Tracheal Activity Recognition Based on Acoustic Signals

Temiloluwa Olubanjo, *IEEE Student Member,* and Maysam Ghovanloo, *IEEE Member*

***Abstract*— Tracheal activity recognition can play an important role in continuous health monitoring for wearable systems and facilitate the advancement of personalized healthcare. Neck-worn systems provide access to a unique set of health-related data that other wearable devices simply cannot obtain. Activities including breathing, chewing, clearing the throat, coughing, swallowing, speech and even heartbeat can be recorded from around the neck. In this paper, we explore tracheal activity recognition using a combination of promising acoustic features from related work and apply simplistic classifiers including K-NN and Naive Bayes. For wearable systems in which low power consumption is of primary concern, we show that with a sub-optimal sampling rate of 16 kHz, we have achieved average classification results in the range of 86.6% to 87.4% using 1-NN, 3-NN, 5-NN and Naive Bayes. All classifiers obtained the highest recognition rate in the range of 97.2% to 99.4% for speech classification. This is promising to mitigate privacy concerns associated with wearable systems interfering with the user's conversations.**

I. INTRODUCTION

Increasing healthcare costs and an aging world population have recently motivated a considerable amount of research effort on wearable health-monitoring systems (WHMS) [1]. It is envisaged that preventive measures with personalized diagnostic approaches would be more cost effective and sustainable for the healthcare system [2]. Wearable sensors facilitate remote patient monitoring and have the potential to extend the reach of specialists in urban areas to more rural areas [3]. Unlike wearable systems such as wristbands, armbands, waist-gears, stomach patches, insole-based activity monitors etc., neck-worn devices provide access to a unique set of health-relevant data and activities that other wearable devices simply cannot access.

Tracheal activities are among signals that if recognized properly can contribute significantly to health monitoring. Common tracheal events that can provide insight into an individual's health and well-being include breathing, chewing, swallowing, coughing, clearing the throat and speech. An acoustic heartbeat signal can also be detected when a sensor is placed on the neck [4]. Various sensor types have been used in literature for tracheal activity recording and analysis including accelerometers [5], [6], electromyography (EMG) sensors [8], flex/piezoelectric sensors [9], and acoustic sensors [4], [6], [10], [11].

A wearable breathing monitoring system was presented in [4] to detect breathing cessation that can be caused by respiratory diseases, neuromuscular diseases, epilepsy, sudden infant death syndrome and sudden adult death syndrome. Corbishley and Rodríguez-Villegas obtained the largest acoustic breathing signal power, especially at low airflow, from a neck-worn device. Using a microphone and conical bell for signal recording, their algorithm detected breathing periods for five subjects with an accuracy of 91.3% [4].

Chewing and swallowing, which are easily recordable from neck-worn devices, are also common physiological tracheal activities of interest for monitoring ingestion behavior. In response to a dramatic increase over the last decades of overweight and obese population in the U.S., methods of food intake detection using only the time series of swallows was investigated in [12] to potentially mitigate consequent threats to life expectancy. Researchers achieved an accuracy of up to 89.4% and 93.9% for group and individual food intake models respectively [12]. Swallowing data is also used for diagnosis of a swallowing disorder called dysphagia. Dysphagia is most common in individuals with neurological impairments such as brain-stem stroke, head/neck injuries, and spinal cord injury with anterior cervical fusion [7]. Patients suffering from dysphagia are likely to choke or aspirate due to the entry of food into the airway below the true vocal folds [5]. Sejdic et al. propose an approach that achieves an accuracy of >90% for classification of swallowing accelerometry recordings containing healthy swallows and penetration aspiration in dysphagic patients [5].

Another tracheal event that can provide insight for health monitoring is coughs. Coughing is a normal protective reflex which clears the respiratory tract and prevents entrance of noxious materials into the respiratory system [11]. Coughing is not usually frequent in healthy subjects, but is a common symptom of many respiratory diseases, including asthma, gastro-esophageal reflux (GOR), postnasal drip, bronchiectasis and chronic bronchitis [13], [14]. Matos et al. achieved 82% average cough detection rate with a false alarm rate of seven events per hour by classifying based on events above an energy threshold relative to each recording's average energy [11]. Similar to coughing, clearing the throat is also another protective mechanism to remove an irritant in the throat [15]. Clearing the throat can be a symptom for dry throat, enlarged tonsils, enlarged adenoids and upper respiratory tract infection. The accuracy of acoustic monitoring for detecting cough and throat clearing was investigated in [15]. The authors found that both event

Research is supported by NSF Graduate Research Fellowship, Achievement Rewards for College Scientists (ARCS) Foundation and a seed grant from the Institute of People and Technology (IPAT) and Children's Healthcare of Atlanta (CHOA).

Temiloluwa Olubanjo is with the GT-Bionics Lab, School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA, USA (e-mail: tolubanjo3@gatech.edu).

Maysam Ghovanloo* is with the GT-Bionics Lab, School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA, USA (e-mail: mgh@gatech.edu).

profiles in pressure topography revealed similar qualitative pattern of pressurization with more vigorous pressure changes and a greater rate of repetitive pressurizations in coughs.

Since the human speech is a combination of linguistics and emotions, several researchers continue to work on automatic emotion recognition using audible paralinguistic cues from speech [16], [17]. A speaker's emotional state expresses itself in speech through paralinguistic features such as pitch, speaking rate, voice quality and laugher [17]. In [16] and [17], automatic emotion recognition was explored using Gaussian mixture models and artificial neutral networks, respectively.

The aforementioned research endeavors work to detect and utilize specific tracheal events in order to draw certain inferences and/or conclusions for health monitoring purposes. Little research effort has been committed to recording and discriminating common tracheal activities from one another to facilitate the potential benefit of a neck-worn WHMS. This paper focuses on detecting and classifying five common and easily replicable tracheal activities namely chewing, clearing the throat, coughing, swallowing and speech from acoustic recordings of a neck-worn WHMS. Since power consumption is of primary concern in wearable systems, we explore the potential of a sub-optimal sampling frequency for tracheal activity recognition. More specifically, we assess whether a sampling rate of 16 kHz is sufficient to preserve enough characteristics of the tracheal activities of interest to facilitate classification. In section II, we present the data collection method, experimental procedure, feature extraction and classifiers used. Data analysis is presented in section III, results and discussion in section IV, and the conclusion along with future works in section V.

II. Methodology

A. Data Collection

Experimental data for five healthy subjects (2 males and 3 females, ages 20 - 35 years old) was adopted from a previous study where a sampling rate of 16 kHz was used [6]. The data was recorded with IASUS NT3 throat microphone [18] placed over the suprasternal notch while subjects were in a sitting position. From the tracheal activities of interest, swallowing has a bandwidth up to 1.5 kHz, speech, up to 4 kHz [10], and chewing, up to ~ 6 kHz depending on the substance being chewed. On the other hand, coughing and clearing the throat can reach frequencies up to ~ 15 kHz [19]. Therefore, according to the Nyquist theorem, 16 kHz is sufficient to preserve important characteristics of chewing, swallowing and speech but possibly not all of the characteristics of coughing and clearing the throat [6], [10].

B. Experimental Procedure

To account for physiological variations, experimental data was collected from subjects on two different days. Data from day 1 was used for training while data from day 2 was used for testing. Table I shows a summary of the dataset used

TABLE I. Data Summary for five subjects

Activity	Day 1 - Training	Day 2 - Testing	Total Number
Chew	65	61	126
Clearing Throat	51	48	99
Cough	51	52	103
Swallow	51	55	106
Speech	121	95	216
Total Number of Tracheal Activities			**650**

in this experiment. The 'chewing' activity consisted of each subject chewing two crackers while the 'swallow' activity consisted of each subject swallowing some water and a few dry swallows when they were audible and visibly recognizable during activity labeling. The speech activity consisted of each subject reading the same text.

C. Feature Extraction

Tracheal activities were isolated and annotated from the continuous recording by listening to audio stream, visually inspecting the signal, and validating the event label with the experimental procedure. Acoustic features were extracted from each isolated activity using a window size of 1000 samples (62.5 msecs) with 50% overlap. In an effort to achieve good clustering of features from each activity, we compiled promising features that have been used to obtain acceptable performance from related works [5], [6], [10], [11], [22]. A total of 47 features from the time, frequency, and cepstral domains were used for training and testing of each classifier. Features from each window frame per tracheal event were then averaged to obtain one real number to represent each activity.

Time domain features:

- *Windowed Energy (W.E.)*: Frame-level feature that provides short-term characteristics of the windowed frame [20]. W.E. has been shown to be a relevant feature for real-time swallowing detection in [6].
- *Total Variation (TV)*: Essentially the Manhattan or L_1 norm of derivatives. TV was initially introduced in [21] for image denoising and reconstruction.
- *Zero Crossing Rate (ZCR):* A measure of the noisiness of the signal. ZCR is commonly used to differentiate voiced and unvoiced speech signals [19], but has also been used for activity recognition in [10].

Frequency domain features:

- *Power Spectral Density (PSD):* Total spectrum power and sub-band power were used as features in [10]. Similar to sub-band power, PSD describes how the power of a signal is distributed over different frequencies.
- *Spectral Centroid:* The centroid is a measure of the spectral shape. High values of the centroid correspond to 'brighter' acoustic structures with more energy in the high frequencies [20].

- *Spectral Roll off:* A measure of the skewness of the spectral shape [19]. It is commonly used in speech recognition and audio classification and has also been used for activity recognition in [10].

Time-Frequency domain features:

- *Discrete Wavelet Transform (DWT):* DWT has been used for emotion recognition from speech [16] and swallowing detection [5], [6]. Delta coefficients of the Coiflet 4 wavelet at decomposition level 3 were shown to be reliable for swallowing detection in [6] and [22].

Cepstral domain features:

- *Mel Frequency Cepstral Coefficients (MFCC):* An important parametric representation commonly used in speech recognition. MFCC has also been used for non-speech recognition such as cough detection [11]. In this study, all 39 coefficients were extracted and used.

D. Classifiers

K-Nearest Neighbor (K-NN) and Naive Bayes classifiers were used in this study. All classifiers were implemented in MATLAB using the Statistical Pattern Recognition Toolbox [24]. Euclidean distance was used to determine the nearest neighbors for K-NN while normal distribution was used for the Naive Bayes classifier. The chosen value of K governs the degree of smoothing; hence, there is an optimum choice for K that is neither too large nor too small [23]. For this reason, odd-number values of K ranging from 1 to 5 were explored and compared to find the optimum K.

III. Data Analysis

Standard information retrieval statistics was used to evaluate performance of the proposed tracheal activity recognition algorithm. A confusion matrix was used to evaluate performance of each classifier on subject-dependent bases. The F_1 score was then calculated for each activity to compare classifier performances:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

A *Recall* performance of 1 means that the event of interest was correctly classified on all occasions while a *Precision* performance of 1 means that there were no false positives. Therefore, the best possible F_1 score is 1.

IV. Results and Discussion

In K-NN classifier, each new data point is assigned to the class having the largest number of representatives from the K nearest points in the training dataset [23]. Therefore, to avoid a tie situation in the majority voting scheme, we focused on odd-number values of K. Classifier results for each activity using K- NN, K = 1, 3 and 5, and Naive Bayes are shown in Fig. 1. The 1-NN classifier achieved the highest F_1 score of 91.2% and 90.2% for chewing and swallowing classification respectively.

A notable property of 1-NN classifier is that in the limit N $\rightarrow \infty$, the error rate is never more than twice the

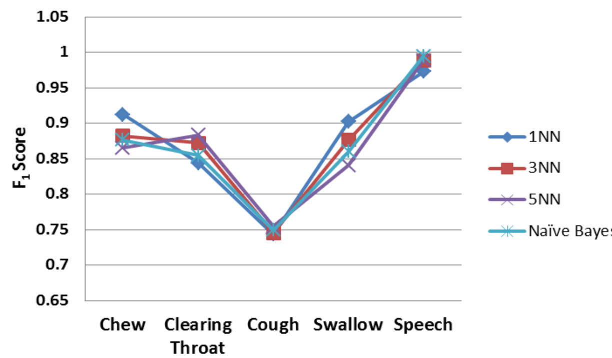

Figure 1. F_1 scores for 1NN, 3NN, 5NN and Naive Bayes Classifiers

minimum achievable error rate of an optimal classifier i.e., one that uses the true class distribution [23].

The 3-NN classifier achieved the highest F_1 score of 87.2% for classification of clearing the throat while 5-NN achieved the highest F_1 score of 75.4% for classification of coughing. Naive Bayes classifier achieved the highest F_1 score of 99.4% for classification of speech.

It is important to note that all classifiers in this study achieved the highest recognition rate for classification of speech. Since the mel frequency scale is a variant of the critical band scale, which is based on perceptual studies and intends to select frequency bands with equal contribution to speech articulation [11], we expect that having MFFCs as a feature for classification contributed to this high classification performance for speech. The ability to classify speech with almost perfect accuracy can mitigate privacy concerns associated with audio-based WHMS by ensuring that the user's conversations can be eliminated before access is provided for further analysis on tracheal events of interest for health monitoring purposes.

Table II shows the confusion matrix obtained for subject 2 using 5-NN. These results show that in the events of misclassification, swallowing events were more commonly confused as chewing while coughing events were more commonly confused as clearing the throat.

Table III shows a summary of classifier performance for

TABLE II. 5-NN Confusion Matrix for Subject 2

		Predicted Class					
		Chew	Clearing Throat	Cough	Swallow	Speech	**Recall (%)**
Actual Class	Chew	12	0	0	1	0	**0.92**
	Clearing Throat	0	9	0	0	0	**1**
	Cough	0	2	9	0	0	**0.81**
	Swallow	2	0	0	8	0	**0.8**
	Speech	0	0	0	0	33	**1**
Precision (%)		**0.85**	**0.81**	**1**	**0.88**	**1**	

TABLE III. SUMMARY OF CLASSIFIER PERFORMANCE

	1-NN	3-NN	5-NN	Naive Bayes
Mean F1 score	0.874	0.873	0.866	0.867

all tracheal activities considered in this study. Each classifiers average performance ranged from 86.6% to 87.4%. These classification results lead us to infer that although a sampling rate of 16 kHz is not sufficient to preserve all important characteristics in tracheal activities, it is sufficient to obtain good classification performance for tracheal activity recognition.

Comparing the results obtained in this work with the results presented by Yatani and Truong in [10] for tracheal activity recognition, our highest mean F_1 score of 87.4% outperforms their highest F_1 score of 79.5% with support vector machine as the classifier. Using 5-NN, in this work we achieved an F_1 score of 86.6% while in [10] they achieved an F_1 score of 75.2% using leave-one-sample-per-participant out. Similarly, in this work we achieved an F_1 score of 86.7% using Naive Bayes while in [10] they achieved an F_1 score of 72.2% using the same classifier. These results lead us to infer that the list of features used in this study may be more comprehensive and therefore more effective than those used in [10]. Although it is important to note that the authors of [10] considered 12 activities in their study which is more extensive than the activities considered here and therefore a limitation of this study.

Since most tracheal activities are non-stationary signals that vary with time, taking advantage of the sequential relationship between activity window frames may improve results obtained in this work. Also, a combination of frame-level and event-level features can allow for short-time and long-time representation of each tracheal activity and allow for even better classification. Results presented in this work are based on classification of clean signals recorded in a relatively quiet environment; further studies are needed for tracheal activity recognition in noisy environments.

V. CONCLUSION AND FUTURE WORK

We have shown that a sub-optimal sampling rate of 16 kHz is sufficient for tracheal activity recognition in wearable health monitoring systems where power consumption is critical. Using a combination of discriminative features from previous work in this area, we obtained average performances in the range of 86.6% and 87.4% for classification of five tracheal activities including chewing, clearing the throat, coughing, swallowing and speech.

Amongst the tracheal events considered, all classifiers achieved the highest recognition for speech classification. This is promising to mitigate privacy concerns associated with neck-worn WHMS interfering with the user's conversations. Comparing K-NN and Naive Bayes classifiers for application in wearable systems, we found that 1-NN, 3-NN, 5-NN and Naive Bayes performed similarly for tracheal activity recognition. Future work includes real-time event detection and activity recognition from continuous recording of daily activities with a wearable neck-wear system in uncontrolled environments.

REFERENCES

[1] A. Pantelopoulos, N.G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. On System, Man and Cyber.,*vol. 40, no. 1, pp. 1-12, Jan. 2010.

[2] C. Poon, M. Wang, P. Bonato, D. Fenstermacher, "Editorial: special issue on health informatics and personalized medicine," *IEEE Trans. Biomed. Eng.,* vol. 60, pp. 143-146, Jan. 2013.

[3] S. Patel, H. Park, P. Bonato, L. Chan, M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Jour. of Neuroengr. and Rehab.,* vol. 9, pp. 21-37, Apr. 2012.

[4] P. Corbishley, E. Rodriguez-Villegas, "Breathing detection: Towards a miniaturized, wearable, battery-operated monitoring system," *IEEE Trans. Biomed. Eng.,* vol. 55, pp. 196-204, Dec. 2007.

[5] E. Sejdic, C.M. Steele, T. Chau, "Classification of penetration-aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects," *IEEE Trans. on Biomed. Eng.,* vol. 60, pp. 1859-1866, July 2013.

[6] T. Olubanjo and M. Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," To be presented at the *IEEE Int. Conf. on Acoustics, Speech and Signal Processing,* May 2014.

[7] A. Yadollahi, Z. Moussavi, "Feature selection for swallowing sounds classification,"*Eng. in Med. and Bio. Conf.,* Aug. 2007, pp. 3172-75.

[8] O. Amft, G. Troster, "Methods for detection and classification of normal swallowing from muscle activation and sound," *Perv. Health Conf.,* 2006, pp. 1-10.

[9] C. Erketin, I. Aydogdu, Y. Seçil, N. Kiylioglu, S. Tarlaci, T. Ozdemirkiran, "Oropharyngeal swallowing in craniocervical dystonia," *J. Neurol Neurosur. Psy.,* vol 73., pp. 406-411, Oct. 2002.

[10] K. Yatani, K.N. Truong, "BodyScope: A wearable acoustic sensor for activity recognition," *UbiComp,* 2012, pp. 341-350.

[11] S. Matos, S. Birring, I.D., Pavord, H. Evans, "Detection of cough signals in continuous audio recordings using hidden Markov models," *IEEE Trans. Biomed Eng.,* vol. 53, pp. 1078-83, June 2006.

[12] P. Lopez-Meyer, O. Makeyev, S. Schuckers, E. Sazonov, E.L. Melanson, M.R. Neuman, "Detection of food intake from swallowing sequences by supervised and unsupervised methods," *Annals of Biomed. Eng.,* vol. 38, pp. 2766-74, Aug. 2010.

[13] K.F. Chung, "Assessment and measurement of cough: The value of new tools," *Pul. Pharm. & Therap.,* vol. 15, pp. 267-72, June 2002.

[14] J. Korpás, J. Sadlonová, M. Vrabec, "Analysis of the cough sound: An overview," *Pulm. Pharmacol.,* vol. 9, pp. 261-8, Oct. 1996.

[15] Y. Xiao, D. Carson, L. Boris, J. Mabary, Z. Lin, F. Nicodème, M. Cuttica, P.J. Kahrilas, J.E. Pandolfino, "The acoustic cough monitoring and manometric profile of cough and throat clearing," *Diseases of the Esophagus,* vol. 27, pp. 5-12, Jan. 2014.

[16] F. Shah, R. Sukumar, B. Anto, "Automatic emotion recognition from speech using artificial neural networks with gender-dependent databases," *Adv. in Comp., Cont. and Tele. Tech.,* 2009, pp. 162-164.

[17] K.P. Truong, D.A. van Leeuwan, "Automatic discrimination between laughter and speech," *Speech Comm.,* vol. 49, pp. 144-158, Feb 2007.

[18] iAsus, "iASUS NT3 Throat Mic," http://www.iasus-concepts.com/

[19] E. Larson, T. Lee, S. Liu, M. Rosenfeld, S. Patel, "Accuracy and privacy preserving cough sensing using a low-cost microphone," *UbiComp,* 2011, pp. 375-384.

[20] S. Theodoridis, K. Koutroumbas, *Pattern Recognition,* 4th Ed. Elsevier 2009.

[21] L. Rudin, S.J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp-259-268, 1992.

[22] 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 Biomed. Eng.*, vol. 57, no. 3, pp. 626–33, Mar. 2010.

[23] C.M. Bishop, *Pattern Recognition and Machine Learning,* Springer, Science+Business Media, LLC, New York, NY, USA 2006.

[24] Statistical Pattern Recognition Toolbox for MATLAB, ONLINE: http://cmp.felk.cvut.cz/cmp/software/stprtool/.