Sleep Quality through Vocal Analysis: a Telemedicine Application

1st Federica Amato
Politecnico di Torino
Turin, Italy
federica.amato@polito.it

2nd Irene Rechichi
Politecnico di Torino
Turin, Italy
irene.rechichi@polito.it

3rd Luigi Borzì
Politecnico di Torino
Turin, Italy
luigi.borzi@polito.it

4th Gabriella Olmo
Politecnico di Torino
Turin, Italy
gabriella.olmo@polito.it

***Abstract*—Voice is a reservoir of valuable health data. Recent studies highlighted its efficacy in predicting sleep quality, and its potential as biomarker of neurodegeneration. This study assesses the feasibility of a Telemedicine system for the evaluation of sleep quality through brief vocal recordings. Machine Learning models were employed in the binary classification between good and poor sleepers, with great performance in scoring poor sleep quality – 88% and 85% F-1 score on a 5-fold Cross Validation (CV) for females and males, respectively. Moreover, the correlation between perceived sleep quality and a validated global score was studied, as well as the influence of external factors and sleep-wake schedule.**

***Index Terms*—Voice, Sleep, Tele-medicine, Artificial Intelligence, Machine Learning**

I. Introduction

Voice Analysis is receiving much attention due to the enormous amount of clinical information contained in the vocal signal. Speech production occurs through synergistic articulating movements that shape the excitation source to convey the final sound. This complex process incorporates a large amount of data of clinical interest, besides achieving the main objective of transmitting the information. These data can be implemented into Machine Learning (ML) algorithms, providing tools to support the clinical practice. Although vocal analysis can be applied to any pathology that directly or indirectly affects the vocal apparatus, recent evidence shows its high potential in identifying sleep disorders [1]–[3].

Sleep is a transitory state of altered consciousness, which not only serves a restorative function, but also plays a pivotal role in the removal of metabolic waste products from the Central Nervous System (CNS) [4]. This occurs through the glymphatic activity [5] during Slow Wave Sleep (SWS) – i.e., the deepest stage of sleep [6]. Low Sleep Quality (SQ) leads to increased fatigue, and may also be a predictor of neurodegeneration [7]. Sleep is commonly assessed through Polysomnography (PSG), an invasive diagnostic test which consists in recording biosignals during sleep through a huge number of electrodes. Evidence from [8] shows poorer voice quality in signals recorded under stressful conditions (e.g., sleep deprivation).

In this work the authors investigated speech samples of 47 healthy subjects recorded through professional equipment while executing several vocal tasks. The results revealed worse Harmonic to Noise Ratio (HNR) values after 24 hours of sleep deprivation, especially in the female subgroup.

In a similar work [3], Kim et al. investigated the ability of voice analysis to predict SQ. 203 English healthy native speakers were recruited and asked to answer a set of questionnaires using mobile devices and collect a set of voice samples (free speech, sentence, and text). Regression performance was evaluated in a 5-fold CV in terms of Concordance Correlation Coefficient (CCC) between real and predicted scores. This value was equal to 0.5 for the SQ index. Promising results were also reported for Obstructive sleep apnea (OSA) detection [2], on speech samples of 45 Portuguese subjects (25 OSA, 20 controls) performing a free monologue and a read text tasks. Vocal features, including Fundamental frequency (F0), HNR and cepstral coefficients, were computed and input to a majority voting ensemble of Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and k-Nearest Neighbor (KNN). The model yielded 88% True Positive Rate (TPR) and 80% True Negative Rate (TNR).

Our study aims at implementing a model for the automatic binary classification of subjects characterized by good and poor sleep quality. In this regard, the COVID-19 pandemic highlighted the need for safe and reliable remote monitoring systems in healthcare. Assessing SQ through Telemedicine, without the need of costly and invasive exams, would have significant impact on the quality of life. Therefore, the main objective of this work was to test the feasibility of a tele-monitoring application, ensuring social-distancing in the medical domain. The study involved the collection of vocal samples from volunteers, along with items from two sleep questionnaires. In Section III, we describe the analysis of vocal signals and sleep questionnaires; in Section IV we present the results.

II. Materials

The data were collected from Italian speakers through a simple WA available on Desktop and Mobile Web browsers. We designed the WA specifically to guide users through performing a voice test and two sleep questionnaires. The test consisted in reading a phonemically balanced text proposed in [9] to capture different facets of pathological voices. It is long enough to stress resistance, it contains intricate phonetics to assess the ability to pronounce complex sounds, and requires changes in expression during reading [9]. Here, we report the text employed (Italian language):

IL RAMARRO DELLA ZIA *Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi, detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana che saltando dalla strada finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxi. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.*

After the vocal task, the WA required the users to fill in a questionnaire to assess sleep quality and an additional survey – detailed in Section III-B2 – that addressed the subjects' daily habits and Sleep–Wake cycle. Furthermore, we added a short section to collect the age, gender, and level of education of the speakers.

135 anonymous volunteers (55 males) participated; among these, 70 subjects (37 males) performed all tasks – i.e., the recording and the sleep questionnaires – and we focused on this subgroup to perform the analysis. Given the influence of the gender on most of the vocal features extracted, we split the dataset into two groups (males and females). Then, we applied the same workflow to each cluster. Data collection, in accordance with the Declaration of Helsinki, was approved by the Ethics Committee of the A.O.U Città della Salute e della Scienza di Torino (approval number 00384/2020). Informed consent for observational study was obtained; demographic and clinical data were noted anonymously. Table I reports an overview of the collected information. Regarding Age, values displayed in Table I, do not include 10 subjects (8 females) which failed in entering their age on the online form. Given that this subset represents the 14% of the whole group (and 23% of the female subjects), for the subsequent analysis (*cf.* Section III-A2) the missing values were replaced with the median of the whole group – per each gender.

Data analysis and classification was carried out in Python; Praat (by Paul Boersma and David Weenink, Phonetic Sciences, University of Amsterdam) was mainly employed for preprocessing and feature extraction.

III. Methods

A. Speech Analysis

1) Signal pre-processing: This section describes the preprocessing steps carried out on the vocal signals, performed through the software Praat. The recordings were firstly downsampled to 16 kHz and a de-noising filter with Praat default hyperparameters was applied to each signal; the signal amplitude was normalized in the range [0, 1] to prevent the speaker-microphone distance from affecting the model. It is worth noting that we manually removed initial and final silence regions; hence, no further preparatory steps were required. Finally, we employed the Praat software to detect voiced regions' start- and end-point.

2) Feature extraction: Raw vocal signals do not provide much information unless a proper feature extraction procedure is implemented. Given the absence of a specific set of features with proven high correlation with the application at hand, we decided to extract a total number of 96 speech features and inspect their effectiveness. We derived *timing measures* from the entire vocal signal and more specific features from the voiced regions. The first group is intended to detect changes in the rhythmic organization of the speech and encompasses the Number of Pauses (NP), the Duration of Pause Intervals (DPI) defined in [10] as the median duration of unvoiced intervals exceeding 30 ms [11], and the Rate of Speech Timing (RST), proposed in [10] to evaluate the capability of alternating voiced, unvoiced and paused regions. The latter include *periodicity measures* – i.e, F0 and the first three Formants; *noise measures* – i.e, HNR, Cepstral Peak Prominence (CPP), and Glottal to Noise Excitation ratio (GNE) – and *amplitude related measures* (i.e., Intensity). Moreover, we extracted spectral and cepstral features (12 Mel Frequency Cepstral Coefficients (MFCC)), 13 Perceptual Linear Prediction (PLP)) which proved to be effective for the application at hand in similar works [2], [3]. In more detail, after identifying and merging voiced regions, we framed each signal into 25 ms windows with 10 ms overlap [3] and extracted features from each segment. Finally, we grouped them into one feature vector, and calculated five statistics (i.e., mean value, median value, standard deviation, kurtosis, and skewness). Table II reports an overview of the features employed. Z-score normalization was applied to the whole feature set to scale features to the same range. This, besides being a general good practice, is particularly important if Euclidean distances have to be computed in subsequent analysis (e.g., similarity measures).

B. Sleep Analysis

1) Sleep Quality Assessment: The shortened sPSQI [13], a validated, 13-item survey, was employed to assess overall sleep quality, adapted from the original Pittsburgh Sleep Quality Index (PSQI) [14]. The obtained global score is used to discriminate between *good* and *poor* sleepers. Differently from the standard PSQI, the sPSQI includes only self-reported questions. As introduced in Section I PSG is the gold standard test to diagnose Sleep Disorders; however, the PSQI is a reliable tool commonly used in reasearch and the medical practice. The index evaluates sleep quality over five components: Sleep Latency, Duration, Efficiency, Disturbances and Daytime Dysfunction. The sPSQI score was obtained from the 70 available subjects. In our dataset, the score ranges in (0,10), with 10 indicating a negative extreme, with a global score of 6.21 ± 1.72 (median: 6.0). When presenting the shortened PSQI score [13], the Authors suggested a cut-off value of 4 for detecting poor sleepers. According to the data collected in the present study, the range, and distribution, we set the cut-off value at 5.0, identifying 26 good sleepers and 44 poor sleepers.

2) Sleep Features: A second sleep survey (SLEEPS) was administered to the subjects through the WA. The survey includes 21 self-reported items that provide an overview of

TABLE I: Demographic Characteristics in the analysed population

	Age	Education Level	Employment	Remote Working/Learning	sPSQI	SLEEPS score	Disease Insomnia, OSA, Covid-19
Females	41.4 ± 18.1	Middle School: 3 (9%) Secondary School: 7 (21%) Bachelor's/Master's Degree: 23 (70%)	Students: 25 (76%) Workers: 21 (64%)	9 (27%)	6.38 ± 1.51	2.29 ± 1.04	1 OSA 1 Covid-19
Males	36.9 ± 14.5	Middle School: 2 (5%) Secondary School: 10 (27%) Bachelor's/Master's Degree: 25 (68%))	Students: 26 (71%) Workers: 27 (73%)	12 (33%)	6.05 ± 1.88	2.39 ± 0.92	

the subjects' sleep-wake schedule. The aim of this survey is to investigate the link between the subjects' habits and their quality of sleep. The set of questions included in SLEEPS is displayed in Table III. One item in the SLEEPS assesses the Perceived Sleep Quality (SLEEPQ), which is usually evaluated through a sleep diary in actigraphy studies and compared to the actual value [15]. Sleep was assessed on a 5-point rating scale (i.e., *Excellent - Above Average – Average - Below Average - Very Poor*). All items were scored on a Likert scale of 0-4 where 4 represented the negative extreme, following the protocols established in [14] and also applied in the design of the sPSQI (*cf.*, Section III-B1). Binary or numerical answers were adapted accordingly.

C. Feature selection

Prior to feature selection we performed an early fusion in the feature space in order to merge vocal and sleep features extracted for each subject. Then we applied the feature selection workflow which was adapted from a from a similar work, which however involves Parkinson's disease (PD) subjects [16].

A correlation-based approach was implemented; this procedure selects the most significant (i.e., those with high feature-target correlation) and non-redundant features (i.e., those with low inter-feature correlation). First, we evaluated Pearson's correlation r between features and target (r_{fo}), investigating its absolute value for each feature and retaining only the most significant ones (i.e, $r > 0.4, P - value < 0.02$). Then we computed the intra-feature correlation (r_{ff}). For couples of features showing inter-correlation higher than intra-correlation (i.e, $r_{ff} > r_{fo}$), the feature less correlated to the class was removed.

TABLE II: List of features extracted. **V:** voiced, **UNV:** unvoiced. **TW:** This work

Feature (type and name)	Region analyzed	Application in SQ assessment
Periodicity: F0, Formants, Bandwidths	V	[3], [8], [12]
Intensity	V	[12]
Noise: HNR, GNE, CPP	V	[3], [8], [12] + TW
Spectral: mean, std, skew., kurt, roll off, slope, decrease	V	[8] + TW
MFCC, ΔMFCC, ΔΔMFCC	V	[3]
PLP, ΔPLP, ΔΔPLP	V	[3]
Timing NP, DPI, RST	V + UNV	TW

D. Classification

Automatic binary classification between subjects characterized by good and poor sleep quality was performed. The label was obtained by setting a threshold on the continuous sPSQI score, as commonly done in the clinical practice [17]. Therefore, the quality threshold was set at 5.0 (*cf.* III-B1, and all subjects with continuous sPSQI score above-threshold were labelled as poor sleepers.

To avoid weak generalization capability of the model, we randomly split the database into two subsets: 80% to be used during the training/validation phase and 20% to be used as the test set. We implemented feature selection, model selection and model optimization on the training/validation set only, while the remaining 20% of subjects, employed in the testing phase, underwent no further optimization.

The extracted Speech and Sleep features were merged into a single vector and inputted to the classifier. Then, the performances of 7 ML classifiers were compared; the classifier featuring the highest F-1 score was optimized. Given the unbalanced dataset – with higher cardinality for poor sleepers, we

TABLE III: Items and Scores of the SLEEPS Questionnaire

Item	Score
I. General Data	
Covid-19 diagnosis	Scale
OSA diagnosis	Scale
Insomnia	Scale
University	Y/N
Work	Y/N
II. Work and Study Habits	
Remote Working/Learning	Y/N
Hours of Screen Time	Scale
End of use Time of Electronic Devices	Numeric
Blue Light Filter	Y/N
III. Leisure Time Habits	
Time spent away from home during workdays	Numeric
Time spent outside over the weekend	Numeric
Excercise hours (outdoors, per week)	Numeric
Excercise hours (indoors, per week)	Numeric
Time spent working on a hobby (per week)	Numeric
IV. Sleep Habits	
Nocturnal awakenings	Scale
Getting up at night	Scale
Morning drowsiness	Scale
Morning fatigue	Scale
Fatigue, poor concentration and impaired performance	Scale
Difficulty falling asleep	Scale
Perceived sleep quality	Scale

considered the F-1 score to be a reliable metric for performance evaluation. We tested several approaches – such as Naive Bayes (NB), KNN, SVM and Random Forest (RF) – as well as ensemble methods, such as Adaptive Boosting (ADA), Gradient Boosting (GB), and Bagging ensemble (BAG) classifiers.

Given the random splitting procedure employed, we performed each experiment 10 times on 10 randomly extracted subsets, and averaged the performance metrics for comparing the classifiers. Hyperparameters optimisation (Grid Search approach) was applied to the best model; we evaluated accuracy, F1-score, precision, sensibility, specificity and the Area Under the Curve (AUC) as an average over 10 iterations, to further assess the stability of the final mode.

IV. Results and discussion

A. Feature Relevance

In this section we investigated the significance of the features extracted according to the Pearson correlation coefficients between each feature and the class. As far as concerns the female population, the three most correlated features resulted ΔΔMFCC12 std (r: -0.70, p<0.001), ΔMFCC12 std (r: -0.67, p<0.001), and ΔΔMFCC10 std (r: 0.60, p<0.001), highlighting the importance of the MFCC coefficients for the application at hand. For the male population, the analysis confirmed the importance of the MFCC coefficients found for the female group (ΔMFCC3 mean (r: 0.50, p: 0.001), and ΔΔ MFCC6 mean (r: -0.46, p<0.004) and introduced the significance of the SLEEPQ parameter measured through the sleep questionnaire (SLEEPQ (r: -0.56, p<0.001).

B. Population Inspection

The scores collected through the SLEEPS were analysed along with the global sPSQI score. A recent study suggested moderate correlation of poor sleep quality and Covid-19-related factors [18] (positivity to the virus, proximity with COVID+ people). SQ was evaluated through the Medical Outcomes Study-Sleep Scale (MOS-SS) Index II score [19]. Among the observed factors, positivity to Covid-19 resulted in being fairly significant. Hence, we examined the sPSQI in our sample and its distribution compared to the Covid-19 parameter – i.e., past (P) positivity, current positivity (C) or never diagnosed (N). Males who were previously affected by the virus presented with worse sPSQI scores, resulting in poorer sleep quality. Instead, regarding the N-condition, the score was equally distributed in the data range. No similar trend was observed in the Female group. Fig. 1 shows the comparison between sPSQI and SLEEPQ. As per Males, the two scores concur. Instead, some subjects in the Female group failed in rating their sleep, showing overrating tendencies. As appreciable, 35% scored their sleep as Average, instead featuring a sPSQI of 7.0 $\pm$ 1.53 and 32.4% of the subjects scored sleep as Above Average and resulted in a sPSQI of 6.0 $\pm$ 0.95 – all below average.

Finally, the items collected through the SLEEPS were ranked according to their correlation with SLEEPQ. As expected, the frequency of nocturnal awakenings, the occurrence of insomnia and hours of sleep were highly correlated with the perceived sleep quality. No significant correlation was found with remote working (or learning), or the use of a blue light filter. Instead, reasonable correlation was found with the end of use time of electronic devices, morning drowsiness and difficulty falling asleep.

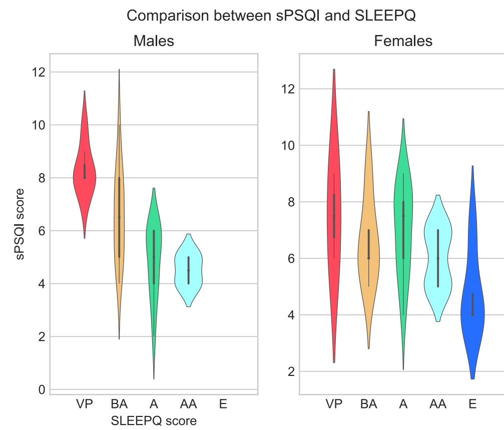


Fig. 1: Comparison between SLEEPQ score and sPSQI in the Males and Females groups. Quality: **VP**: very poor, **BA**: below average, **A**: average, **AA**: above average, **E**: excellent.

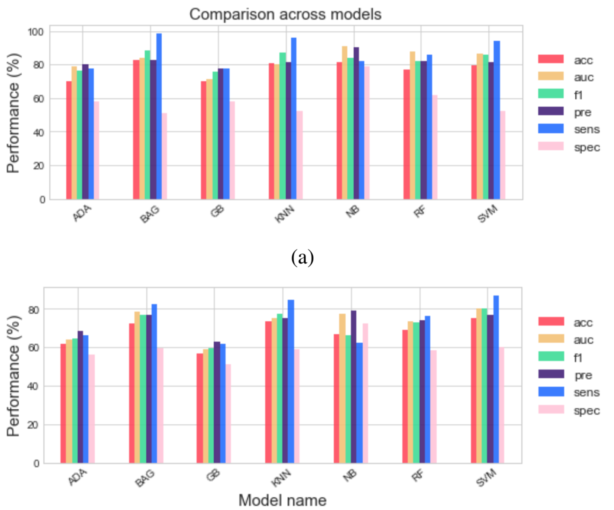
C. Classification Results

This section presents the results of the binary classification performed as described in Section III-D, along with a discussion. Fig. 2 displays the comparison of the classification F-1 score on the 7 models tested. The values refer to a 5-fold CV and are averaged over 10 iterations.

Regarding females, the best performance was achieved through the BAG classifier (88% $\pm$ 3.4), the KNN classifier (87% $\pm$ 3.7) and the SVM classifier (86% $\pm$ 3.7). In both subgroups, no model clearly outperformed; hence, on the three best models (BAG, KNN, SVM) we performed the hyperparameters tuning. The considered parameters were: number-of-estimators (2 to 50, steps of 2), distance metrics, K (from 2 to $N_{samp}/2$), C (10, 100, 1000) and gamma (0.1, 0.001, 0.0001), kernel, for the three classifiers respectively. A SVM (C = 10, gamma = 0.001, kernel = RBF) and a KNN (k=7, Chebyshev distance) emerged as the best models for the female and male subgroups, respectively.

The final performance of the optimized models is reported in Table IV. No impairment in performance is observed when moving from validation data to completely new samples contained in the test set, suggesting lack of overfitting and good generalization. The different cardinality in the dataset classes inevitably leads to low classification specificity. However, the good overall performance suggests that this may be compensated when testing the proposed algorithm on larger datasets.

Table V reports an overview of the features resulting from the feature selection procedure, for the male and female subgroups.



(a)

(b)

Fig. 2: Performance comparison across the 7 model tested. **(a)** Female subgroup, **(b)** Male subgroup.

V. Conclusions and Future Work

This study presented a workflow for classifying SQ based on vocal analysis, through a Telemedicine system and ML techniques. Vocal signals were recorded on personal computers or smartphones. Despite the lack of professional microphones and task-training, the employed ML models proved efficient, with F-1 scores of 88% and 85% for females and males, respectively. Higher performance in the female subgroup may be due to the intrinsic structure of the female vocal apparatus, which may be more prone to vocal changes resulting from sleep disturbances, as also noted in [8]. This study is not without limitations. First of all, the target was identified through the sPSQI score, which is a clinically validated index which, however, relies only on subjective items. Future work will address this issue, and include in the analysis objective parameters – e.g., actigraphy-derived measures [20]. Finally, enhanced performances may be obtained by enlarging the dataset and performing a stratified analysis per age or physiological characteristics – e.g., similar sleep-wake routine, comorbidities, other physiological or demographic parameters.

TABLE IV: Classification performance of the optimized models. Average over 10 iterations employing a 5-fold CV are reported.

	Female		Male	
	Validation	**Test**	**Validation**	**Test**
Acc.	83 ± 4.4	86 ± 9.0	82 ± 6.0	84 ± 11.3
Pre.	84 ± 3.4	85 ± 8.2	83 ± 4.6	88 ± 11.1
F1	88 ± 3.0	91 ± 5.8	85 ± 4.8	87 ± 9.1
Sens.	96 ± 3.4	98 ± 6.0	91 ± 5.2	88 ± 13.3
Spec.	60 ± 12.6	55 ± 26.7	69 ± 10.5	77 ± 21.3
AUC	0.92 ± 0.03	0.76 ± 0.14	0.84 ± 0.07	0.82 ± 0.12

TABLE V: Overview of feature selected in the final model.

Female	**Male**
12^{th} MFCC; 10^{th} 12^{th} ΔMFCC	3^{rd} Formant; 1^{st} MFCC
5^{th}, 6^{th}, 10^{th}-13^{th} $\Delta\Delta$ MFCC	3^{rd} , 6^{th}, 7^{th} ΔMFCC
4^{th} ΔPLP ; 5^{th}, 8^{th} $\Delta\Delta$ PLP	1^{st} PLP ; 1^{st} Δ PLP
Spectral Flux; Spectral Roll-off point	1^{st} $\Delta\Delta$ PLP

References

[1] B. R. Rocha and M. Behlau, "The Influence of Sleep Disorders on Voice Quality," *Journal of Voice*, vol. 32, no. 6, pp. 771.e1–771.e13, 2018.

[2] M. C. Botelho, I. Trancoso, A. Abad, and T. Paiva, "Speech as a biomarker for obstructive sleep apnea detection," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5851–5855, 2019.

[3] S. Kim, N. Kwon, H. O'Connell, N. Fisk, S. Ferguson, and M. Bartlett, "How are you? Estimation of anxiety, sleep quality, and mood using computational voice analysis," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, pp. 5369–5373, 2020.

[4] N. L. Hauglund, C. Pavan, and M. Nedergaard, "Cleaning the sleeping brain–the potential restorative function of the glymphatic system," *Current Opinion in Physiology*, vol. 15, pp. 1–6, 2020.

[5] L. Xie, H. Kang, Q. Xu, M. J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D. J. Christensen, C. Nicholson, J. J. Iliff, *et al.*, "Sleep drives metabolite clearance from the adult brain," *science*, vol. 342, no. 6156, pp. 373–377, 2013.

[6] L. M. Hablitz, H. S. Vinitsky, Q. Sun, F. F. Stæger, B. Sigurdsson, K. N. Mortensen, T. O. Lilius, and M. Nedergaard, "Increased glymphatic influx is correlated with high eeg delta power and low heart rate in mice under anesthesia," *Science advances*, vol. 5, no. 2, p. eaav5447, 2019.

[7] A. Iranzo, J. L. Molinuevo, J. Santamaría, M. Serradell, M. J. Martí, F. Valldeoriola, and E. Tolosa, "Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study," *The Lancet Neurology*, vol. 5, no. 7, pp. 572–577, 2006.

[8] M. Icht, G. Zukerman, S. Hershkovich, T. Laor, Y. Heled, N. Fink, and L. Fostick, "The "Morning Voice": The Effect of 24 Hours of Sleep Deprivation on Vocal Parameters of Young Adults," *Journal of Voice*, vol. 34, no. 3, pp. 489.e1–489.e9, 2018.

[9] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22199–22208, 2017.

[10] J. Rusz, J. Hlavnicka, T. Tykalova, M. Novotny, P. Dusek, K. Sonka, and E. Ruzicka, "Smartphone Allows Capture of Speech Abnormalities Associated with High Risk of Developing Parkinson's Disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 8, pp. 1495–1507, 2018.

[11] J. R. Orozco-Arroyave, *Analysis of speech of people with Parkinson's disease*, vol. 41. Logos Verlag Berlin GmbH, 2016.

[12] J. Garefelt, S. Gershagen, G. Kecklund, H. Westerlund, and L. G. Platts, "How does work impact daily sleep quality? a within-individual study using actigraphy and self-reports over the retirement transition," *Journal of Sleep Research*, p. e13513, 2021.

[13] O. A. Famodu, M. L. Barr, I. Holásková, W. Zhou, J. S. Morrell, S. E. Colby, and M. D. Olfert, "Shortening of the pittsburgh sleep quality index survey using factor analysis," *Sleep disorders*, vol. 2018, 2018.

[14] D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The pittsburgh sleep quality index: a new instrument for psychiatric practice and research," *Psychiatry research*, vol. 28, no. 2, pp. 193–213, 1989.

[15] M. Goelema, M. Regis, R. Haakma, E. Van Den Heuvel, P. Markopoulos, and S. Overeem, "Determinants of perceived sleep quality in normal sleepers," *Behavioral sleep medicine*, vol. 17, no. 4, pp. 388–397, 2019.

[16] F. Amato, L. Borzì, G. Olmo, J. Rafael, and O. Arroyave, "An algorithm for Parkinson ' s disease speech classification based on isolated words analysis," *Health Information Science and Systems*, 2021.

[17] J. S. Carpenter and M. A. Andrykowski, "Psychometric evaluation of the pittsburgh sleep quality index," *Journal of psychosomatic research*, vol. 45, no. 1, pp. 5–13, 1998.
[18] C. Franceschini, A. Musetti, C. Zenesini, L. Palagini, S. Scarpelli, M. C. Quattropani, V. Lenzo, M. F. Freda, D. Lemmo, E. Vegni, *et al.*, "Poor sleep quality and its consequences on mental health during the covid-19 lockdown in italy," *Frontiers in psychology*, vol. 11, p. 3072, 2020.
[19] A. L. Stewart and J. E. Ware, *Measuring functioning and well-being: the medical outcomes study approach*. duke university Press, 1992.
[20] S. Lemola, T. Ledermann, and E. M. Friedman, "Variability of sleep duration is related to subjective sleep quality and subjective well-being: an actigraphy study," *PloS one*, vol. 8, no. 8, p. e71292, 2013.