



# Assessing the Clinical and Functional Status of COPD Patients Using Speech Analysis During and After Exacerbation

Wolfgang Mayr <sup>1,\*</sup>, Andreas Triantafyllopoulos<sup>2,3,\*</sup>, Anton Batliner <sup>2,3</sup>, Björn W Schuller<sup>2-5</sup>, Thomas M Berghaus<sup>1,6</sup>

<sup>1</sup>Department of Cardiology, Respiratory Medicine and Intensive Care, University Hospital Augsburg, Augsburg, Germany; <sup>2</sup>Chair of Health Informatics (CHI), Department of Clinical Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany; <sup>3</sup>Munich Center for Machine Learning (MCML), Munich, Germany; <sup>4</sup>Group on Language Audio, & Music (GLAM), Imperial College, London, UK; <sup>5</sup>Munich Data Science Institute (MDSI), Munich, Germany; <sup>6</sup>Medical Faculty, Ludwig Maximilians University of Munich, Munich, Germany

\*These authors contributed equally to this work

Correspondence: Wolfgang Mayr, Department of Cardiology, Respiratory Medicine and Intensive Care, University Hospital Augsburg, Stenglinstrasse 2, Augsburg, D-86156, Germany, Email wolfgang.mayr@uk-augsburg.de; Andreas Triantafyllopoulos, Chair of Health Informatics, Department of Clinical Medicine, Klinikum rechts der Isar, Technical University of Munich, Ismaninger Straße 22, Munich, 81675, Germany, Email andreas.triantafyllopoulos@tum.de

**Background:** Chronic obstructive pulmonary disease (COPD) affects breathing, speech production, and coughing. We evaluated a machine learning analysis of speech for classifying the disease severity of COPD.

**Methods:** In this single centre study, non-consecutive COPD patients were prospectively recruited for comparing their speech characteristics during and after an acute COPD exacerbation. We extracted a set of spectral, prosodic, and temporal variability features, which were used as input to a support vector machine (SVM). Our baseline for predicting patient state was an SVM model using self-reported BORG and COPD Assessment Test (CAT) scores.

**Results:** In 50 COPD patients (52% males, 22% GOLD II, 44% GOLD III, 32% GOLD IV, all patients group E), speech analysis was superior in distinguishing during and after exacerbation status compared to BORG and CAT scores alone by achieving 84% accuracy in prediction. CAT scores correlated with reading rhythm, and BORG scales with stability in articulation. Pulmonary function testing (PFT) correlated with speech pause rate and speech rhythm variability.

**Conclusion:** Speech analysis may be a viable technology for classifying COPD status, opening up new opportunities for remote disease monitoring.

**Keywords:** COPD, pathological speech, feature interpretation, personalization, digital health

## Introduction

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity. Patients typically present dyspnoea, coughing, wheezing, chest tightness, fatigue, and increased sputum production, while also being at risk of exacerbation. Exacerbations may be associated with acute airway infections but can also occur independently and may increase in frequency and intensity with disease progression.<sup>1,2</sup>

The diagnosis and classification of COPD is typically based on the severity of air-flow limitation in pulmonary function testing (PFT) and takes into account symptoms and risks of exacerbation.<sup>3,4</sup> PFT remains the gold standard for diagnosing and monitoring COPD.<sup>5</sup> In addition, clinical symptoms can be used to monitor COPD status, like sputum production, dyspnoea, and exercise capacity. Here, various questionnaires or scales, like the COPD Assessment Test (CAT) or the BORG Scale, are used in clinical routine. The CAT is a cost-effective and convenient tool, showing good sensitivity and specificity to detecting disease severity and exacerbation frequency in patients with COPD.<sup>6</sup>

The downside of PFT and clinical assessment is the need for specific tools and trained personnel, as well as patient cooperation. In addition, these tests are expensive and time consuming.<sup>5</sup> Self-reported symptomatology, on the other hand, is often highly subjective. For those reasons, alternative diagnostic methods are highly desirable as they can mitigate costs while maintaining clinical validity.<sup>7</sup> For COPD, in particular, the automatic analysis of audio offers a promising avenue for research, given the important role of respiration in speech production and the evident manifestation of COPD in speech or coughing sounds – especially during exacerbations.<sup>8–10</sup>

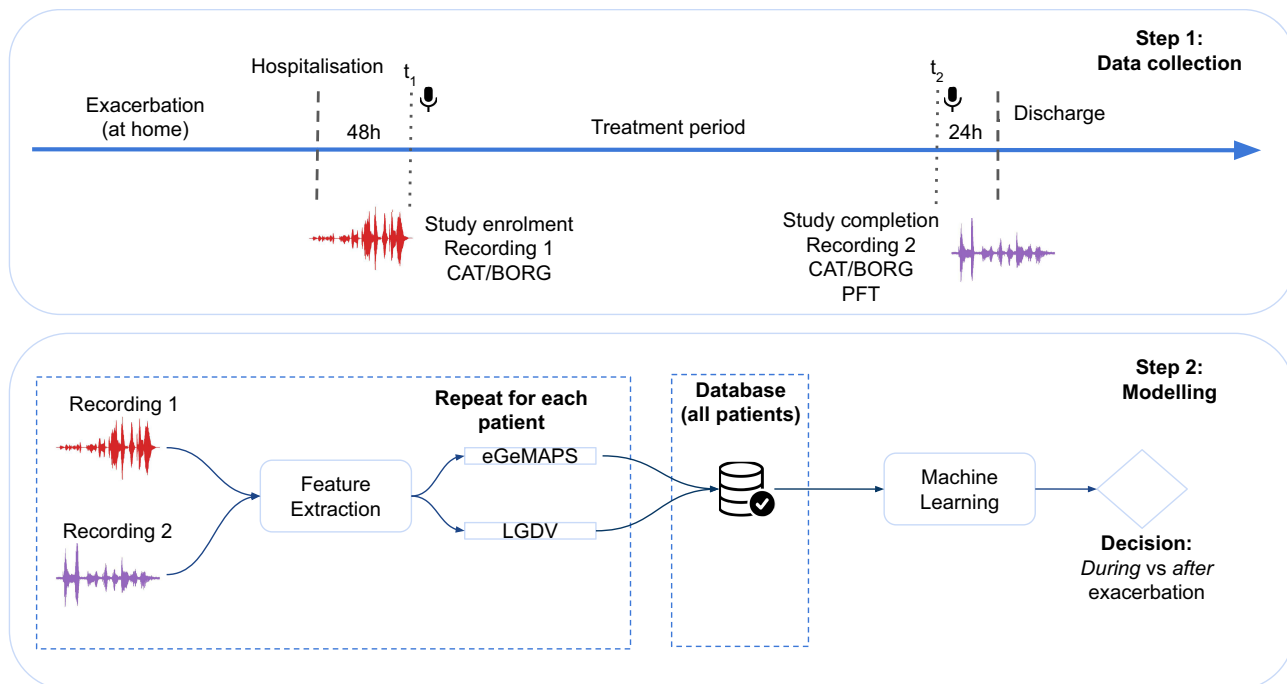
Previous work has already shown that artificial intelligence (AI)-based speech analysis can be successfully used to distinguish between pre- and post-treatment speech following an exacerbation episode.<sup>10</sup> In this research project, we extend our previous work by considering a larger patient cohort and a more comprehensive analysis of speech patterns, as well as linking these patterns to clinical variables that are typically collected for COPD patients.

## Methodology

In **Figure 1** data collection, assessment of disease severity, speech feature extraction, and machine learning (ML) experiments are shown.

### Data Collection

In this single centre study, non-consecutive patients with an acute COPD exacerbation were prospectively enrolled at the University Hospital of Augsburg within 48 hours after emergency admission between October 2020 and February 2023. The local ethics committee approved the study on June 24, 2020 (BKF 2020-34). Only patients older than 18 years, with a diagnosed COPD, able to sit and read a short text without any need for ventilation or oxygen supply during the time of recording, were included. Exclusion criteria were relevant other respiratory diseases, ear, nose, and throat (ENT) comorbidities, acute left- or right heart failure or any palliative situation.



**Figure 1** Visual overview of our methodology. Patients were admitted to the University Hospital Augsburg following an acute exacerbation. They are recruited to our study within 48 hours of hospitalization, at which point a first recording is made and the CAT/BORG questionnaires are filled. They remain hospitalized for a duration determined by their treating physician. A second recording is done within 24 hours before discharge, along with a second round of CAT/BORG questionnaires and pulmonary function testing (PFT). For each patient, we extract a set of speech features for each of their two recordings. The corresponding dataset of speech features for all patients is used as input to a machine learning classification system which distinguishes between speech collected at enrolment ( $t_1$ ) vs discharge ( $t_2$ ).

**Abbreviations:** CAT, COPD Assessment task; eGeMAPS, Geneva Minimalistic Acoustic Descriptor Set; LGDV, Local and Global Duration Variability features.

Clinical baseline data were extracted from medical records and through standardized patient interviews. For the assessment of clinical symptoms, we used the BORG CR10 Scale<sup>11,12</sup> and the COPD Assessment Test (CAT).<sup>13</sup> These tests were carried out both on admission and before discharge.

Two recordings – one during and the second after an acute exacerbation – took place bedside with the portable Zoom® H5 recorder and a Sony® ECM-144 lapel microphone. All patients obtained their individual home medication and a standard treatment of their exacerbation, mainly LABA and LAMA, sometimes in addition to ICS, systemic corticosteroids, and antibiotics. Some patients needed non-invasive ventilation, which was paused for the time of recording, as well as oxygen supply. The two recordings were performed under controlled settings in a quiet surrounding without any background noises.

Patients were required to produce a) a set of sustained vowels (/a:/, /e:/, /i:/, /o:/, /u:/); b) a few spontaneous utterances; c) (forced) coughing; d) normal breathing; e) reading a standardized text “Der Nordwind und die Sonne” (The Northwind and the Sun, NuS).

## Clinical Assessment of Disease Severity

In PFT, either body plethysmography or spirometry was performed once for each patient under stable clinical conditions. Forced expiratory volume in one second (FEV1% predicted), specific total airway resistance (sR<sub>tot</sub>), total lung capacity (TLC% predicted), vital capacity (VCmax% predicted), and residual volume (R % predicted) were recorded. The BORG CR10 Scale was used to evaluate the level of shortness of breath among the stages 0 (nothing at all) and 10 (extremely strong).<sup>14</sup> The COPD Assessment Test (CAT) was used to provide a simple quantified measure of the patient’s symptoms as well as quality of life indicators like sleep or energy level.<sup>13</sup> Answer possibilities were from 0 (not at all) up to 5 (all the time), resulting in a maximum possible value of 40. Due to the simple use and easy evaluation for both patients and medical staff, we used the BORG scale and the CAT questionnaire to evaluate shortness of breath. The BORG scale is well known to most of the patients and, thus, could be better evaluated than the mMRC dyspnoea scale even during severe exacerbations. Clinical data, such as current smoking status, pack-years, existing long-term oxygen therapy and/or ventilation, pre- and inner clinical ventilation therapy, sputum production, the number of acute exacerbations in the last 2 years, triggers for the current exacerbation, and the duration of the current hospitalization, were extracted from the patient’s medical record or through standardized interviews.

## Statistical Analysis

The demographic and clinical characteristics of this study population were shown as either frequencies and percentages or means with standard deviations (SD). All statistical analyses were conducted using the IBM Statistical Package for Social Science SPSS (Version 29.0.1.0). The test for normality was done using descriptive statistics. For all correlations between quantitative variables, we used the more conservative Spearman’s  $\rho$ , which does not presuppose a normal distribution. The correlation of speech features with the BORG and CAT scales was done twice, once by only using the values obtained during exacerbation, and once by using the difference of the values obtained after and during exacerbation. The correlation of speech features with PFT was conducted only once, namely after acute exacerbation. The following scale is often used for the assessment of correlations:<sup>15</sup> [0.0–0.2], very low; [0.2–0.4], low; [0.4–0.6], medium; [0.6–0.8], good; and [0.8–1.0], excellent. Additionally, we compute p-values by using Student’s *t* statistic with two degrees of freedom. Due to the intrinsic problems of Null Hypothesis Testing,<sup>16</sup> we employ p-values with Benjamini–Hochberg adjustment as descriptive measures, not as criteria deciding between hypotheses. By that, we provide a traditional measure for readers expecting p-values < 0.05 to be statistically significant. We carried out a sample size estimate using GPower. This resulted in a sample size of 26–82 for correlation coefficients between 0.30 and 0.60.

## Speech Processing

### Preprocessing

Recordings were first manually segmented into their constituent parts (sustained vowels, read speech, coughing, etc). Read speech recordings along with their transcripts were uploaded to the web tool MAUS (Munich Automatic Segmentation) for an automatic segmentation with known word sequences (forced alignment).<sup>17,18</sup>

## Speech Features

We first computed speech features for each audio using the open-source Speech and Music Interpretation by Large-space Extraction toolkit (openSMILE).<sup>19</sup> We extracted the extended Geneva Minimalistic Acoustic Descriptor Set (eGeMAPS)<sup>20</sup> – a small set of 88 acoustic parameters that has previously been shown to contain relevant information for respiratory diseases,<sup>21</sup> yielded the best performance in Triantafyllopoulos et al,<sup>10</sup> and also showed to be competitive in.<sup>22,23</sup> We complemented eGeMAPS with a small, interpretable set containing Local and Global Duration Variability features (LGDV).<sup>11</sup> As LGDV explicitly model temporal information, we removed the six temporal features included in eGeMAPS to avoid confounding: rate of loudness peaks per second; mean length and standard deviation of continuous voiced/unvoiced segments; rate of voiced segments per time.

## Machine Learning (ML) Experiments

The main goal of this study is to classify the speech of patients in two states of the disease: during and after an acute exacerbation. This is done by training an ML model using the speech features we extracted. Additionally, we train an identical ML model when using the self-reported CAT and BORG scores as input features. Note that we opt for a learning paradigm (instead of the more standard rule-based approach) to be comparable to the speech-based results.

We evaluate our performance using accuracy. Accuracy is computed as the percentage of samples classified correctly. Given the perfect class balance in our data, this is equivalent to unweighted average recall, ie, the average sensitivity for each class.

**Standardization:** As speaker standardization was found to substantially improve performance in previous work,<sup>23</sup> we used it in our study as well. It entails the computation of standardization parameters (mean and standard deviation) separately for each speaker.

**Modelling:** We employed a nested leave-one-speaker-out cross-validation, whereby data from every speaker is used exactly once for testing, each time using the data of all other speakers for training. Conceptually, leave-one-speaker-out is similar to the common practice of leave-one-out validation. However, due to intra-speaker dependencies, we do not leave out one sample at a time, but all samples of a speaker, thus ensuring no information leakage between training and testing. This means we train a total of 50 models. Each model is trained on data from 49 speakers and used to make predictions on the remaining one speaker. The process is repeated for all speakers. This process was used to train a support vector machine (SVM) classifier, where we optimize the cost parameter ( $\{.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$ ) and kernel function ( $\{\text{linear, polynomial, radial basis function (RBF)}\}$ ) in a grid search manner. These parameters are always optimized on the development partition, which is created by splitting the training speakers into two speaker-disjoint sets; for each of the 50 “external” folds, we perform an additional “internal” cross-validation using 2 folds and pick the parameters that yield the best performance over those 2 folds. Once the optimal set of parameters has been identified (based on development set performance), we train a final model on the entire training data for each fold.

**Feature Importance:** We also identify the most important features for each of the two feature sets (eGeMAPS/LGDV) by training identical models in the same setup as above, but using instead only one feature at a time.<sup>23</sup>

## Results

### Cohort

We recruited a total of 63 patients who got admitted with an acute exacerbation of COPD to the University Hospital of Augsburg. We used the data of 50 participants who completed two recordings, one within 48 hours after admission to the hospital and one 24 hours before discharge. Thirteen patients were sorted out because of a missing second recording (eg, due to early discharge, death, and others) or severe difficulties in reading.

### Clinical Data

Clinical data and PFT results are shown in [Table 1](#). Fifty-two percent of our patients were male, while all of them had a history of smoking. The mean duration of COPD disease is 9.6 years with a minimum duration of 3 and a maximum duration of 22 years. Thirty-two percent of the participants were diagnosed with COPD GOLD 4, 44% with GOLD 3, 22% with GOLD 2, 0% with GOLD 1 (information for 1 participant was missing). According to the new COPD-ABE

**Table 1** Available Clinical and Patient Metadata Statistics

Variable	N	Mean (Standard Deviation)
Age [years]	50	69 (10)
Pack-years	50	51 (24)
Neck circumference [cm]	47	42.2 (9.86)
Weight [kg]	50	72.3 (23.8)
Height [cm]	50	167.9 (9.1)
Hospitalisation [days]	49	9.3 (4.2)
FEV1 [%]	46	39.8 (14.6)
TLC [%]	43	122.4 (21.5)
FEV1/FVC	46	48.3 (12.2)
RV [%]	43	205.9 (65.4)
VCM <sub>max</sub> [%]	46	65.9 (21.2)
sR <sub>tot</sub> [kPa * s]	43	5.3 (3.1)
Variable	N	N per category
NIV	50	No (23); Yes (27)
LTOT	50	No (12); Yes (38)
COPD GOLD Stage	49	A (0); B (11); C (22); D (16)
Infection	50	No (21); Yes (19)

**Notes:** We include the total count (N) of patients for which the information was available. For numerical variables, we include the mean ( $\pm$  standard deviation). For categorical variables (NIV/GOLD/LTOT/Infection), we include the number of patients in each category. PFT indicators were collected prior to discharge. Metadata include: amount of patients undergoing non invasive ventilation (NIV); COPD GOLD stage (GOLD) classified by PFT; forced expiratory volume in 1 second (FEV1); total lung capacity (TLC); residual volume (RV); vital capacity (VCM<sub>max</sub>); specific airway resistance (sR<sub>tot</sub>); long-term oxygen therapy (LTOT).

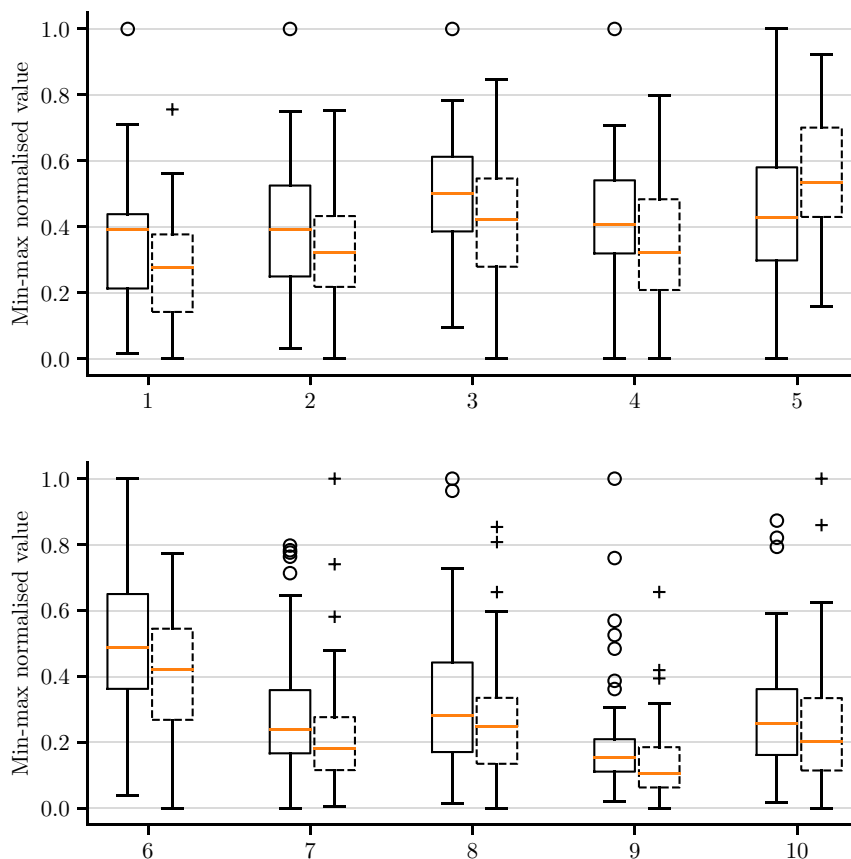
classification, all included participants were categorized into patients group E because they were all hospitalized due to an acute exacerbation.<sup>1</sup> The mean exacerbation rate in the last two years before recruitment was 2.3 per year ( $\pm 3.4$ ) with a maximum of 17 exacerbations and a minimum of 0.

All of the patients were suffering from an acute exacerbation at the time of inclusion. Ninety-eight percent received treatment with systemic steroids, 58% received additional antibiotics. Non-invasive ventilation therapy (NIV) was performed at 54% of the patients, 32% were already provided with ventilation at home and 76% with long-term oxygen therapy (LTOT). The mean time to the first in-hospital measurements, except PFT, was 1.34 ( $\pm 1.21$ ) days. PFT measurement was recorded once for each patient under stable conditions in order to achieve representative values. The mean time interval between the 1st and the 2nd recording was 7.8 days ( $\pm 4.2$ ). The mean CAT score at the time of recruitment was 28 ( $\pm 5.44$ ), and the perceived dyspnoea after the BORG scale was 4.40 ( $\pm 2.78$ ). At the time of discharge, both scales showed an improvement, to 22.38 ( $\pm 5.97$ ) in the CAT Score and 2.10 ( $\pm 1.79$ ) in the BORG score.

## Speech Analysis and Clinical Assessment

Using CAT-BORG as the only two input features in an SVM, we obtain an accuracy of 65% (95% bootstrap CI is [57%–73%]). Our speech features yield an accuracy of 84% (95% bootstrap CI is [76%–91%]) instead, thus much higher than the CAT-BORG scales. As discussed above, these results are all obtained using leave-one-speaker-out cross-validation.

Figure 2 shows the five most important features for the speech analysis, along with their changes from during exacerbation to after exacerbation status (see caption for feature explanation). After exacerbation, we observe a decrease in voice breathiness, as shown by the lowering of spectral flux, which is computed as a difference in the spectrogram energies between successive frames. Patients exhibit an improved airflow and more stable phonation, as shown by the reduced variance in loudness. Moreover, they articulate more clearly, as denoted by lower formant bandwidth spread; formants model different concentrations of acoustic energy in the vocal tract, and thus the lower bandwidth spread indicates that their spectral energy is distributed in a tighter space, which is then perceived as more clear articulation. Finally, patients read with a faster speaking tempo, manifesting as more loudness peaks per second. The LGVD features additionally captured improvements in reading rhythm and tempo both at the local (rPVIc) and global level ( $\delta C/CV(C)$ ),



**Figure 2** Top-5 most important features for eGeMAPS (top) and LGDV (bottom). We show value distributions (y-axis) of normalized feature values for the top-5 eGeMAPS (top) and top-5 LGDV features (bottom). Each boxplot pair in the x-axis denotes one feature (see below). For each feature, we show a boxplot and outliers (values beyond 95% of the distribution) for data during exacerbation (left, solid line for outliers) and after exacerbation (right, dashed line, + for outliers). eGeMAPS, Geneva Minimalistic Acoustic Descriptor Set, LGDV, Local and Global Duration Variability features. Features are (increasing number): (1) CV(F): the coefficient of variation of spectral flux, a marker of breathiness;<sup>10</sup> after exacerbation: less breathiness. (2) CV(L): The coefficient of variation of loudness, measuring the intensity of the speech signal, and thus indicative of phonation stability and airflow;<sup>10</sup> after exacerbation: less variation caused by improved airflow and phonation stability. (3/4) CV(F1-F0)/CV(F2-F0): the coefficient of variation of the amplitude of the first and second formants (relative to F0); formants model different concentrations of acoustic energy in the vocal tract (oral and nasal cavity), caused by different tongue and jaw positions; broader formant bandwidth is indicative of dysphonic speech,<sup>24</sup> meaning higher formant dispersion and mutual masking of neighbouring vowels;<sup>25</sup> after exacerbation: narrower bandwidth yields clearer articulation. (5) L/s: Loudness peaks per second, models speaking tempo;<sup>26</sup> after exacerbation: faster speech. (6) %P: the percentage of pause durations relative to total duration; shows how many breaks patients take while speaking; after exacerbation: better phonation capacities yield less breaks. (7)  $\delta C$ : the standard deviation of total consonantal durations; captures global rhythm stability; after exacerbation: more stable. (8) rPVIc: the raw pairwise variability index of the duration of successive consonantal segments; captures local rhythm stability; after exacerbation: more stable. (9) CV(C): the coefficient of variation of the total consonantal duration; captures global rhythm stability; after exacerbation: more stable. (10) Ts: The total duration (excluding pauses) the patient needs to complete reading the entire story; after exacerbation: less time.

as well as fewer pauses and less overall speaking time. Moreover, we note that some of those features can achieve good performance in isolation, with the first two LGVD features achieving 82% and 80% and the first two eGeMAPS both yielding 78%, respectively, although still below the combination of all features.

**Table 2** shows the Spearman  $\rho$  of the difference ( $\delta$ ) of the most important speech parameters between during and after exacerbation states with the difference in self-reported CAT/BORG scores (also during and after exacerbation) and the total time interval (in days) between the two recordings. We highlight Spearman's  $\rho$  values that exceed  $|\pm 0.2|$ : I) The difference in BORG is positively correlated with less dispersion in the formant space, and thus with a clearer articulation.<sup>27</sup> II) There is a positive correlation between the time interval between the two recordings and more stable reading patterns. III) Finally, a higher difference in CAT scores manifests as a more stable reading rhythm.

In **Table 3**, we additionally show the Spearman  $\rho$  of the absolute CAT/BORG scores and the speech features recorded during exacerbation. Correlations are overall lower than the above, indicating that speech features do not capture patient state as it is reflected in the CAT and BORG scores.

In addition, we compute the Spearman's  $\rho$  between the absolute values of speech parameters after exacerbation and PFT values. These are shown in **Table 4**, where we also highlight Spearman  $\rho$  values larger than  $|\pm 0.2|$ . We observe I) a

**Table 2** Spearman's  $\rho$  Between a) the Total Time Interval Between the During and After Exacerbation. Recordings (Days), b) the CAT/BORG Differences ( $\delta$ ), c) and the Difference ( $\delta$ ) in Speech Features. Between the Two Measurement Points (During vs After Exacerbation)

Value	Days	BORG	CAT
<b>CV(F)</b>	0.077	0.181	0.079
<b>CV(L)</b>	0.012	0.074	0.102
<b>CV(F1-F0)</b>	0.034	0.259	0.095
<b>CV(F2-F0)</b>	0.078	0.201	0.003
<b>L/s</b>	0.007	-0.127	-0.096
<b>%P</b>	-0.036	0.264	0.103
<b><math>\delta C</math></b>	0.137	0.074	0.323
<b>rPVIc</b>	0.211	0.032	0.272
<b>CV(C)</b>	-0.100	0.108	0.378
<b>Ts</b>	0.131	0.021	0.123

**Notes:** Difference values of during and after exacerbation were used. Effects with  $\rho \geq |0.2|$  highlighted in bold. No results were significant at  $\alpha = 0.05$  after Benjamini-Hochberg for multiple hypothesis testing. The features that are included in the table are: CV(F): the coefficient of variation of spectral flux, a marker of breathiness.<sup>10</sup> CV(L): The coefficient of variation of loudness, measuring the intensity of the speech signal, and thus indicative of phonation stability and airflow.<sup>10</sup> CV(F1-F0)/CV(F2-F0): the coefficient of variation of the amplitude of the first and second formants (relative to F0); formants model different concentrations of acoustic energy in the vocal tract (oral and nasal cavity), caused by different tongue and jaw positions; broader formant bandwidth is indicative of dysphonic speech,<sup>24</sup> meaning higher formant dispersion and mutual masking of neighboring vowels.<sup>25</sup> L/s: Loudness peaks per second, models speaking tempo.<sup>26</sup> %P: the percentage of pause durations relative to total duration; shows how many breaks patients take while speaking.  $\delta C$ : the standard deviation of total consonantal durations; captures global rhythm stability. rPVIc: the raw pairwise variability index of the duration of successive consonantal segments; captures local rhythm stability. CV(C): the coefficient of variation of the total consonantal duration; captures global rhythm stability. Ts: The total duration (excluding pauses) the patient needs to complete reading the entire story.

**Table 3** Spearman's  $\rho$  Between the Absolute CAT/BORG Values and the Speech Features Recorded During Exacerbation

Value	BORG	CAT
CV(F)	-0.063	0.135
CV(L)	-0.030	0.002
CV(F1-F0)	0.147	0.058
CV(F2-F0)	0.105	0.020
L/s	-0.127	-0.104
%P	0.138	0.031
$\delta$ C	-0.026	0.208
rPVIc	-0.020	0.127
CV(C)	0.188	0.173
Ts	-0.126	0.164

**Notes:** Effects with  $\rho \geq |0.2|$  highlighted in bold. No results were significant at  $\alpha = 0.05$  after Benjamini-Hochberg for multiple hypothesis testing. The features that are included in the table are: CV(F): the coefficient of variation of spectral flux, a marker of breathiness.<sup>10</sup> CV(L): The coefficient of variation of loudness, measuring the intensity of the speech signal, and thus indicative of phonation stability and airflow.<sup>10</sup> CV(F1-F0)/CV(F2-F0): the coefficient of variation of the amplitude of the first and second formants (relative to F0); formants model different concentrations of acoustic energy in the vocal tract (oral and nasal cavity), caused by different tongue and jaw positions; broader formant bandwidth is indicative of dysphonic speech,<sup>24</sup> meaning higher formant dispersion and mutual masking of neighboring vowels.<sup>25</sup> L/s: Loudness peaks per second, models speaking tempo.<sup>26</sup> %P: the percentage of pause durations relative to total duration; shows how many breaks patients take while speaking.  $\delta$ C: the standard deviation of total consonantal durations; captures global rhythm stability. rPVIc: the raw pairwise variability index of the duration of successive consonantal segments; captures local rhythm stability. CV(C): the coefficient of variation of the total consonantal duration; captures global rhythm stability. Ts: The total duration (excluding pauses) the patient needs to complete reading the entire story.

**Table 4** Spearman's  $\rho$  Between Most Important Speech Parameters and PFT Both Collected Within 24 Hours Prior to Discharge

Difference	FEV1 [%]	FEV1/FVC	TLC [%]	SR. TOT.	VCM <sub>max</sub> [%]	RV [%]
CV(F)	-0.018	0.068	0.126	-0.002	-0.061	-0.024
CV(L)	-0.124	0.080	0.110	0.063	-0.128	0.052
CV(F1-F0)	0.011	-0.080	0.045	-0.000	0.022	0.002
CV(F2-F0)	0.043	-0.088	0.054	-0.062	0.069	-0.013
L/s	0.126	-0.019	-0.063	-0.065	0.085	-0.057
%P	-0.162	-0.037	0.273	0.100	-0.108	0.176
$\delta$ C	0.042	-0.151	-0.035	-0.087	0.110	-0.051
rPVIc	0.030	-0.200	-0.040	-0.082	0.136	-0.051

(Continued)



**Table 4** (Continued).

Difference	FEV1 [%]	FEV1/FVC	TLC [%]	SR. TOT.	VCMax [%]	RV [%]
<b>CV(C)</b>	-0.037	-0.234	0.061	0.032	0.024	-0.009
<b>Ts</b>	0.133	0.136	-0.092	-0.152	0.102	-0.124

**Notes:** Effects with  $\rho \geq |0.2|$  highlighted in bold. Absolute values after exacerbation were used. No results were significant at  $\alpha = 0.05$  after Benjamini-Hochberg for multiple hypothesis testing. The features that are included in the table are: CV(F): the coefficient of variation of spectral flux, a marker of breathiness;<sup>10</sup> CV(L): The coefficient of variation of loudness, measuring the intensity of the speech signal, and thus indicative of phonation stability and airflow.<sup>10</sup> CV(F1-F0)/CV(F2-F0): the coefficient of variation of the amplitude of the first and second formants (relative to F0); formants model different concentrations of acoustic energy in the vocal tract (oral and nasal cavity), caused by different tongue and jaw positions; broader formant bandwidth is indicative of dysphonic speech,<sup>24</sup> meaning higher formant dispersion and mutual masking of neighboring vowels.<sup>25</sup> L/s: Loudness peaks per second, models speaking tempo.<sup>26</sup> %P: the percentage of pause durations relative to total duration; shows how many breaks patients take while speaking.  $\delta C$ : the standard deviation of total consonantal durations; captures global rhythm stability. rPVIc: the raw pairwise variability index of the duration of successive consonantal segments; captures local rhythm stability. CV(C): the coefficient of variation of the total consonantal duration; captures global rhythm stability. Ts: The total duration (excluding pauses) the patient needs to complete reading the entire story.

positive correlation between TLC [%] and the percentage of pauses; II) a negative correlation between FEV1/FVC and rhythm stability. Collectively, these observations can be interpreted as more stable reading patterns and fewer pauses for patients that show improved PFT results at discharge.

## Discussion

Our classification results are higher than those reported by other groups in prior works, such as Nallanthighal et al<sup>23</sup> who achieved an accuracy of 75% using a large feature set and 70% using the eGeMAPS features we used here, and slightly higher than our own previous work, which reached a maximum accuracy of 82% on a subset of this cohort.<sup>10</sup> Moreover, we go beyond these previous studies by contrasting speech-based predictions to standard CAT/BORG questionnaires. Speech features show a higher performance than those, as CAT/BORG are only able to distinguish between the two states with an accuracy of 65%, whereas speech features achieve 84%. Those features can be interpreted to indicate a reduction in voice breathiness, a more stable loudness and phonation, a decrease in pauses, and a more stable reading rhythm in recompensated COPD patients.

Our interpretability findings are generally in line with previous works. For example, Merkus et al<sup>28</sup> also found a lower articulation rate and increased inhalations for COPD patients compared to healthy controls.<sup>28</sup> They also found more inhalations per syllable for patients at exacerbation vs at stable state, similar to Nallanthighal et al.<sup>23</sup> This is consistent with our findings, as we found a less breathy voice and a faster reading rhythm with fewer pauses in recompensated patients. Bommel et al<sup>22</sup> also used eGeMAPS and found increased loudness in speech from COPD patients (in a stable state) vs healthy controls and higher formant bandwidths in exacerbated vs stable COPD patients, which is similar to our own findings.

Going beyond previous works, we further connect speech characteristics to differences in CAT/BORG scores and PFT values. The improvement in CAT correlates with a more stable reading rhythm, whereas the improvement in BORG goes together with a more stable articulation. Correlation is higher when considering the differences between before/during exacerbation, rather than the values during exacerbation. This can be attributed to the fact that both overall speech quality and CAT/BORG scores improve after the exacerbation episode is finished, and, thus, treatment mediates both variables, whereas the values during exacerbation correspond to a different characterization of patient state than obtained via CAT/BORG. PFT values are not highly correlated with speech characteristics, with TLC [%] showing the highest (positive) correlation with pause rate, indicating that increased TLC leads to more pauses for breathing. Additionally, patients with higher FEV1/FVC show a negative correlation with local and global rhythm variability, showcasing how patients with airflow limitations struggle to maintain a regular reading rhythm. In summary, while correlation values between PFT and speech features are generally low, they still show trends that match clinical expectations.

The generalizability of our findings is limited by the small size and the high degree of homogeneity of our cohort. Given only 50 patients all collected in a single center, we can only reliably claim that the classification of COPD patient speech during and after exacerbation states is possible in recording conditions that closely mirror ours.

An additional confounder is that our cohort mainly consisted primarily of severely ill patients, with 77% of patients categorized into COPD GOLD 3 or higher. There was a large range in the number of days patients stayed at the hospital (4–19 days) which might have resulted in an inadequate contrast between the two conditions during and after exacerbation. In addition, it is possible that not all patients had fully recovered at discharge. Another limitation is that we used the BORG scale to evaluate the shortness of breath in resting condition.

Nevertheless, using a combination of expert-based speech features, we were able to achieve an accuracy of 84%, which exceeds the accuracy achieved when relying on self-reported CAT/BORG scores and that seen in previous works.<sup>10,23,28</sup> Our findings demonstrate that speech analysis is a viable technology for classifying COPD status, opening up new opportunities for the remote monitoring of patients both in the clinic and at home.

Overall, this contribution represents a step forward in the automatic assessment of COPD patients, which can greatly support clinicians and health care providers. Our work shows that the classification of a patient's speech between a during and after exacerbation state is feasible. This line of research can enable the automatic, even "remote", monitoring of patients. Future work should investigate whether speech-based monitoring can facilitate the detection of an upcoming exacerbation episode before it manifests in self-reported symptomatology.

## Data Sharing Statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Ethics Approval and Consent to Participate

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All patients gave written informed consent. The Ethics Committee of the Faculty of Medicine of the Ludwigs Maximilian University Munich approved the study on June 24, 2020 (BKF 2020-34).

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Disclosure

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

## References

1. Sethi S, Evans N, Grant B, Murphy T. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med*. 2002;347(7):465–471. doi:10.1056/NEJMoa012561

2. Sethi S, Sethi R, Eschberger K, et al. Airway bacterial concentrations and exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2007;176(4):356–361. doi:10.1164/rccm.200703-417OC
3. Agusti A, Celli BR, Criner GJ, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Am J Respir Crit Care Med.* 2023;207(7):819–837. doi:10.1164/rccm.202301-0106PP
4. Singh D, Agusti A, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019. *Eur Respir J.* 2019;53(5):1900164. doi:10.1183/13993003.00164-2019
5. Zubaydi F, Sagahyoon A, Aloul F, Mir H, Mahboub B. Using mobiles to monitor respiratory diseases. *Informatics.* 2020;7(4):56. doi:10.3390/informatics7040056
6. Varol Y, Ozacar R, Balci G, Usta L, Taymaz Z. Assessing the effectiveness of the COPD Assessment Test (CAT) to evaluate COPD severity and exacerbation rates. *COPD.* 2014;11(2):221–225. doi:10.3109/15412555.2013.836169
7. Triantafyllopoulos A, Kathan A, Baird A, et al. HEAR4Health: a blueprint for making computer audition a staple of modern healthcare. *Front Digit Health.* 2023;12(5):1196079. doi:10.3389/fgth.2023.1196079
8. Eyben F, Scherer KR, Schuller BW, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans Affect Comput.* 2015;7(2):190–202. doi:10.1109/TAFFC.2015.2457417
9. Lee SE, Rudd M, Kim TH, et al. Feasibility and utility of a smartphone application-based longitudinal cough monitoring in chronic cough patients in a real-world setting. *Lung.* 2023;201(6):555–564. doi:10.1007/s00408-023-00647-1
10. Triantafyllopoulos A, Fendler M, Batliner A, et al. Distinguishing between pre- and post-treatment in the speech of patients with chronic obstructive pulmonary disease. In: Proceedings of Interspeech. Incheon, South Korea; 2022. 3623–3627. doi:10.21437/Interspeech.2022-10333.
11. Borg AG. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc.* 1982;14(5):377–381. doi:10.1249/00005768-198205000-00012
12. Williams N. The Borg rating of perceived exertion (RPE) scale. *J Occup Med.* 2017;67(5):404–405. doi:10.1093/occmed/kqx063
13. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Leidy NK. Development and first validation of the COPD Assessment Test. *Eur Respir J.* 2009;34(3):648–654. doi:10.1183/09031936.00102509
14. Hareendran A, Leidy NK, Monz BU, Winnette R, Becker K, Mahler DA. Proposing a standardized method for evaluating patient report of the intensity of dyspnea during exercise testing in COPD. *Int J Thron Obstruct Pulmon Dis.* 2012;7:345–355. doi:10.2147/COPD.S29571
15. Schuller BW, Batliner A. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing.* Wiley; 2013.
16. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *AM STAT.* 2016;70(2):129–133. doi:10.1080/00031305.2016.1154108
17. Nallanthighal VS, Herma A, Strik H. Detection of COPD exacerbation from speech: comparison of acoustic features and deep learning based speech breathing models. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2022. 9097–9101. doi:10.1109/ICASSP43922.2022.9747785.
18. Schiel F. Automatic Phonetic Transcription of Non-Prompted Speech. In: Proceedings of ICPhS14. San Francisco; 1999. 607–610.
19. Eyben F, Wöllmer M, Schuller BW. Opensmile: the Munich versatile and fast open-source audio feature extractor. In: Proceedings of 18th ACM international conference on Multimedia; 2010. 1459–1462. doi:10.1145/1873951.1874246.
20. Crooks MG, Brinker AD, Hayman Y, et al. Continuous cough monitoring using ambient sound recording during convalescence from a COPD exacerbation. *Lung.* 2017;195(3):289–294. doi:10.1007/s00408-017-9996-2
21. Bartl-Pokorny KD, Pokorny FB, Batliner A, et al. The voice of COVID-19: acoustic correlates of infection in sustained vowels. *J Acoust Soc Am.* 2021;149(6):4377. doi:10.1121/10.0005194
22. Van Bommel L, Harmsen W, Cucchiari C, Strik H. Automatic selection of the most characterizing features for detecting COPD in speech. In: Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings 23. Springer; 2021. 737–748. doi:10.1007/978-3-030-87802-3\_66.
23. Kisler T, Reichel U, Schiel F. Multilingual processing of speech via web services. *Comput Speech Lang Virtual Special Issues.* 2017;45:326–347. doi:10.1016/j.csl.2017.01.005
24. Ishikawa K, Webster J. The Formant bandwidth as a measure of vowel intelligibility in dysphonic speech. *J Voice.* 2023;37(2):173–177. doi:10.1016/j.jvoice.2020.10.012
25. De Cheveigné A. Formant bandwidth affects the identification of competing vowels. In: Proceedings of ICPhS14. San Francisco, CA, USA; 1999. 2093–2096.
26. Eyben F. *Real-Time Speech and Music Classification by Large Audio Feature Space Extraction.* Springer; 2015.
27. Honig F, Batliner A, Bocklet T, et al. Are men more sleepy than women or does it only look like – automatic analysis of sleepy speech. In: Proceedings of ICASSP. Florence, Italy; 2014. 995–999. doi:10.1109/ICASSP.2014.6853746.
28. Merkus J, Hubers F, Cucchiari C, Strik H. Digital eavesdropper – acoustic speech characteristics as markers of exacerbations in COPD patients. In: Proceedings of RRaPID workshop of the 12th International Conference on Language Resources and Evaluation (LREC2020). Marseille, France; 2020. 78.

International Journal of Chronic Obstructive Pulmonary Disease

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

**Dovepress**  
Taylor & Francis Group