REAL-TIME SWALLOWING DETECTION BASED ON TRACHEAL ACOUSTICS

Temiloluwa Olubanjo and Maysam Ghovanloo

Georgia Institute of Technology Department of Electrical and Computer Engineering GTBionics Laboratory, Atlanta GA, USA

ABSTRACT

Wearable systems play an important role in continuous health monitoring and can contribute to early detection of abnormal events. The ability to automatically detect swallowing in real-time can provide valuable insight into eating behavior, medication adherence monitoring, and diagnosis and evaluation of swallowing disorders. In this paper, we have developed a real-time swallowing detection algorithm based on acoustic signals that combines computationally inexpensive features to achieve comparable performance with previously proposed methods using acoustic and non-acoustic data. With data from four healthy subjects that includes common tracheal events such as speech, chewing, coughing, clearing the throat, and swallowing of different liquids, our results show an overall recall performance of 79.9% and precision of 67.6%.

Index Terms— Activity recognition, swallowing detection, acoustic analysis, wearable sensor

1. INTRODUCTION

Swallowing sound patterns differ from person-to-person. Even swallowing sounds recorded from a subject being fed the same bolus texture and size, do not have a unique temporal and durational pattern [1]. Yet, there are certain characteristics within the human swallowing sound that make it identifiable to a trained listening ear. Real-time swallowing detection can play a prominent role in wearable sensor advancement for health monitoring purposes. Research has shown that such wearable systems can provide valuable insight into daily eating behavior [2], [3], activity monitoring [4], [5], and medication adherence monitoring [6].

Currently, one of the primary tools for swallowing assessment is cervical auscultation; listening to the swallowing sounds using a stethoscope [1]. This method is a very subjective solution that depends on the skills and experience of the examiner. In addition, the amount of data collected from even one sensor running throughout an entire day is very tedious to analyze [7]. Therefore, there is need for a non-invasive, objective, and in this case, real-time method for swallowing detection and analysis. A swallowing detection algorithm should be effective in retrieving correct events and omitting non-swallow events while maintaining low processing effort [8]. Various types of sensors have been used for swallowing detection purposes including accelerometers [9], electromyography (EMG) sensors [8] and flex/piezoelectric sensors [10].

At the present time, there is no accurate, inexpensive, nonintrusive means for objective monitoring of ingestive behavior in free-living conditions [11]. The ability to accurately and reliably evaluate dietary intake is of value across a range of populations, including the obese, athletes and even the general population, for the purpose of regulating calorie intake to maintain an optimal body mass for health reasons [12]. EMG sensors and a microphone was used in [8] to classify the human swallowing activities; their sensors could detect swallowing events, and differentiate two levels of volume and viscosity. In addition, significant research effort is being focused on utilizing wearable cameras for dietary monitoring because cameras have become low-cost and small enough that the user can always wear them [4], [13]. An important challenge that limits widespread use of wearable cameras relates to activating the camera only during a user's meal time in order to mitigate privacy concerns [14], [15]. Our work on real-time swallowing detection has the potential to be used as a camera trigger; when the frequency of swallows increases, it may be assumed that the user is either eating or drinking something. This can save image storage space, reduce processing efforts on retrieved images and privacy concerns associated with taking pictures at fixed time intervals throughout an entire day.

Progress is also being made towards the implementation of a chronic disease monitoring system which will aid both clinicians and patients by automating the process of capture, transmission, storage, processing and display of information related to health and chronic diseases [12]. Dysphagia diagnosis and evaluation, and monitoring medication adherence for patients living with chronic illnesses are two important applications that can benefit from real-time swallowing detection. Detection of normal swallowing can contribute towards identification of abnormal swallows that can lead to aspiration, which refers to food or liquid entry into the airway for patients with dysphagia [1], [16]. For the purpose of improving medication adherence for patients living with chronic illnesses, doctors often make vital medical decisions based on a patients' own report of their compliance to administered medication or based on results of indirect monitoring methods [6]. Since the process of normal swallowing may be divided into three distinct phases: oral, pharyngeal and esophageal phases [10], [17], the ability to detect the initial phase of swallowing sounds in real-time can be beneficial for medication adherence monitoring where ingestion of tagged medications can be detected by a patient's wearable neckwear [6].

To the best of our knowledge this paper is the first of its kind to present work on acoustic-based real-time swallowing detection. We focus on using computationally inexpensive features for automatic swallowing detection in order to save power. Specifically, we are interested in quantifying real-time swallowing detection from a continuous recording that includes other tracheal sounds such as speech, coughing, clearing the throat, and chewing. We also explore swallowing detection independent of the substance being swallowed by including liquids of different viscosities in our experiment. The experimental procedure, data collection process and swallowing detection methodology are presented in Section 2. In Section 3, we present the data analysis method used and results achieved from this study. This is followed by our conclusion and direction for future work towards testing real-time swallowing detection in free-living conditions in Section 4.

2. EXPERIMENT

2.1. Data Acquisition

Acoustic data was collected with a throat microphone placed over the suprasternal notch of the trachea with a sampling rate of 16 kHz. We used the IASUS NT3 throat microphone because it only picks up vibrations generated by the wearer's larynx [18]. It therefore has a low response in the outward direction and is insensitive to background noise [19]. Fig. 1 shows that according to the Nyquist theorem, a sampling frequency of 16 kHz is sufficient to preserve important characteristics of the swallowing sound that reach a maximum frequency of about 1.5 kHz [4].

Experimental data was recorded from six subjects (3 males, 3 females, ages 20 - 35 years old) with no history of swallowing disorders as they were instructed to perform a variety of activities. Recordings from two subjects were excluded from data analysis due to an incomplete experiment and the throat microphone not maintaining contact during one experiment. This study was approved by the Institutional Review Board of Georgia Institute of Technology and all participants signed a written consent prior to the experiment.

2.2. Feature Extraction

The goal of an efficient real-time swallowing detection algorithm is to use minimal computational resources to achieve high swallowing detection accuracy. This implies the use of computationally inexpensive features to discriminate swallowing from other common tracheal sounds. To achieve real-



Fig. 1. Spectrogram of common tracheal events

time swallowing detection, we used four easy-to-compute features namely: windowed energy, peak frequency, Shannon entropy and wavelet entropy. The wavelet decomposition was computed using Coiflet 4 wavelet as in [11]. Advantages of the Coiflet wavelet include near-linear phase, good amplitude response, and fast computation [20]. LabVIEW's Advanced Signal Processing Toolkit [21] was used to extract coefficients of the Discrete Wavelet Transform (DWT). The delta coefficients at decomposition level 3 of the wavelet transform were then converted into a scalar feature using Shannon's entropy.

Windowed energy (W.E) is described as:

$$W.E(X) = \sum_{i=0}^{n} |X_t(i)|^2$$
(1)

where $X_t(i)$ is the discrete sample amplitude at time t, and n is the number of samples per window frame.

Peak frequency (P.F) is described as:

$$P.F(X) = argmax_{[f=0,fmax]}|F_M(f)|^2$$
(2)

where *fmax* is the highest available frequency in the signal and F_M represents the Fourier transform of the signal [16].

Shannon entropy (S.E) in [21] is described as:

$$S.E(X) = \sum_{i=0}^{n} X_t^2(i) * \log(X_t^2(i))$$
(3)

where $X_t(i)$ is the discrete sample amplitude at time t, and n is the number of samples per window frame.

The acoustic data was processed and features were calculated with a non-overlapping 500 ms window frame. A swallowing event is detected when the pre-selected features per window frame falls within the subject-dependent threshold range (R):

$$if[W.E(X) \in R_{W.E} \cap P.F(X) \in R_{P.F} \\ \cap S.E(X) \in R_{S.E} \cap W.S.E(X) \in R_{W.S.E}]; \\ O_1(X) = 1; otherwise; O_1(X) = 0; \quad (4)$$

where $R_{W.E}$, $R_{P.F}$, $R_{S.E}$, $R_{W.S.E}$ represent the subjectdependent threshold ranges for windowed energy, peak frequency, Shannon entropy and wavelet Shannon entropy respectively. The algorithm's output is represented by $O_1(X)$.

2.3. Training and Testing

The experiment consisted of two data collection sessions conducted on different days to account for physiological variations. Data collected from the first session was used for training. We manually labeled each activity in the training data set according to the observed tracheal event determined by listening to the audio stream, visually inspecting the signal, and validating the event label with the experimental procedure that each subject was instructed to follow. Activities included in the training data set are outlined in Table 1. Although we were more interested in real-time detection of the swallowing sound from a continuous tracheal recording, we included head tilts and head turns as activities in our training phase to ensure that our algorithm was robust towards this type of motion artifact. We then set the subject-dependent feature threshold ranges based on the training data to be used for real-time swallowing detection in the testing experiment.

Activities included in the testing phase are outlined in Table 2. The experimental procedure for testing consisted of two parts. In Part I, each subject was instructed to perform 10 coughs, 10 clearing the throat, 10 swallows of water, 5 dry swallows and 30 seconds of speech. Part II of the testing experiment consisted of the subject reading a standard passage, commonly used in speech pathology, with each of the non-speech activities shown in Table 2 randomly embedded throughout the text. This part of the experiment is intended to imitate a more realistic situation where a subject might be speaking and have to pause in the middle of a speech to swallow, cough, clear their throat etc. Below is a sample of Part II of our experimental procedure for testing:

"{Clear throat - 1 time} You wished to know all {Swallow water - 2 gulps} about my grandfather. {Cough -3 times}. Well, he is nearly ninety-three years old; {Swallow water - 1 gulp}} he dresses himself in an ancient {Clear throat - 2 times} black frock coat {swallow yogurt-3 times}, usually minus several buttons"

In parenthesis are activities the subject was instructed to complete and outside the parenthesis is text the subject was instructed to read.

 Table 1. TRAINING ACTIVITIES FOR 4 SUBJECTS

Training Activity	Total or Duration
Dry swallowing	20
Swallowing water	40
Chewing crackers	148.32 secs
Coughing	40
Clearing the throat	40
Speech	480 words
Head tilts	20
Head turns	20

Table 2. TESTING ACTIVITIES FOR 4 SUBJECTS

Testing Activity	Total or Duration	
Dry swallowing	55	
Swallowing water	66	
Swallowing yogurt	20	
Swallowing orange juice	12	
Swallowing coke	16	
Chewing crackers	290 secs	
Coughing	64	
Clearing the throat	68	
Speech	1168 words	

3. DATA ANALYSIS AND RESULTS

3.1. Data Analysis

Tables 1 and 2 show that a total of 229 swallows were used in this study. Sixty swallowing events (35.5%) from the total number of swallows were used as part of the training data set. The following definitions were used to analyze our results:

- True Positives = Correctly detected swallows
- False Negatives = Undetected swallows
- False Positives = Incorrectly detected swallows

True negatives are ill-defined because they constitute a continuum of times where a swallow does not occur and the algorithm does not detect one [9]. We then used standard measures for information retrieval to assess performance [22]:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$
(5)

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$
(6)

A recall performance equal to 1 means that all prominent events were detected, and a precision performance equal to 1 means that there were no false positives. It is important to note that for the aforementioned potential application areas in Section 1, recall has priority over precision.

3.2. Results

Selection of an appropriate subject-dependent threshold range is critical for our proposed real-time swallowing detection algorithm. CLEAR 2007 [23] performance evaluation method was used to score detection of relevant acoustic events. Temporal coincidence of the annotated signal and output of the algorithm did not affect performance. A swallowing sound event was considered correctly detected if there was at least one detected swallowing event that the temporal center was situated between time stamps of an annotated swallowing event, or if the temporal center of an annotated swallowing sound event laid in between the time stamps of at least one detected swallowing event.

Recall and *Precision* performances were calculated for all subjects using (5) and (6). Table 3 shows a summary of the achieved results for each subject from our experiment. On average, our real-time swallowing detection algorithm achieved 79.9% recall and 67.6% precision. As shown in Table 4, our overall results are comparable with the best recall performance presented in [4] and [8]. We chose to compare our results with [4] and [8] because both of these papers include other tracheal sounds such as coughing, speaking, eating solid foods and drinking different liquids in their experiment.

Yatani and Truong developed *BodyScope*, a wearable acoustic sensor [4]. They achieved an overall performance of 79.4% recall and 79.6% precision for classification of 12 activities using support vector machines as the classifier for leave-one-sample-per-participant-out. More specifically, their swallowing detection recall performance of 78.0% and precision of 66.1% can be seen in gathered samples for drinking events each of which contained one gulping sound from the subject drinking water [4].

On the other hand, Amft and Troster achieved a high 84% recall performance at the expense of high false positives that contributed to a low precision performance of only 18% using feature-level fusion from infra-hyoid EMG (IH-EMG) and microphone (SND) as sensors [8]. Using only the acoustic signal for detection and classification of normal swallowing sounds, they achieved a performance of 73% recall and 15% precision. In a 2nd pass using IH-EMG and SND, the authors improved their precision to 30% but this decreased their recall to 57% [8].

It is important to note that our proposed swallowing detection algorithm in comparison to those presented by [4], [8] is designed for real-time detection; therefore, it is applicable to wearable systems that collect continuous recordings for a long duration and need to respond to swallowing events.

	Experiment			
	Part I		Part II	
	Recall	Precision	Recall	Precision
Subject 1	0.867	0.619	0.880	0.733
Subject 2	0.727	0.500	0.650	0.813
Subject 3	0.824	0.667	0.652	0.600
Subject 4	0.944	0.586	0.852	0.885
Average	0.84	0.593	0.759	0.758

Table 4. COMPARISO	ON WITH	RELATED	WORK
--------------------	---------	---------	------

	This Work	[8] 2006	[4] 2012
Recall	0.799	0.84	0.780
Precision	0.676	0.18	0.661
Real-Time	Yes	No	No

4. CONCLUSION

In this paper, we have presented an algorithm that uses computationally inexpensive features for real-time swallowing detection. From this experiment, we achieved 79.9% recall and 67.6% precision in real-time; this is comparable to performance of related research efforts that detect swallowing offline from a continuous recording which includes other tracheal sounds such as speech, coughing, chewing and head motion. To the best of our knowledge, this work is the first of its kind that presents acoustic-based real-time swallowing detection. Our algorithm is intended for potential application in wearable systems that can provide insight into eating behavior, medication adherence monitoring, and diagnosis and evaluation of dysphagic patients.

Future work includes improving results in this study by exploring more robust classifiers with methods used in speech recognition that may more accurately discriminate swallowing sound from other tracheal sounds. In addition, we will conduct in-the-wild studies where we will test our improved algorithm on continuous recordings for an entire day.

5. ACKNOWLEDGMENT

This work was in part supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1148903 and a seed grant from the Institute for People and Technology (IPaT) at Georgia Institute of Technology. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors of this paper also acknowledge the Achievement Rewards for College Scientists (ARCS) Foundation for partial support of this work.

6. REFERENCES

- M. Aboofazeli and Z. Moussavi, "Analysis of temporal pattern of swallowing mechanism," in *Eng. in Med. and Bio. Soc. IEEE*, 2006, pp. 5591–5594.
- [2] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, "Noninvasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiol. Meas.*, vol. 29, no. 5, pp. 525–41, May 2008.
- [3] O. Amft and G. Tröster, "On-body sensing solutions for automatic dietary monitoring," *Pervasive Computing*, *IEEE*, pp. 62–70, 2009.
- [4] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," New York, NY, USA, 2012, UbiComp, pp. 341–350, ACM.
- [5] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. N. Patel, "Accurate and privacy preserving cough sensing using a low-cost microphone," New York, NY, USA, 2011, UbiComp, pp. 375–384, ACM.
- [6] Y. M. Choi, T. Olubanjo, A. Farajidavar, and M. Ghovanloo, "Potential barriers in adoption of a medication compliance neckwear by elderly population," in *Eng. in Med. and Bio. Soc. IEEE*, 2013, pp. 4678–4681.
- [7] H. Ghasemzadeh, S. Ostadabbas, E. Guenterberg, and A. Pantelopoulos, "Wireless medical-embedded systems: A review of signal-processing techniques for classification," *Sensors Journal, IEEE*, vol. 13, no. 2, pp. 423–437, 2013.
- [8] O. Amft and G. Tröster, "Methods for detection and classification of normal swallowing from muscle activation and sound," in *Pervasive Health*, 2006, pp. 1–10.
- [9] S. Damouras, E. Sejdic, C. Steele, and T. Chau, "An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry," *Signal Processing, IEEE Trans. on*, vol. 58, no. 6, pp. 3352–3359, 2010.
- [10] C. Ertekin, I. Aydogdu, Y. Seçil, N. Kiylioglu, S. Tarlaci, and T. Ozdemirkiran, "Oropharyngeal swallowing in craniocervical dystonia," *Journal of Neurol.*, *Neurosurg., and Psych.*, vol. 73, no. 4, pp. 406–11, Oct. 2002.
- [11] E. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior," *IEEE Trans. on Biomedical Engineering*, vol. 57, no. 3, pp. 626–33, Mar. 2010.

- [12] W. Walker and D. Bhatia, "Towards automated ingestion detection: Swallow sounds," in *Eng. in Med. and Bio. Soc. IEEE*, 2011, pp. 7075–7078.
- [13] G. O'Loughlin, S. J. Cullen, A. McGoldrick, S. O'Connor, R. Blain, S. O'Malley, and G. D. Warrington, "Using a wearable camera to increase the accuracy of dietary analysis," *American Journal of Prev. Med.*, vol. 44, no. 3, pp. 297–301, Mar. 2013.
- [14] A. R. Doherty, S. E. Hodges, A. C. King, A. F. Smeaton, E. Berry, C. J. Moulin, S. Lindley, P. Kelly, and C. Foster, "Wearable cameras in health. the state of the art and future possibilities," *American Journal of Prev. Med.*, vol. 44, no. 3, pp. 320–3, Mar. 2013.
- [15] E. Thomaz, A. Parnami, J. Bidwell, I. Essa, and G. D. Abowd, "Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras," New York, NY, USA, 2013, Ubi-Comp, pp. 739–748, ACM.
- [16] I. Jestrovic, J. M. Dudik, B. Luan, J. L. Coyle, and E. Sejdic, "The effects of increased fluid viscosity on swallowing sounds in healthy adults," *Biomedical Engineering Online*, vol. 12, no. 1, pp. 90, Sept. 2013.
- [17] J. A. Logemann, Evaluation and Treatment of Swallowing Disorders 2nd Ed., PRO-ED, Austin, TX, 1998.
- [18] iASUS, "iASUS NT3 Throat Mic," Available ONLINE: http://www.iasus-concepts.com/product?pid=101.
- [19] B. Denby, T. Schultz, K. Honda, T. Hueber, J. Gilbert, and J. Brumberg, "Silent speech interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270–287, Apr. 2010.
- [20] S. Fu, X. Liu, B. Muralikrishnan, and J. Raja, "Wavelet analysis with different wavelet bases for engineering surfaces," in *Proc. 16th Annu. Meeting Amer. Soc. Precision Eng.*, Raleigh, NC, 2001, pp. 249–252.
- [21] National Instruments, "LabVIEW Advanced Signal Processing Toolkit," Available ONLINE: http://sine.ni.com/nips/cds/view/p/lang/en/nid/209055.
- [22] Y. Yang, T. Pierce, and J. Carbonell, "A study of retrospective and on-line event detection," in *Proc. of 21st Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, New York, NY, USA, 1998, SIGIR, pp. 28–36, ACM.
- [23] A. Temko, R. Malkin, C. Zieger, D. Macho, C. Nadeu, and M. Omologo, "Clear evaluation of acoustic event detection and classification systems," in *Multimodal Technologies for Perception of Humans*, R. Stiefelhagen and J. Garofolo, Eds., vol. 4122 of *Lecture Notes in Computer Science*, pp. 311–322. Springer Berlin Heidelberg, 2007.