Features of Audio Frequency Content of Respiration to Distinguish Inhalation from Exhalation

Souhail Katti, Federica Aveta, Saurav Basnet, and Douglas E. Dow(✉)

Electrical and Computer Engineering, Wentworth Institute of Technology, Boston, MA 02115, USA

{kattis,avetaf,basnets,dowd}@wit.edu

Abstract. The life-sustaining function of respiration becomes impaired by diseases that occur more with old. A system that monitors the inhalations and exhalations of the respiratory cycle could raise an alert when abnormal patterns or prolonged disruptions are detected. Noninvasive methods are suitable to chronically monitor respiration. Methods include analyzing audio sounds generated during respirations and analyzing changes in the volume of the thorax or abdomen. In casual observations of eupneic breathing, inhalation often sounds different from exhalation, though may be quite similar. One of the challenges for signal processing is to distinguish inhalation from exhalation based on only the audio. The purpose of this study was to find a method of analyzing the audio frequency content that could differentiate the inhalation and exhalation. Volunteer subjects were recruited to record audio during eupneic respiration for analysis. To classify the timing of each inhalation and exhalation, both respiratory sounds and volume changes of the thorax were simultaneously recorded. The audio files were analyzed by Fast Fourier Transform (FFT) to determine the frequency content. Features of the frequency power spectrum were found that appear promising for distinguishing inhalation and exhalation. Such differences could be used to characterize audio respiratory signals and improve the monitoring of individuals at risk for impaired respiratory function.

Keywords: Fourier · FFT · LabView · MATLAB · eupneic · breathing · chronic monitoring

1 Introduction

The population in the world is aging, and the number of people aged over 65 years is growing [1]. Old age is associated with an increased risk of chronic health conditions that may degrade respiration including sleep apnea, respiratory disease and cardiovascular diseases [2, 3]. Sleep disorder breathing (SDB) includes impairments that involve obstructions or narrowing of the upper airway during sleep. SDB is responsible for numerous problems, including fragmented sleep, hypertension and traffic accidents [4].

Y. Chen et al. (Eds.): BICT 2023, LNICST 512, pp. 143–155, 2023.
https://doi.org/10.1007/978-3-031-43135-7_14

Episodes of sleep apnea results in intermittent deregulation of oxygen saturation, which have short and long-term consequences [5]. Sleep apnea increases the risk for high blood pressure, heart problems, type 2 diabetes, metabolic syndrome, liver problems [6].

The diagnosis of SDB typically takes place in clinical settings using intrusive instrumentation, such as spirometer or polysomnography. Due to the clinical setting and procedures, patients may not exhibit their normal pattern of sleep and respiration, thus hampering diagnosis.

Efforts have been made to develop less intrusive methods to monitor respiration during activities of daily living or sleep [7–11]. One method involves recording and analyzing the audio sounds that occur during respiration [12–14]. In one of these studies, audio sounds recorded during respiration were analyzed to classify as normal, wheezes or crackles [12]. Another study analyzed respiratory audio to determine the respiration rate and then determine the duration of exhalations [13]. Another study analyzed respiratory audio to find spectral features to characterize the respiration as normal or abnormal [13]. These studies did not analyze the audio to find features toward distinguishing the behavior of inhalation from the exhalation, which was the purpose of our study.

Factors that may influence the audio pattern include the rate of airflow, occurrence of turbulence in airways, the respiratory rate, individual differences in the anatomy, differences in behavior or condition of the airway, and whether the person is breathing through their nose or mouth. The audio profile of an inhalation may be similar to an exhalation, but may have distinguishing features [14]. For respiratory function to be derived from only analyzing audio signals, the analysis would need to distinguish whether a burst of sound is for an inhalation or an exhalation. Analysis for the timing of respiration would need to distinguish and identify each burst of sound generated during an inhalation and each burst generated during an exhalation.

The purpose of our study was to find features of the audio frequency content of respiration that could be used to distinguish inhalations from exhalations. The findings of this study would be useful for monitoring respiration and health.

2 Methods and Materials

2.1 Recording of Respiratory Activity

In this study audio recordings were made of volunteer subjects while they underwent eupneic cycles of breathing. Of the 13 volunteer subjects for the recording sessions 54% identified as male and 46% as female. Six subjects were aged between 20 to 30 years old, six were between 30 to 60 years old and one subject was 98 years old. Figures 1 and 2 show the demographics of the volunteers.

Since audio sounds generated during inhalations or exhalations may vary in amplitude and frequency content due to the respiration rate, depth of breathing, mode of breathing through the nose or mouth. The subjects were asked to breathe in several different ways during the recording sessions.

Each subject had 3 recordings of 1 min each while they underwent eupneic breathing at a moderate level of intensity. Of the 3 recordings for each subject, one recording was made for each of 3 modes. The first mode had the subject inhale and exhale through

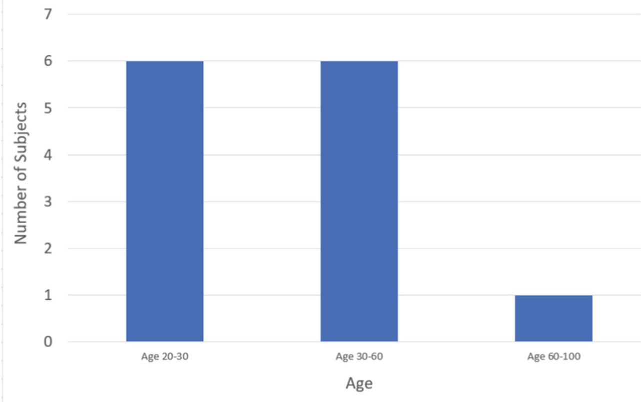


Fig. 1. Ages of the volunteer subjects who underwent recording sessions.

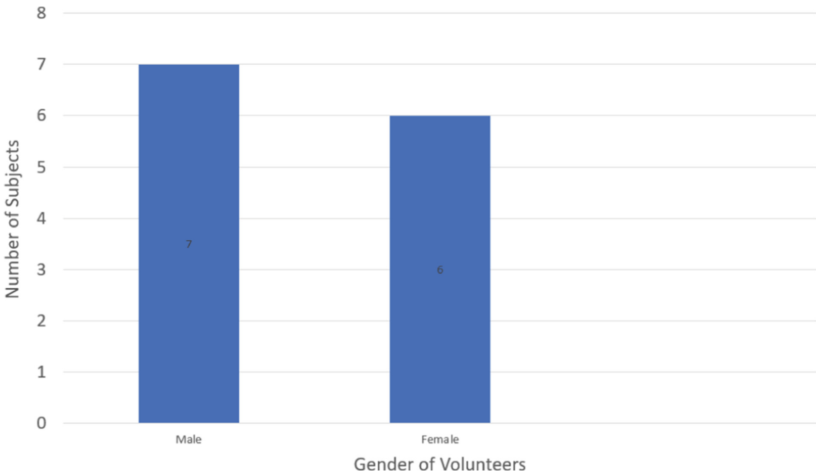


Fig. 2. Gender of the volunteer subjects who underwent recording sessions.

their nose. The second mode had the subject inhale and exhale through their mouth. The third mode had the subject inhale through nose and exhale through mouth.

During the recording sessions, every subject had to undergo 3 recordings while doing eupneic breathing and sitting in a chair. All the subjects had a break of 10 min at the start before they began their recordings to allow for relaxation of their breathing rate prior. Between each of the 3 recordings, the subject had a break of 1 min.

The recording sessions utilized Vernier (Vernier Software and Technology LLC, Beaverton, OR, USA) devices for the microphone, chest belt, and data acquisition module (Fig. 3). The data acquisition module was controlled by a custom LabView (National Instruments, Austin, TX, USA) program running on a Windows (Microsoft, Redmond, WA, USA) laptop personal computer (PC).

Recording sessions were done in one of two indoor environments using a microphone placed at a distance close to the subject's philtrum, the vertical groove between the base of the nose and the upper lip border as shown in Fig. 3.

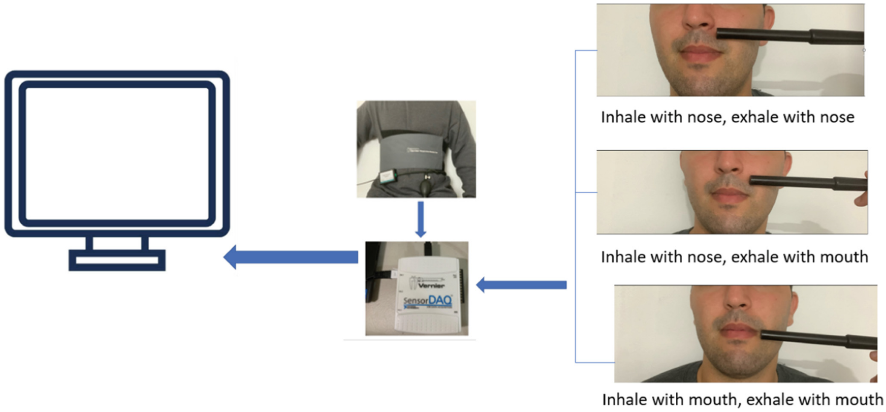


Fig. 3. Setup for the recording sessions. The microphone was placed near the nose and mouth as shown. Vernier devices were used for the microphone, chest belt and data acquisition unit. The recording sessions were controlled by a custom LabView program running on a Windows laptop personal computer (PC).

The microphone was placed on the side between the nose and mouth, depending on which mode was being recorded as shown in Fig. 3. The microphone was connected to channel 1 of the sensor-DAQ and the chest belt was connected to channel 2. The results were displayed in a graph and stored on the PC.

The recording sessions were conducted in an environment with low noise levels in order to better capture the audio sounds generated during respiration. The audio and chest belt recordings were sampled at 24 kHz. A custom LabView program was developed to manage the recording sessions (Fig. 4).

An example of a recording in the LabView program is shown in Fig. 4. The plots show the 60-s recorded signal that was sampled at 24 kHz. The top plot shows the audio signal, and the bottom plot shows the chest belt signals. Cursors in yellow were manually added to this image to help visually correlate the signals for the respiration cycle. The chest belt signal rose during inhalations as the volume of the thorax increased. Then the signal values decreased during exhalations as the volume of the thorax decreased. The peak of the chest belt signal (where the vertical yellow cursors were placed) indicated the transition from inhalation to exhalation. The yellow vertical cursors were placed at approximately the same time in the upper plot showing the audio signal. The audio signal appeared as an almost flat plot line centered at zero of during the quiet periods of no airflow, in between the bursts of inhalation or exhalation. The airflow during inhalations or exhalations appeared as a burst of activity in the audio signal still centered at an amplitude of 0. The smaller audio burst before each yellow cursor corresponded with an inhalation, being at the same time as the rise of the chest belt signal. The larger audio burst after each yellow cursor corresponded with an exhalation, as the chest belt signal fell to lower values.

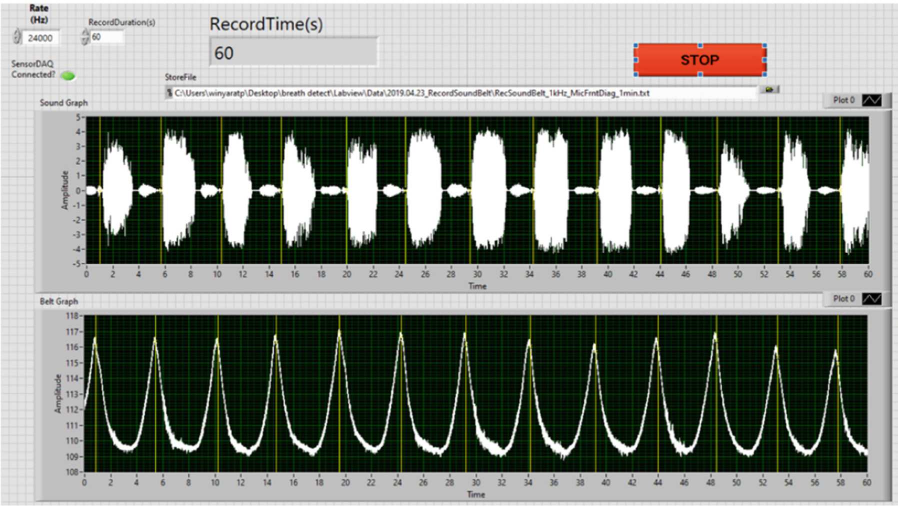


Fig. 4. Recording of breathing in the LabView program that is sampled at 24 kHz with duration of 60 s.

As the purpose of the study was to explore features in the frequency content that could distinguish inhalation from exhalation, the inhalation audio segments were isolated from the exhalation audio segments prior to analysis by FFT. Another Custom LabView program was developed to manually isolate each inhalation and exhalation audio segment. In the example shown in Fig. 5, the white trace was the audio bursts during inhalation or exhalation with a short quiet period in between. The red trace in the background was the chest belt signal with rises indicating inhalations (expanding chest circumference) and falls indicating expirations.

According to Nyquist's theorem, a periodic signal must be sampled at a frequency that is more than twice as high as the signal's highest frequency component to be observed. Thus, the frequency content to be analyzed would be less than 12 kHz. The range of frequencies used in this study was further reduced to the range of 0 to 6 kHz.

The data from the microphone and the chest belt were displayed within the Graphical Interface of the LabView program. The recorded audio and chest belt data were transferred to an Excel sheet and stored as a file.

The cursors were used to manually isolate each inhale or exhale. Cursors were used in the chest belt signal to mark the beginning or end of an inspiration. Then the corresponding audio signal was copied as a list of numbers of the sampled audio values to a growing list of all the inhalations. Each burst of inhalation was concatenated to the end of the growing list. The same was done for each burst of exhalation. Figure 5 shows an example of this. Figure 6A shows an example of concatenated exhales, and Fig. 6B of concatenated inhales. Thus, each audio recording for a volunteer doing eupneic breathing for one of their 3 recordings having different modes would be separated into two audio

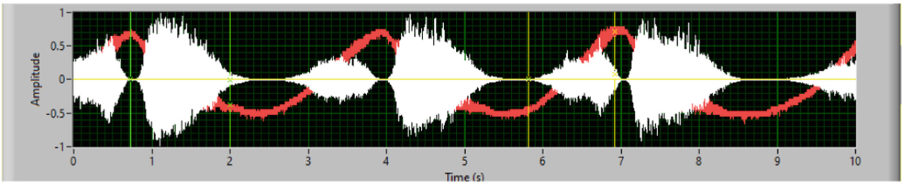


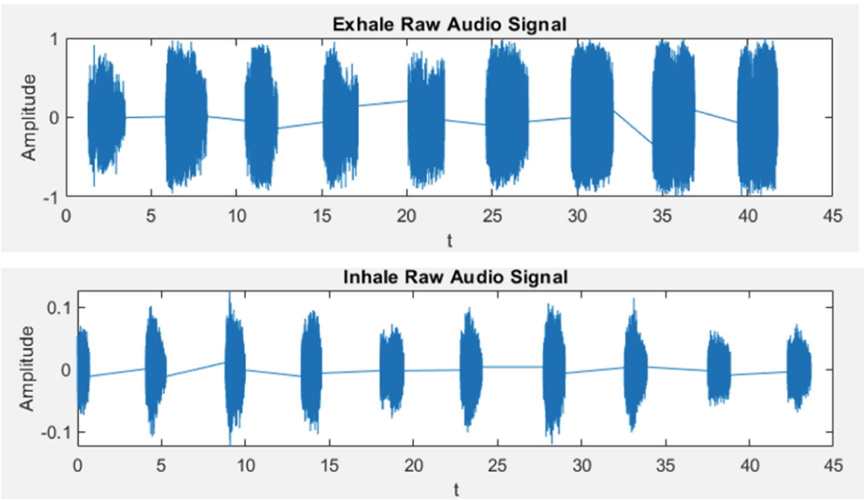
Fig. 5. Example of recorded audio and chest belt signal superimposed together in one graph in the LabView program. The respiration signals were loaded from previously recorded files.

files. One having all the inhalation bursts concatenated one after another, and another file having all the exhalations concatenated one after another. In this way, a frequency analysis could be done for just for the inhalation-related sounds, and another analysis for just the exhalation-related sounds.

2.2 Frequency Domain Analysis

The algorithm developed in this study did a series of steps to form a signal that was called the Smoothed Normalized Spectrum. This signal was the basis for feature extraction to explore features that would help distinguish inhales from exhales. The following section will describe the following steps of the algorithm. The concatenated inhale and exhale signals underwent FFT to form the power spectrum. The power spectrum was normalized for amplitude. The normalized spectrum was smoothed out to form the Smoothed Normalized Spectrum.

A)



B)

Fig. 6. Examples of isolated audio bursts of only exhalations (A) or only inhalations (B). Chest belt recordings of thorax volume changes were used to help move cursors to manually select each exhalation or inhalation. These isolated signals then underwent FFT for analysis of frequency content.

A)

B)

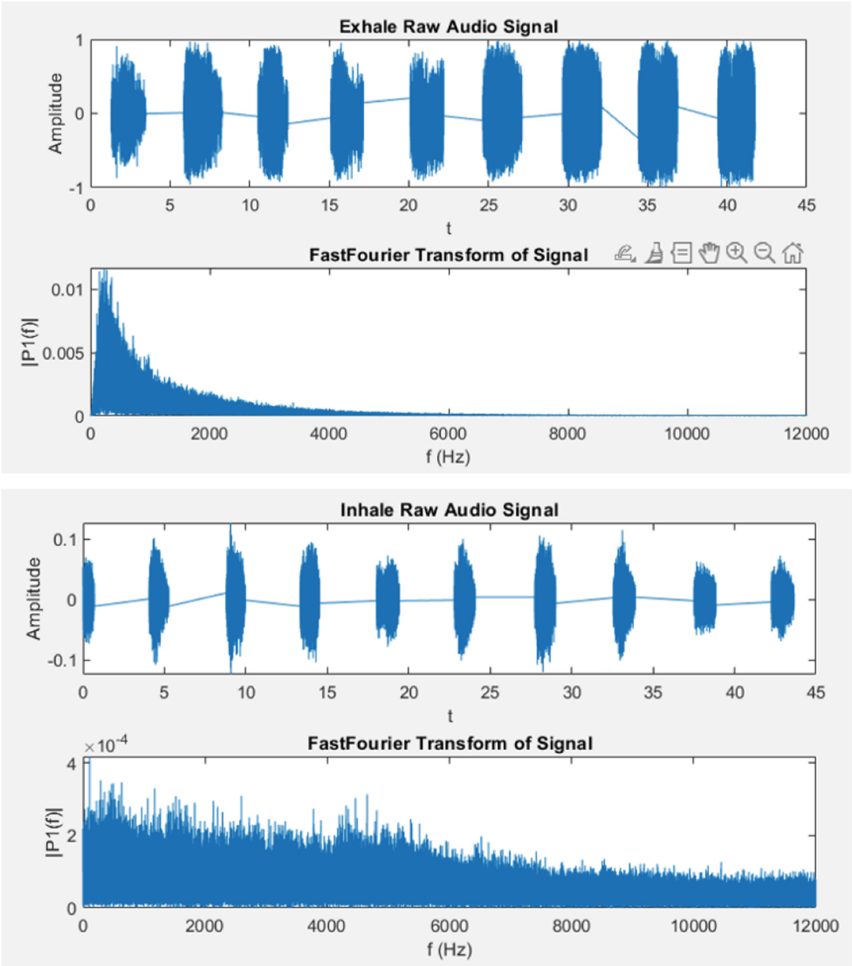


Fig. 7. Example of isolated audio segments that underwent FFT, with the resulting power spectrum. A) shows plots exhalation, and B) shows plots for inhalation. The power spectrum of the exhalation had a different profile than the one for inhalation.

FFT separates a time-domain signal into its distinct spectral components, giving frequency information about the signal. The FFT implements the Discrete Fourier Transformation. The FFT algorithm was implemented within a custom-made MATLAB program on a Windows laptop PC.

The FFT converted the waveform signal in the time domain into the frequency domain. The FFT disassociated the time-based waveform into a number of sinusoidal terms, each with a distinct magnitude, frequency, and phase. The resulting power spectrum was plotted with amplitude of each sinusoidal phase against its frequency as shown in Fig. 7.

To prepare for the exploration of features to find differences between the power spectrum for the inhalations from the power spectrum for the exhalations, the following two steps were done: the power spectrum was 1) normalized for amplitude, and 2) smoothed out. The power spectrum was normalized for amplitude to enhance the comparison of relative frequency content between the inhale and exhale. Without normalization, the relative loudness might have a large influence compared to the frequency content. The MATLAB function "normalize(data)" was utilized.

A Savitzky-Golay filter was used to smooth out the power spectrum. The Savitzky-Golay is a digital filter that has been used to smooth out data on power spectrum data points [15]. This filter uses convolution to the data points with a low-degree polynomial. The result was a smoothed signal (or derivatives of the smoothed signal) at the center of each sub-set. In the plots of Fig. 8, the red trace is the smoothed-out curve for the raw power spectrum shown behind it in blue. The red trace was the Smoothed Normalized Spectrum.

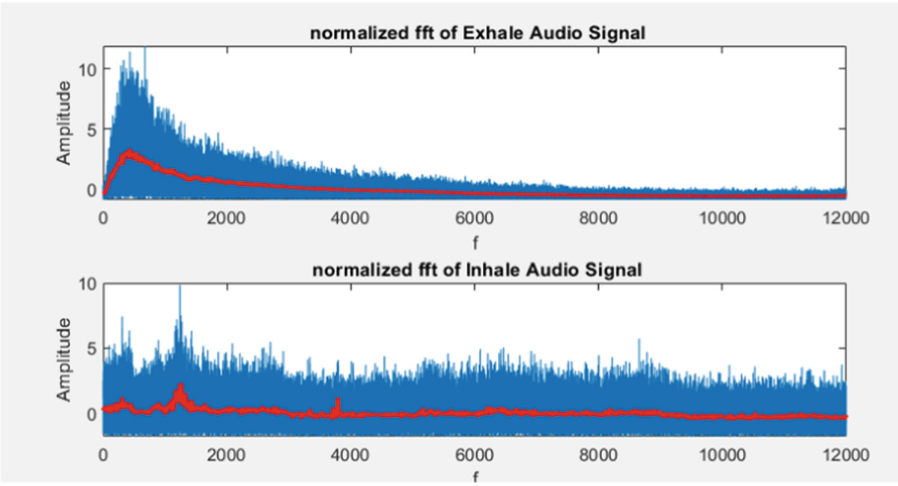


Fig. 8. Normalized FFT of Exhales and Inhales. The plots show the results of the isolated and concatenated exhale audio bursts having undergone FFT, then being normalized for amplitude, followed by being smoothed out with the Savitzky-Golay digital filter. The blue plot shows the normalized power spectrum, and the red plot shows the smoothed-out trace, called the Smoothed Normalized Spectrum. The top plot is for exhalation and the bottom plot is for inhalation.

The Smoothed Normalized Spectrum was used in this study as the basis for feature extraction. Thus, each audio recording of the volunteer subjects breathing resulted in two Smoothed Normalized Spectrums, one for the concatenated inhales and one for the concatenated exhales.

2.3 Feature Extraction

To form features, a curve was fit to each Smoothed Normalized Spectrum, one for inhale and one for exhale. Several types of curves were tried. The cubic polynomial curve appeared able to fit the curve in ways that could distinguish the exhale from the inhale smoothed normalized spectrum. The cubic polynomial was form of this form.

$$Y = Ax^3 + Bx^2 + Cx + D \tag{1}$$

The determined coefficients (A, B, C, D) became the features.

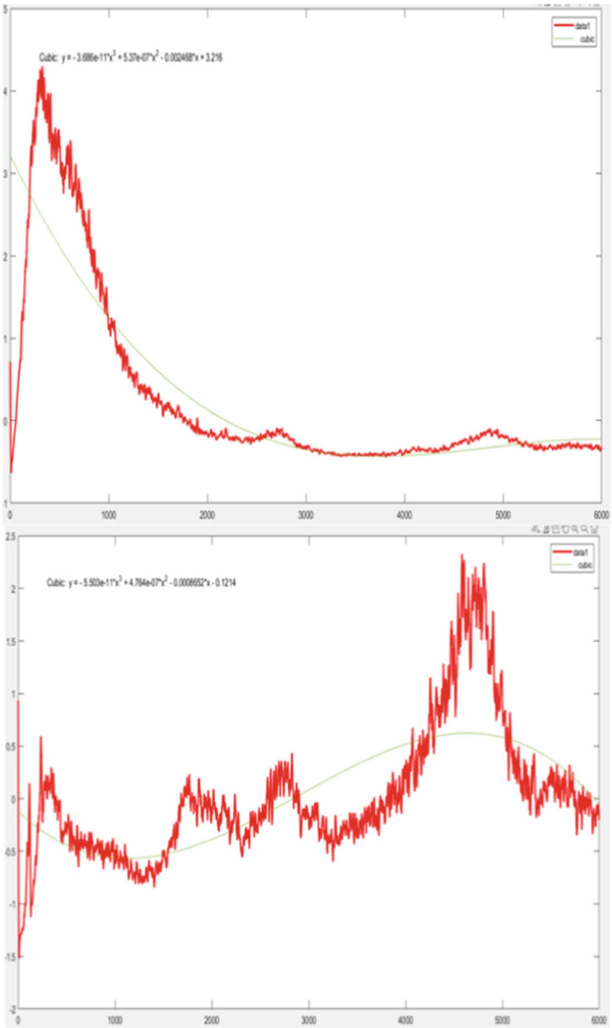


Fig. 9. Fit cubic polynomial to find the features. The red plot shows the normalized, smoothed-out power spectrum of 0–6 kHz. The green plot shows the fitted cubic polynomial. The coefficients (A, B, C, D) of the equation for the fitted cubic polynomial became the features to use for further analysis. The top graph (A) shows the plots for exhalation, and the bottom graph (B) shows the plots for inhalation.

Figure 9 shows an example of the red trace was the normalized, smoothed-out power spectrum. The top graph (Fig. 9A) was derived from the concatenated exhales, and the bottom graph (Fig. 9B) was derived from the concatenated inhales. The green curve was the fitted curve. The coefficients (A, B, C, D) of the fitted polynomial equation were the features used for further analysis. In general, the fitted curve and resulting features for the exhales were quite different than those for the inhales.

3 Results

For each recording of the volunteer subjects having the mode of Inhale from nose and exhale from mouth with a moderate intensity. The feature A, B, C, and D were determined. Plots were made of pairs A-B, A-C, and B-C. Each pair was plotted as a colored dot on the graph: a blue dot for exhale and orange dot for inhale.

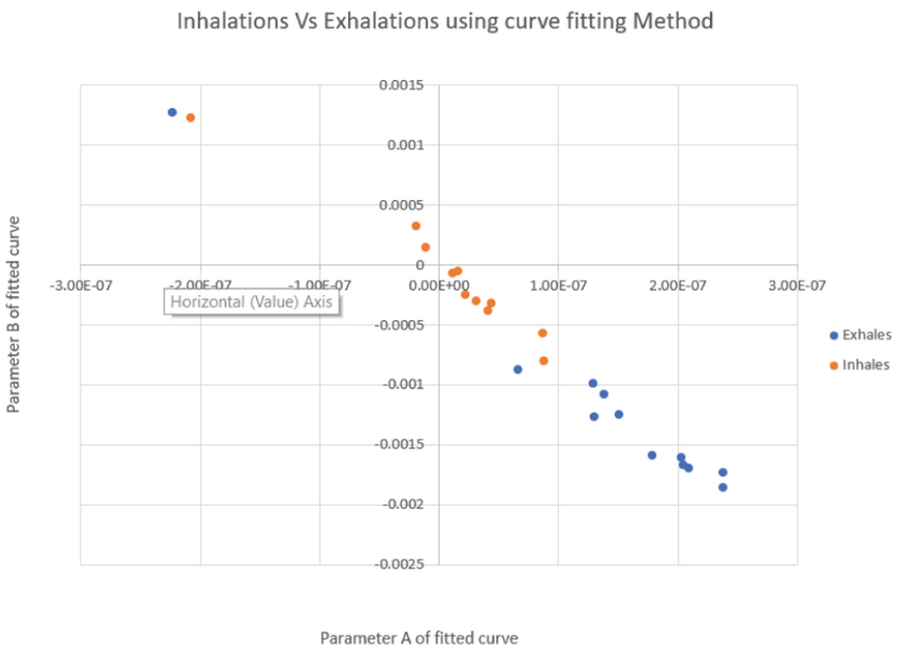


Fig. 10. Plot of features for Coefficients A vs. B.

Figure 10 shows a plot of the feature pairs of A-B. Most of the blue dots (exhale) are separate from the area where the orange dots (inhale) are. Figure 11 shows the feature pairs for A–C. Figure 12 show the feature pairs for B-C. In general, the region of the blue dots (exhales) is separate from the region of the orange dots (inhales).

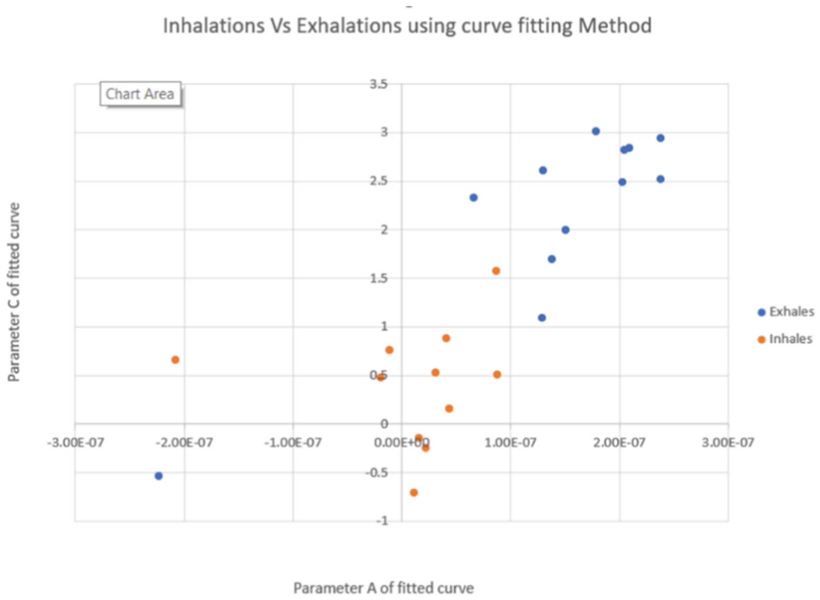


Fig. 11. Plot of features for Coefficients A vs. C.

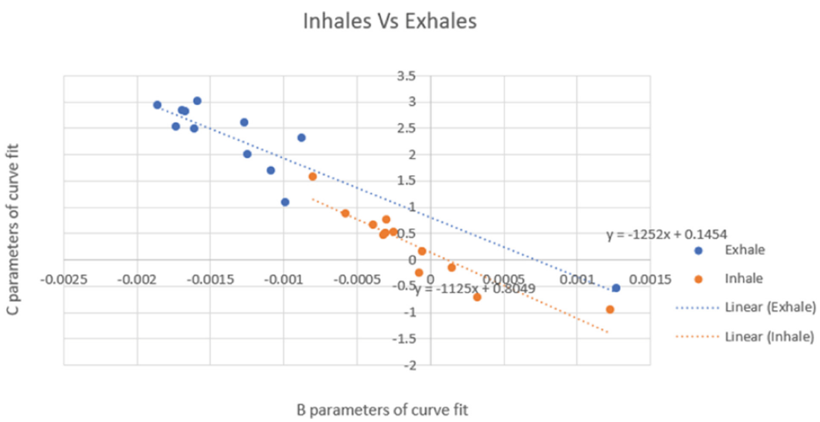


Fig. 12. Plot of features for Coefficients B vs. C.

4 Discussion and Future Directions

An algorithm was developed to determine features for the frequency content of the audio sounds during inhalations and exhalations of breathing. The main steps were converting the breathing signal of concatenated inhales and exhales from the time domain to the frequency domain by applying FFT. The resulting power spectrum was normalized and smoothed to form the Smoothed Normalized Spectrum, which was then fit with a cubic polynomial. The features were the coefficients of the polynomial.

The features were plotted in pairs, and appear to be in separate into regions for the pairs of exhale from inhale. These results look promising in that the features may be able to be used to classify an audio burst as either an exhale or inhale. This would be useful toward monitoring respiration using only audio signals. Further testing would be required to verify these results toward wider application.

Future steps would be to run this analysis on more recordings, such as those having different intensities or modes of respiration. Machine learning could be applied to classify an unknown sound as an exhalation, inhalation or neither. Development of such a method would improve the monitoring of respiration that would enhance diagnosis and treatment of people with chronic respiratory disorders.

References

1. Sanderson, W.C., Scherbov, S.: A new perspective on population aging. Demographic Res. **16**, 27–58 (2007). https://doi.org/10.4054/DemRes.2007.16.2
2. Bloom, D.E., Boersch-Supan, A., McGee, P., Seike, A.: Population aging: facts, challenges, and responses. Benefits Compens. Int. **41**, 22 (2011)
3. Bloom, D.E., Canning, D., Lubet, A.: Global Population Aging: Facts, Challenges, Solutions & Perspectives, vol. 144, pp. 80–92. MIT Press, Daedalus (2015)
4. Zhao, J., Wang, W., Wei, S., Yang, L., Wu, Y., Yan, B.: Fragmented sleep and the prevalence of hypertension in middle-aged and older individuals. Nature Sci. Sleep **13**, 2273–2280 (2021). https://doi.org/10.2147/NSS.S337932
5. Gottlieb, D.J., Punjabi, N.M.: Diagnosis and management of obstructive sleep apnea: a review. JAMA **323**(14), 1389 (2020). https://doi.org/10.1001/jama.2020.3514
6. Azagra-Calero, E., Espinar-Escalona, E., Barrera-Mora, J.M., Llamas-Carreras, J.M., Solano-Reina, E.: Obstructive sleep apnea syndrome (OSAS). Review of the literature. Medicina Oral Patología Oral y Cirugia Bucal e925–e929 (2012). https://doi.org/10.4317/medoral.17706
7. Reyes, B., Reljin, N., Chon, Ki.: Tracheal sounds acquisition using smartphones. Sensors **14**(8), 13830–13850 (2014). https://doi.org/10.3390/s140813830
8. Lu, X., Coste, C.A., Nierat, M.-C., Renaux, S., Similowski, T., Guiraud, D.: Respiratory monitoring based on tracheal sounds: continuous time-frequency processing of the phonospirogram combined with phonocardiogram-derived respiration. Sensors **21**(1), 99 (2020). https://doi.org/10.3390/s21010099
9. Bethel, S.J., Joslin, C.T., Shepherd, B.S., Martel, J.M., Dow, D.E.: Wearable at-home recording system for sleep apnea. Presented at the 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), pp. 149–152. IEEE (2015)
10. Dow, D.E., Petrilli, A.M., Mantilla, C.B., Zhan, W.-Z., Sieck, G.C.: Electromyogram-triggered inspiratory event detection algorithm. Presented at the The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems, pp. 789–794. IEEE (2012)
11. Dow, D.E., Horiguchi, Y., Hirai, Y., Hayashi, I.: Monitoring of respiratory cycles utilizing sensors on sleeping mat. Presented at the ASME International Mechanical Engineering Congress and Exposition, vol. 58363:V003T04A011 (2017)
12. Meza, C.A.G., del Hoyo Ontiveros, J.A., Lopez-Meyer, P.: Classification of respiration sounds using deep pre-trained audio embeddings. Presented at the 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI), pp. 1–5. IEEE (2021)
13. Doheny, E.P., et al..: Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones. Biomed. Signal Process. Control **80**, 104318 (2023). https://doi.org/10.1016/j.bspc.2022.104318

14. Gavriely, N., Palti, Y., Alroy, G.: Spectral characteristics of normal breath sounds. J. Appl. Physiol. **50**, 307–314 (1981)
15. Xu, P., Jia, Y., Jiang, M.: Blind audio source separation based on a new system model and the savitzky-golay filter. J. Electric. Eng. **72**(3), 208–212 (2021). https://doi.org/10.2478/jee-2021-0029