and the mobile recordings are stopped. Following this protocol for recording,
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Explanation</th>
<th>Units</th>
<th>Parameter</th>
<th>Explanation</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of SC1</td>
<td>Duration of first swallow component – duration of swallow waveform at the beginning of the swallow marked by a set of small peaks</td>
<td>seconds</td>
<td>Duration of SSW1</td>
<td>Duration of first swallow wave – the duration of SC1 from the first peak along with the duration of I1</td>
<td>seconds</td>
</tr>
<tr>
<td>Duration of I1</td>
<td>First interval - interval of inactivity after the first sound component, characterized by absence of any event</td>
<td>seconds</td>
<td>Duration of SSW2</td>
<td>Duration of second swallow wave – combined duration of SC2 and I2</td>
<td>seconds</td>
</tr>
<tr>
<td>Duration of SC2</td>
<td>Duration of second swallow component – duration of swallow waveform starting after the first interval covering the peaks</td>
<td>seconds</td>
<td>PI$_{SSW1}$</td>
<td>Peak intensity of first swallow wave</td>
<td>Watt</td>
</tr>
<tr>
<td>Duration of I2</td>
<td>Second interval – period of silence after the second swallow component up to the re-raising peak</td>
<td>seconds</td>
<td>PI$_{SSW2}$</td>
<td>Peak intensity of second swallow wave</td>
<td>Watt</td>
</tr>
<tr>
<td>Duration of SC3</td>
<td>Duration of third swallow component – duration of set of peaks following the second interval</td>
<td>seconds</td>
<td>DPI</td>
<td>Duration of peak intensity – period of swallow from the beginning of the first swallow component to the instant at which the short term energy contour becomes maximum</td>
<td>seconds</td>
</tr>
</tbody>
</table>

Fig. 4. The top row illustrates the swallow waveforms for 2ml, 5ml and 10ml of water consumed. The bottom row shows the corresponding spectrograms.

The average gap between two consecutive swallows is found to be 15s.

The same procedure is repeated for 5ml and 10ml water where the subject is given rest for about 2 minutes before the recording begins for each volume. A total of forty one subjects’ data is recorded, with an age group of the subjects ranging from 20 to 30 years. The PRAAT software is used for recording the swallow sounds and is set to a sampling frequency of 16000Hz. Table I gives the distribution of subjects across age, gender and educational background.

### III. Swallow Sound Phases

Normal swallow can be divided into three main stages namely – oral, pharyngeal and oesophageal phases [7]. The oral phase involves the voluntary chewing of food which is blended with saliva to form a food bolus. The pharyngeal phase begins by the activation of the receptors by the bolus. The regurgitation of food into the mouth is prevented by the blocking of the tongue in the oral cavity. The wind pipe is closed by the vocal folds. The epiglottis and the larynx come together blocking the trachea which otherwise may cause aspiration. In the oesophageal phase, food bolus traverses along the oesophagus by peristalsis. It involves the repositioning of the larynx and hyoid bone [15]. Fig 4 shows exemplary swallow sound waveforms for 2ml, 5ml and 10ml of water consumed and the corresponding spectrograms.

### IV. Parameters Considered for Acoustic Analysis

The three swallow sound phases discussed in section III are represented by three swallow waves (SSW1, SSW2 and SSW3 in Fig 4). A graphical user interface is implemented in MATLAB using which the entire swallow sound is segmented into five components namely – first swallow component (SC1), first interval (I1), second swallow component (SC2), second
interval (I2) and third swallow component (SC3). The total duration (TD) is the duration of all these components [9]. This gives four points between the start and end of each swallow. Fig 3 gives the schematic of the three swallow sound phases along with the five swallow components and the total duration of swallow. Clearly, the duration of SC1 from the first peak along with the duration of I1 gives duration of SSW1. Duration of SSW2 is the combined duration of SC2 and I2. SSW3 represents the same region as SC3. Table II gives the description of twelve parameters used in the acoustic analysis of swallow sounds.

From Table II, duration of the peak intensity (DPI) is the time interval between the start of the swallow signal and the instant at which the short-time energy contour is maximum. The peak intensity (PI) is the maximum value of the short time energy contour. Fig 5 shows the swallow waveform and its corresponding short-time energy where DPI and PI are illustrated.

![Swallow waveform and short-time energy contour](image)

**Fig. 5.** Swallow waveform (top) and its corresponding short-time energy contour (bottom) for a window size of 5 ms and shift of 1 ms.

### V. RESULTS AND DISCUSSION

The annotations are carried out on recordings from forty one subjects each swallowing three volumes of water six times, thereby producing a total of 738 (41x3x6) swallows. Potential errors could arise from human errors during annotations. After careful inspection, inappropriate swallows are discarded resulting in a total of 432 valid swallows. There is no gain control for the device used for recording. Table III gives the distribution of the valid swallows across volumes for forty one subjects. It is clear that all recorded six swallows are valid swallows for only 3, 5 and 7 subjects (among 41 of them) for 2ml, 5ml, and 10ml volumes respectively. Similarly, only one among six swallows is found to be valid in case of five subjects in case of 2ml volume. This is true for 5ml as well as 10ml volume cases. Fig 6 shows the mean of the parameters along with their corresponding errorbars representing standard deviation, separately for 2ml, 5ml and 10ml water consumed.

Dry swallow (only saliva) is not considered as it is difficult to segment the swallow into its sub-components due to interference from undesired signals, especially the heart rate.

Duration of SC1 for 2ml is 2.45% higher than that for 5ml. This duration, however, decreases for 10ml where it is lower than 5ml by 6.2%. Hence, no particular volumes specific trend is found in the duration of SC1. Duration of I1 decreases by 26.7% from 2ml to 5ml but shows an increase of 11.5% from 5ml to 10ml. There is a decreasing trend observed in the duration of SC2 with increase in volume. Duration of SC2 decreases by 16.6% from 2ml to 5ml and by 3.07% from 5ml to 10ml. An overall drop of 19.16% is observed in the duration of SC2 from 2ml to 10ml. The duration of I2 is found to decrease by 4.59% from 2ml to 5ml and by 0.8% from 5ml to 10ml. Hence, there is a decreasing trend in the duration of I2 with volume. Duration of SC3 does not show any specific trend with volume of water consumed. TD is found to decrease by 11.25% from 2ml to 5ml and by 1.54% from 5ml to 10ml of water consumed. This implies that the total swallow duration decreases with increase in volume of liquid consumed. This is due to the increase in the velocity of the bolus head with increase in volume. It traverses the pharynx faster and, hence, produces a shorter swallow sound [14].

The duration of SSW1 decreases by 18.98% from 2ml to 5ml. A 5.5% drop in SSW1 is observed when the volume changes from 5ml to 10ml. The duration of SSW2 decreases with volume by 12.62% and 2.24% from 2ml to 10ml through 5ml. In contrast, PI_SSW1 shows an increasing trend with volume. It increases by 29.2% and 13% from 2ml to 5ml, and 5ml to 10ml respectively. This is in contrast with the observations provided by Tsutoshi et al. where no significant correlation is found for PI_SSW1 with variation in volume. PI_SSW2 increases by 29.1% from 2ml to 5ml and by 19.7% from 5ml to 10ml and an increase of 54.6% from 2ml to 10ml. This could be caused by increased muscle effort while swallowing a greater bolus volume. Hence, it shows an increasing trend with the volume of water consumed which is in good agreement with previous study [7].

DPI is found to be maximum for 2ml and decreases by 25.5% for 5ml followed by a drop of 7.3% from 5ml to 10ml. PI increases with volume where it increases by 36.8% from 2ml to 5ml and by 48.7% from 5ml to 10ml.

We further understand the significance of the difference in each of these parameters using the t-test. T-test on all three pairs of volume changes are performed. Table IV shows the p-values of the t-test for different parameters. It is clear from the table that the duration of SC2 and DPI change significantly (p-value<10^{-5}) from 2ml to 10ml and they do not show any significant difference from 5ml to 10ml. However, PI varies significantly from 5ml to 10ml and from 2ml to 10ml, but not for the transition from 2ml to 5ml. PI_SSW1 shows a significant change from 2ml to 10ml but not for the other two changes in volume. PI_SSW2 changes significantly for all three combinations of volumes and, hence, it is clear that PI_SSW2 is the best parameter to discriminate different volumes.
VI. CONCLUSION

In this paper, cervical auscultation is applied for three volumes of water and various parameters related to the swallowing function are analyzed. Swallow recordings from 41 healthy subjects are collected and segmented into five components using the graphical user interface in MATLAB. Twelve parameters are computed from the annotated images and their trends with respect to 2ml, 5ml and 10ml of water swallowed are plotted. The most important parameter in studying the swallowing sounds is PI_SSW2 which shows an increasing trend across all volumes. This could imply that a greater volume corresponds to increased muscle effort while swallowing. Additionally, duration of SC2, PI_SSW1, DPI and PI can be applied to study the acoustic swallow signals.

There is scope for extending this study further by obtaining the significant parameters for subjects afflicted with dysphagia. The extent of variability of acoustic swallow signals of a particular volume produced by the same subject can be analyzed. In this study, the swallow waveforms from Eko core, Littmann and Life-Line Paediatric-Al stethoscopes are compared. Since Eko core and Littmann stethoscopes are tuned for heart and lung sounds, they are not effective for the study of swallowing sounds. This implies that further filtering and processing techniques need to be applied to devices to explicitly record swallow sounds. An automatic method can be devised for partitioning the start and end of swallows along with internal segmentation into relevant components. Another domain of study could be the variation of swallow waveforms with bolus consistencies. Neural networks can be implemented for automatic classification of bolus volumes and consistencies of swallowing. Finally, there can be a classification model that differentiates normal and dysphagic swallows caused due to different kinds of impairments.

REFERENCES