DETECTION OF COPD EXACERBATION FROM SPEECH: COMPARISON OF ACOUSTIC FEATURES AND DEEP LEARNING BASED SPEECH BREATHING MODELS

Venkata Srikanth Nallanthighal[1,2], *Aki Härmä*[1], *and Helmer Strik*[2]

[1] Philips Research, Eindhoven, The Netherlands
[2] Centre for Language Studies (CLS), Radboud University Nijmegen

ABSTRACT

Respiration is a primary process involved in speech production. We can often hear if a person has respiratory difficulty, thus making speech a good pathological indicator for respiratory conditions. This is more relevant to conditions like chronic obstructive pulmonary disease (COPD). Patients with COPD suffer from voice changes with respect to the healthy population. Medical professionals observe that the speech of COPD patients during stable periods differs from the speech during exacerbation. In this paper, we investigate this detection of COPD exacerbation from speech in three approaches: acoustic features identification using a statistical approach, low-level descriptive features with classification, and speech breathing models based on deep learning architectures to estimate the patients' breathing rate. Our analysis indicates that each of these approaches indeed results in a clear distinction of speech during exacerbation and stable periods of COPD.

Index Terms— COPD exacerbation, speech technology, statistical approach, deep neural networks, speech breathing

1. INTRODUCTION

The importance of respiratory sensing needs little justification, especially in-home monitoring of patients suffering from respiratory conditions. The COVID-19 pandemic demonstrated the necessity of remote digital health assessment tools for telehealth services. This is particularly pertinent for elderly and vulnerable populations who already have a chronic disease. We have demonstrated recently that it is possible to use neural networks to estimate the breathing parameters from a speech audio signal [1, 2, 3]. The development of acoustic sensing technology is vital for telehealth call services especially for patients with respiratory symptoms [4].

COPD is an umbrella term used to describe progressive lung diseases characterized by airflow limitation. The prevalence of COPD worldwide is estimated at roughly 12%, but the percentage differs significantly between different subgroups [5]. Most COPD patients suffer from stage II COPD (70%), while stage I, III and IV make up 16%, 11% and 3% of the COPD population. The four most significant predictors of COPD are years and intensity of smoking, age, sex, and BMI. Most patients suffering from COPD are smokers with a low BMI, over 50 years old, and male [5]. Taking into account the three million annual deaths globally, COPD is currently the fourth leading cause of death in high-income countries, and it is expected to be the third leading cause in 2020 due to a higher life expectancy and increasing air pollution [6]. Exacerbations are often recognized at a late stage, which delays the treatment [7]. It is important to diagnose the patient quickly and correctly because early treatment with Prednisone shortens the duration and the seriousness of the exacerbation and could prevent COPD patients from being hospitalized[8]. Moreover, many patients do not fully recover from an exacerbation.

Medical professionals from the Department of Pulmonary Diseases of Dekkerswald (Radboud University Medical Centre) have indicated that they can estimate the condition of a patient who suffers from COPD by listening to their speech [9, 10]. The medical professionals observed that the speech of COPD patients during stable periods differs from the speech during exacerbations. The information regarding the acoustic speech characteristics of COPD patients could be of great value if the professional's observations are correct.

This study is proposed to justify the above professional's observations. We investigate the detection of COPD in three approaches, from basic voice-based feature identification to deep learning models detecting the breathing rate of the patients. The structure of the paper is as follows. Section 2 describes the experiment, data, and three approaches. In section 3, we introduce the methodology behind each of the three approaches in detail. Section 4 presents results and comparisons, and Section 5 concludes our observation.

2. APPROACH

We present the following three approaches for detecting COPD exacerbation from speech.

1. We investigate the acoustic features and identify the relevant features for the exacerbation detection using a statistical approach.

2. We use traditional openSMILE toolkit [11] for low-

level descriptive features and perform SVM classification from 20s snippets of speech.

3. We use speech breathing models developed by the authors using a different speech respiratory database to compare breathing rates of both stable and exacerbation conditions.

2.1. Experiment and Dataset

The voices of 11 native speakers of Dutch (7 males, 4 females) were recorded twice in a treatment room of the lung department of the Radboud University Medical Centre from August 2016 until April 2017. All participants had officially been diagnosed with COPD. The participants were hospitalized due to an exacerbation and they had to stay in the hospital for 2 to 23 days (M = 8.82, SD = 6.11). The participants were requested to participate in the experiment after first receiving urgent care for their exacerbation. Patients suffering from additional lung diseases were excluded from participation. Detailed participant information, such as the COPD stage, number of previous exacerbations, age, and years since diagnosis, is missing due to the retrospective nature of the research. Furthermore, 5 recordings of 5 healthy, adult speakers (4 males, 1 female) from the Corpus Gesproken Nederlands (CGN) were selected in order to compare (characteristics of) their speech with the speech of the COPD patients. These 5 recordings were obtained between 1998 and 2004.

The recordings were made using two Relitech microphones, and each recording consisted of two parts: a sustained vowel and read speech. To simulate more natural situations with control over the experiment, we use read speech [12]. All recordings from each patient are the storytelling of 'De Koning' (Bomans, 1946) or 'Papa en Marloes' [13]. The length of the recording is around 4 minutes. The recordings of 11 COPD patients reading aloud of storytelling is collected when the patient has exacerbation and also when the same patient is stable after treatment. Hence it is a total of 11 COPD stable recordings and 11 COPD exacerbation recordings.

3. METHODS

3.1. Statistical approach for feature selection

To analyze how various acoustic measures differed for COPD patients in exacerbation and in a stable condition, several one-way analyses of variance (ANOVAs) were conducted. The acoustic features investigated for this approach are indicated in Table 1 [14]. A one-way multivariate analysis of variance (MANOVA) is used to investigate the effect of condition on speech features. A MANOVA extends the traditional analysis of variance (ANOVA) possibilities as multiple dependent variables can be simultaneously determined while the joint error rate does not increase [15]. Therefore, the risk of rejecting a true null hypothesis (type I error) is minimized. A MANOVA was preferred over ANOVA because the current analysis included multiple dependent variables which were moderately correlated. The analysis is conducted on the story reading recordings of COPD patients collected during exacerbation and in stable condition, as described in section 2.1.

Table 1: Acoustic features and their descriptions used in statistical approach

Feature	***Description***
Mean intensity(dB)	Perceived loudness in dB
Mean frequency(Hz)	Perceived pitch in Hz
Pitch variability (Hz)	Range in variation in level and extent of pitch in Hz
Mean center (Hz) of gravity	Weighted mean frequency in Hz
Formants	F1, F2, F3 in Hz
Speaking rate	Speech tempo (in syllables per second)
Syllables per breath group	The number of syllables produced on one breath
Jitter (%)	Vocal fold frequency variability (frequency perturbation)
Jitter ppq5	Five-point perturbation quotient
Shimmer	Variability of the peak-to-peak amplitude
Shimmer apq3	3-point amplitude perturbation quotient
Shimmer apq5	5-point amplitude perturbation quotient
HNR	Proportion of the harmonic sound to noise in the voice

3.2. openSMILE features with SVM for COPD exacerbation detection

A state-of-art feature set is obtained from the speech recordings using the openSMILE toolkit [11], the so-called COMPARE feature set. It calculates 6373 acoustic features using diverse functionals over low-level descriptor (LLD) contours. A full description can be found in [16].

The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) contains a rudimentary ideal set intended to be applied in different fields of automatic voice analysis, such as paralinguistic or clinical speech analysis. Researchers delivered the feature set as a minimalistic group of voice parameters in interdisciplinary work. The principal objective was to provide a common baseline for future evaluation of systems, eliminating the differences in internal parameters. This sets a standard for implementing the same features across the research groups. The GeMAPS consists of 62 parameters originated from 18 LLD descriptors divided into the following categories:

1. frequency: pitch and its jitter, and first 3 formants;

2. energy/amplitude: shimmer, loudness and harmonics-tonoise ratio;

3. spectral: different ratios and indices showing relations between energy bands and peaks.

The extended GeMAPS (eGeMAPS) is an alternative version which adds 26 extra parameters to the basic set. They are obtained from cepstral coefficients along with dynamic information. The implementation of GeMAPS is publicly available with the openSMILE toolkit. The system provided for the classification of speech into COPD exacerbation or stable conditions consists of Support Vector Machine (SVM) with linear kernels and epsilon-insensitive hinge loss.

3.3. Speech breathing model for breathing rate detection

Our earlier work established that breathing signal can be estimated from speech using deep learning architectures. We collected a speech and breathing signal database consisting of 40 participants and trained neural network models to establish the relation between speech and respiration. An exhaustive study is presented in [1, 2, 3].

For breathing signal estimation using Long short-term memory recurrent neural network (LSTM-RNN) architecture, log Mel spectrogram is a preferred input spectral feature representation of speech signal [1] and fixed time window of 4s is optimum for training the model. Speech signals of the fixed time window length of 4s were processed by a pre-emphasis filter to spectrally flatten speech signals and boost higher frequencies. Spectral features are extracted from these windows using log Mel spectrograms [17]. As estimating breathing signal from speech using LSTM-RNN is a regression problem, we use the following two metrics for evaluation and comparision: Correlation and mean squared error(MSE) of estimated breathing signal and the actual respiratory sensor signal. The model with higher correlation and lesser MSE would give the best estimation. From these estimated breathing signals, breathing events are identified d to get breathing rate which is an important respiratory parameter to detect the pathological condition of a person. The estimated breathing rate is used as a parameter to observe the difference between stable and exacerbation conditions for each COPD subject.

4. RESULTS

4.1. Statistical approach for feature selection

It is observed from Table 2 that the following features have a p-value less than 0.05, which implies that they are statistically significant. Shimmer and its perturbation quotients play an important role in detecting exacerbation. It is also observed that the shimmer and its perturbation quotients are significantly increased in exacerbation conditions. This could be because of insufficient and unstable pulmonary support for a speech during exacerbation. Syllables per breath group could also be used as an important feature for detecting exacerbation. It is observed that the syllables per breath is significantly less in exacerbation conditions which is quite intuitive.

Table 2: Overview of the means, standard deviations, F-value and p-value (in exacerbation vs. stable) for all features.

Parameter	Mean(SD) Exa.	Mean(SD) stable	F	p
Mean intensity(dB)	64.51 (7.12)	63.38 (7.82)	0.119	.721
Mean frequency(Hz))	153.78 (30.12))	190.52 (90.12)	1.23	.271
Pitch variability(Hz)	586.53 (238.99)	720.2 (508.30)	0.616	.444
Mean center of gravity (Hz)	439.12 (152.32)	504.34 (203.60)	0.592	0.453
F1(Hz)	529 (110.2)	554 (95.2)	0.96	.34
F2(Hz)	1702 (110.4)	1731 (91.2)	0.382	.545
F3(Hz)	2860 (150.2	2872 (162.3)	0.25	.885
Speaking rate (syllables/sec)	12.78 (3.18)	17.43 ()5.64	4.160	.064
Syllables per breath group	9.50 (1.70)	12.78 (3.18)	7.413	**.015**
Jitter (%)	1.45 (0.96)	0.84 (0.42)	3.310	.084
Jitter ppq5	0.88 (152.3)	0.45 (203.6)	3.348	.089
Shimmer	8.72 (5.12)	4.39 (1.36)	5.62	**.034**
Shimmer apq3	4.52 (2.13)	1.97 (0.78)	4.19	**.044**
Shimmer apq5	5.23 (3.36)	2.48 (0.81)	5.693	**.029**
HNR	15.09 (5.11)	19.46 (2.67)	5.424	.036

4.2. openSMILE features with SVM for COPD exacerbation detection

openSMILE toolkit [11] is used for extracting acoustic features. Speech recordings are divided into signals of fixed window length of 20-second snippets and given as input for extracting features. Traditional Support Vector Machine (SVM) with linear kernels and epsilon-insensitive hinge loss is used for this classification task. The classification framework is designed by training the model on 8 subjects and by testing on 3 subjects. Cross-validation is performed, and results are shown in Table 3.

Using the ComPare2016 feature set of 6373 features, we achieve an accuracy of 75.12% and sensitivity of 0.85. Comparable accuracies are obtained for GeMAPS and eGeMAPS feature set with significantly less features (62 and 88, respectively). This suggests that the eGeMAPS feature set is relevant in exacerbation detection and thus saves computational time and power. This study also proposes an interesting fact that these models perform blind detection on entirely new subjects for exacerbation detection. As our training and testing data split is based on subject , these models are tested on subjects who aren't included in training set.

Table 3: Accuracy and sensitivity for openSMILE features with SVM classification for COPD exacerbation detection

Low level descriptor	Accuracy	Sensitivity
ComPare2016(6373)	75.12%	0.85
GeMAPSv01b (62)	67.47%	0.88
eGeMAPSv02(88)	69.86%	0.84

4.3. Speech breathing model for breathing rate detection

The pre-trained deep learning model using RNN architecture is used to compute the estimated breathing signal for each of the recordings. From these estimated respiratory signals, respiratory breathing events are identified using an Automatic Multiscale Peak Detection Algorithm (AMPD) [18]. The breathing signal is analyzed to get breathing rate which is an important respiratory parameter to detect the pathological condition of a person. *Breathing rate* is an average number of breaths per minute [19].

Breathing rates of each COPD patient is estimated in both stable and exacerbation conditions. These breathing rates are compared with the healthy subjects, as shown in Figure 1. We observe a clear distinction in the breathing rates for healthy subjects and subjects with COPD. Also, we observe that the average breathing rate for stable condition is 10.12 breaths per minute, which is much lesser than the breathing rate for exacerbation condition, 11.84 breaths per minute.

5. CONCLUSIONS

We present an extensive study on methods for using speech as a modality for detecting COPD exacerbation from speech by investigating three approaches. The conclusions are as follows.

1. The statistical approach suggests that the acoustic feature that is most relevant for classification is shimmer and its perturbation quotients. It is also observed that the shimmer and its perturbation quotients are significantly increased in exacerbation conditions. This could be because of insufficient and unstable pulmonary support for speech during exacerbation.

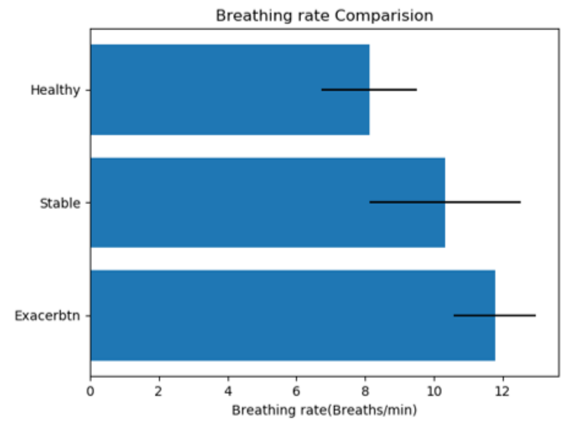

Fig. 1: Breathing rate comparision using Speech breathing model for healthy patients vs COPD patients with stable condition vs COPD patients with exacerbation condition

2. The classification approach using openSMILE features with SVM resulted in an accuracy of 75.12%. It is also observed that a smaller feature set, eGeMAPS with 88 features, also results in comparable results, which could save computational time and power.

3. Speech breathing models: We observe a clear distinction in the breathing rates for healthy subjects and subjects with COPD. Also, we observe that the average breathing rate for stable condition is much lesser than the breathing rate for exacerbation conditions. Thus breathing rate estimated from speech breathing model could be used as an important parameter for COPD exacerbation detection.

The results of each of these approaches support the hypothesis of this study which is based on the medical professional's observations that speech of COPD patients during stable periods differs from the speech during exacerbation. Continuous speech monitoring could facilitate a more thorough and adequate assessment for early recognition of COPD exacerbation and a potentially cost-effective option for long-term remote monitoring in telehealth care applications.

6. ACKNOWLEDGEMENTS

This work was partially supported by the Horizon H2020 Marie Skłodowska-Curie Actions Initial Training Network European Training Network project under grant agreement No. 766287 (TAPAS) and Data Science Department, Philips Research, Eindhoven.

7. REFERENCES

[1] Venkata Srikanth Nallanthighal, Aki Härmä, and Helmer Strik, "Deep Sensing of Breathing Signal During Conversational Speech," in *Proc. Interspeech 2019*, 2019, pp. 4110–4114.

[2] Venkata Srikanth Nallanthighal, Aki Härmä, and Helmer Strik, "Speech breathing estimation using deep learning methods," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1140–1144.

[3] Venkata Srikanth Nallanthighal, Zohreh Mostaani, Aki Härmä, Helmer Strik, and Mathew Magimai-Doss, "Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings," *Neural Networks*, vol. 141, pp. 211–224, 2021.

[4] Sander Boelders, Venkata Srikanth Nallanthighal, Vlado Menkovski, and Aki Härmä, "Detection of mild dyspnea from pairs of speech recordings," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 4102–4106.

[5] Alan D Lopez, K Shibuya, C Rao, CD Mathers, AL Hansell, LS Held, V Schmid, and S Buist, "Chronic obstructive pulmonary disease: current burden and future projections," *European Respiratory Journal*, vol. 27, no. 2, pp. 397–412, 2006.

[6] A Sonia Buist, Mary Ann McBurnie, William M Vollmer, Suzanne Gillespie, Peter Burney, David M Mannino, Ana MB Menezes, Sean D Sullivan, Todd A Lee, Kevin B Weiss, et al., "International variation in the prevalence of copd (the bold study): a population-based prevalence study," *The Lancet*, vol. 370, no. 9589, pp. 741–750, 2007.

[7] John R Hurst, Jørgen Vestbo, Antonio Anzueto, Nicholas Locantore, Hana Müllerova, Ruth Tal-Singer, Bruce Miller, David A Lomas, Alvar Agusti, William MacNee, et al., "Susceptibility to exacerbation in chronic obstructive pulmonary disease," *New England Journal of Medicine*, vol. 363, no. 12, pp. 1128–1138, 2010.

[8] Jaap CA Trappenburg, Evelyn M Monninkhof, Jean Bourbeau, Thierry Troosters, Augustinus JP Schrijvers, Theo JM Verheij, and Jan-Willem J Lammers, "Effect of an action plan with ongoing support by a case manager on exacerbation-related outcome in patients with copd: a multicentre randomised controlled trial," *Thorax*, vol. 66, no. 11, pp. 977–984, 2011.

[9] Anne B Wichmann, Hanna van Dam, Bregje Thoonsen, Theo A Boer, Yvonne Engels, and A Stef Groenewoud, "Advance care planning conversations with palliative patients: looking through the gp's eyes," *BMC family practice*, vol. 19, no. 1, pp. 1–9, 2018.

[10] Maud Ten Koppel, Bregje D Onwuteaka-Philipsen, Lieve Van den Block, Luc Deliens, Giovanni Gambassi, Martijn W Heymans, Marika Kylänen, Mariska G Oosterveld-Vlug, H Roeline W Pasman, Sheila Payne, et al., "Palliative care provision in long-term care facilities differs across europe: results of a cross-sectional study in six european countries (pace)," *Palliative medicine*, vol. 33, no. 9, pp. 1176–1188, 2019.

[11] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[12] Youri Maryn and Nelson Roy, "Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity," *Jornal da Sociedade Brasileira de Fonoaudiologia*, vol. 24, pp. 107–112, 2012.

[13] Jeroen Van de Weijer and I Slis, "Nasaliteitsmeting met de nasometer," *Logopedie en Foniatrie*, vol. 63, no. 97, pp. 101, 1991.

[14] Julia Merkus, FCW Hubers, Catia Cucchiarini, and Helmer Strik, "Digital eavesdropper. acoustic speech characteristics as markers of exacerbations in copd patients," 2020.

[15] James W Grice and Michiko Iwasaki, "A truly multivariate approach to manova," 2008.

[16] Felix Weninger, Florian Eyben, Björn W Schuller, Marcello Mortillaro, and Klaus R Scherer, "On the acoustics of emotion in audio: what speech, music, and sound have in common," *Frontiers in psychology*, vol. 4, pp. 292, 2013.

[17] Kun Han, Yanzhang He, Deblin Bagchi, Eric Fosler-Lussier, and DeLiang Wang, "Deep neural network based spectral feature mapping for robust speech recognition," in *INTERSPEECH*, 2015.

[18] Felix Scholkmann, Jens Boss, and Martin Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[19] Susanne Fuchs, Uwe D. Reichel, and Amelie Rochet-Capellan, "Changes in speech and breathing rate while speaking and biking," in *ICPhS 2015: 18th International Congress of Phonetic Sciences*, 2015.