Breathing Patterns in Speech : Discovering Markers of Health

Gauri Deshpande[1,2]

[1]Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
[2]TCS Research Pune, India

gauri.deshpande@informatik.uni-augsburg.de, gauri1.d@tcs.com

Abstract

Breathing patterns—the signals generated during respiration—are intricately connected to speech production. The respiratory organs contribute to the production of speech signals as well, and hence both breathing patterns and speech have an impact on each other. In this Ph.D. work, time-domain speech representation, coupled with phase-domain decomposed speech components, is investigated as a carrier of respiratory information. This feature set and a novel long-short-term-memory (LSTM)-based deep architecture are introduced to extract the breathing patterns from the speech signals. The speech-breathing data from 100 healthy college going students, while they read a phonetically balanced text is collected to build this model. This Ph.D. work also explores the impact of breathing pattern categories on the performance of the deep model as well as the variability of model performance observed across the 100 speakers. Furthermore, the pre-trained model is utilised to extract breathing patterns from speech data labelled with respiratory disorders and human-confidence levels. The resulting speech-derived breathing patterns serve as a pioneering feature set for detecting respiratory disorders and gauging human-confidence levels.

Index Terms: digital-health, health informatics, affective computing, speech-breathing parameters

1. Research Problem and Motivation

Breathing pattern analysis has found its significance in the diagnosis of the physical and mental well-being of individuals. Several studies are reviewed in [1] and [2] on breathing pattern analysis for the detection of respiratory disorders, including COVID-19. Similarly, psychological states and the breathing process have an impact on each other. In [3], individuals with high self-rated apprehension are found to have more pauses, longer breath groups, and more interjections in their speech. This explains the importance of analysing breathing patterns to understand the physiological and psychological aspects of human health.

The existing techniques to measure breathing patterns include 1) visual inspection, 2) using a spirometer, 3) impedance pneumography, 4) mercury-in-silastic strain gauges, 5) using magnetometers, and 6) respiratory inductive plethysmography (RIP). Visual inspection is the simplest of all, but it is prone to errors. All other techniques require a measurement instrument connected to the individual under observation. For example, in RIP, a transducer is connected over the chest area to convert the changes in lung volume into digital breathing patterns. The acquisition of such patterns to enable further analysis of the signals requires a respiratory belt along with a data acquisition unit. Such conventional transducers used for capturing respiration-related information are intrusive and rely on expensive instruments. The invasive nature of these mechanisms can impede the accurate analysis of breathing patterns affected by psychological states. Similarly, for investigating physiological disorders associated with the respiratory process, infected individuals are required to visit lab setups equipped with sensor-based instruments to analyse their breathing patterns. As an individual needs to visit a clinic for an inspection of the breathing pattern, this is usually done only after the difficulty in breathing becomes severe. The intrusive, expensive, and error-prone mechanisms of capturing breathing patterns present the need for a non-intrusive modality, such as speech, that provides breathing information even outside of a clinical or lab setup.

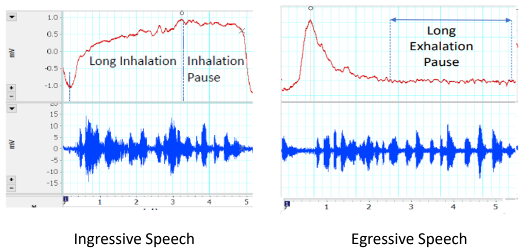

Figure 1: *Two broad categories of the speech-breathing patterns: speech during inhalation called ingressive and speech during exhalation called egressive speech-breathing.*

2. Contributions

The main contributions of this Ph.D. work are as follows:

1. Corpus of speech-breathing data from 100 healthy college-going students.
2. A deep network called SBreathNet is trained with the data from 100 speakers to extract breathing patterns from speech signals.
3. We augment the understanding of speech-breathing patterns. As seen in Figure 1, the right breathing pattern with a sudden inhalation peak followed by exhalation is called egressive speech-breathing. Here, the speech production happens during exhalation. The left side of Figure 1 shows the breathing pattern with a longer inhalation and a sudden drop during exhalation. This is called an ingressive breathing pattern [4]. Here, the speech production happens during inhalation. We introduce the impact of ingressive patterns on the model's performance.
4. Presenting speech-derived breathing patterns (SDBPs) as a novel feature set for the detection of physiological and psychological disorders.

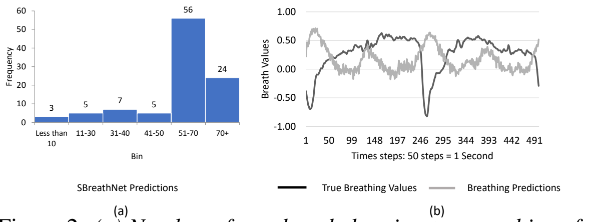

Figure 2: *(a) Number of speakers belonging to seven bins of r-value performance. (b) Breathing predictions for an ingressive speaker.*

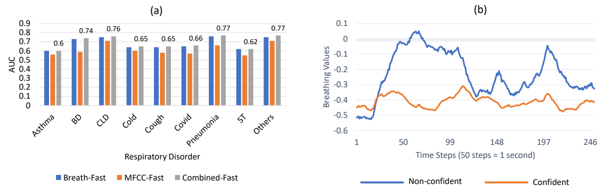

Figure 3: *(a) Detection of a respiratory disorder from the healthy class speech samples while the subjects count the digits with fast speed; measured using the metric area under the curve (AUC). Breath-fast: results using SDBPs as the feature set; MFCC-fast: results using MFCCs as the feature set; and Combined-fast: results with both the SDBPs and MFCCs combined together as a feature set on the counting-fast speech samples. (b) The average breathing patterns for the confident and non-confident classes.*

3. Results

This section presents the results obtained with the SBreathNet architecture for predicting breathing patterns. The performance is calculated using Pearson's correlation coefficient (r-value) and breaths-per-minute error (BPME) as metrics. Further inferences from the pre-trained model, SBreathNet, on two datasets of speech: 1) labelled with respiratory disorders (Coswara [5]) and 2) labelled with human confidence levels (self-built dataset of 51 speakers [6]) are used for the detection of respiratory disorders and human confidence levels, respectively.

3.1. SBreathNet Performance

SBreathNet extracts breathing patterns with an average r-value of 0.61 and a BPME of 2.50. The BPME is found to range between 0.3 and 7.5. The change in BPME across the speakers is not synchronised with the r-value exhibited by them. Speakers with a negative r-value of -0.40 and -0.21 have BPMEs of 3 and 2.1, respectively. This shows that SBreathNet captures the breathing event equally well for speakers with low r-values.

As seen in Figure 2 (a), the number of speakers having an r-value above 0.50 is 80. Similarly, $90\,\%$ speakers have BMPE less than 4. SBreathNet can extract breathing patterns with an r-value above 0.50 for $80\,\%$ speakers and a BPME below 4 for $90\,\%$ speakers. It is observed that 14 out of 20 ($70\,\%$) of the speakers exhibiting an r-value below 0.50 (low-performers) are ingressives. The average r-value of egressive speakers is 0.65 and that of ingressive speakers is 0.37 using SBreathNet predictions. These results suggest that ingressiveness has a considerable impact on the model's performance. Figure 2 (b) shows the 10 s prediction for an ingressive speaker where the breathing events are correctly identified, resulting in a BPME of only 1.2. However, the breathing pattern prediction is inverted such that the inhalation and inhalation pause of true breathing patterns are predicted as expiration for the corresponding time slot. This explains the absence of synchronisation between the r-value and the BPME across the speakers.

SBreathNet performs equivalently well on the speech-breathing dataset of the ComParE challenge organised at Interspeech 2020 [7] when compared with the performance reported by the winners of this challenge [8].

3.2. Applications of speech-derived breathing patterns

The SDBPs obtained using SBreathNet as the pre-built model are used for the detection of respiratory disorders and human-confidence levels using Coswara and a self-built dataset of 51 speakers, respectively. The analysis outcome of SDBPs compared to MFCCs is shown in Figure 3 (a) for the detection of respiratory disorders in individuals counting $1-10$ digits at a fast speed. SDBPs perform better than MFCCs for nine disorders (asthma, breathing difficulty (BD), chronic lung disorder (CLD), cold, cough, COVID-19, pneumonia, sore throat (ST), and others) and perform even better when they are combined together. Figure 3 (b) shows the analysis outcome of SDBPs for detecting confidence levels. An average breathing pattern of the confident and non-confident classes is shown. Empirically, an average area-under-the-curve (AUC) of $75.6\,\%$ is achieved in detecting non-confident individuals from confident ones. This outcome is around $5\,\%$ and $8\,\%$ higher than that exhibited by autoencoder-based representation and MFCCs, respectively.

4. Challenges

It is observed from the results that extracting breathing patterns for ingressive speech is difficult. To collect more data belonging to the ingressive class, we need to understand such speaker characteristics. We asked further questions to the ingressive speakers to understand them better from psychological and physiological perspectives. For all of them, no uniformity of any kind of symptom is observed. Hence, collecting more data from ingressive speakers is a challenge. Likewise, breathing patterns get affected by many factors, and controlling these impacting factors while collecting data for a specific one is a challenge. Distinguishing between similar dysfunctions such as asthma and BD requires a deeper understanding of the domain. Another challenge is to identify the deep exhalations, as they do not produce any sound. Hence, the current model gets confused between the breathing pause and the deep exhales.

5. Future Work

In future work, we will focus on collecting more data and identifying ingressive breathing patterns accurately. Given the nature of breathing patterns as features, their usability would reflect more when applied to lengthy speech samples of duration greater than 30 s. In such cases, the lung volume capacity variations exhibited over a period of time will further strengthen the analysis. We will work with longer speech samples of subjects with respiratory disorders to further reinforce our analysis. Similar to detecting human-confidence levels, we also intend to extend this analysis to other psychological parameters such as emotions, stress, and anxiety. With more data, we will also focus on understanding the markers from speech that help us understand the deep exhalations.

6. References

[1] G. Deshpande, A. Batliner, and B. W. Schuller, “Ai-based human audio processing for covid-19: A comprehensive overview,” *Pattern recognition*, vol. 122, p. 108289, 2022.

[2] G. Deshpande and B. Schuller, “An overview on audio, signal, speech, & language processing for covid-19,” *arXiv preprint arXiv:2005.08579*, 2020.

[3] “Acoustic characteristics of public speaking: Anxiety and practice effects,” *Speech Communication*, vol. 53, no. 6, pp. 867–876, 2011.

[4] R. Eklund, “Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech,” *Journal of the International Phonetic Association*, vol. 38, no. 3, pp. 235–324, 2008.

[5] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, N. R., P. K. Ghosh, and S. Ganapathy, “Coswara - a database of breathing, cough, and voice sounds for covid-19 diagnosis,” in *Proceedings of the 21st Annual Conference of the International Speech Communication Association*. Shanghai, China: INTERSPEECH Communication Association (ISCA), 2020, pp. 4811–4815.

[6] G. Deshpande, G. Yagna, S. Patel, and B. Schuller, “Applying speech derived breathing patterns to automatically classify human confidence,” in *Proceedings of the 31st European Signal Processing Conference (EUSIPCO)*. Helsinky, Finland: EURASIP, 2023.

[7] B. W. Schuller, A. Batliner, C. Bergler, E.-M. Messner, A. Hamilton, S. Amiriparian, A. Baird, G. Rizos, M. Schmitt, L. Stappen, H. Baumeister, A. D. MacIntyre, and S. Hantke, “The interspeech 2020 computational paralinguistics challenge: Elderly emotion, breathing & masks,” in *Proceedings of the 21st Annual Conference of the International Speech Communication Association*. Shanghai, China: INTERSPEECH Communication Association (ISCA), 2020, pp. 2042–2046.

[8] M. Markitantov, D. Dresvyanskiy, D. Mamontov, H. Kaya, W. Minker, and A. Karpov, “Ensembling end-to-end deep models for computational paralinguistics tasks: Compare 2020 mask and breathing sub-challenges,” in *Proceedings of the 21st Annual Conference of the International Speech Communication Association*. Shanghai, China: INTERSPEECH Communication Association (ISCA), 2020, pp. 2072–2076.