Statistical Analysis of Distribution-based Spectral Features of Low-Sampled Snoring Vibrations in Predicting Treatment Outcomes in Obstructive Sleep Apnea

Behrad TaghiBeyglou[1,2], Tasnia Kamal[3], Fernanda Almeida[3] and Azadeh Yadollahi[1,2]

[1] Institute of Biomedical Engineering, University of Toronto, Toronto, Canada
[2] KITE Research Institute, Toronto Rehabilitation Institute – University Health Network, Toronto, Canada
[3] Faculty of Dentistry, University of British Columbia, Vancouver, Canada

Abstract— **This study delves into the statistical significance of spectral characteristics of snoring vibrations in the prediction of the efficacy of oral appliance devices in the treatment of obstructive sleep apnea. By analyzing data from 20 participants who underwent at-home sleep apnea testing both before and after a 5-month utilization of mandibular advancement devices, we established that specific distribution-based descriptive spectral features can predict the efficacy of oral appliances. Our analysis revealed that among 20 highly correlated features from an initial set of 192 features, only two features are significantly different between the group of responders and non-responders. Using these two features and a linear regression model, a predictive accuracy of 75%, coupled with a sensitivity of 67% and specificity of 82% was achieved. Our findings are also aligned with previous clinical outcomes on the snoring sounds, which share a lot of similarities with the snoring vibration signals.**

Keywords— **Sleep apnea, Oral appliances, Acoustic analysis, Snoring vibration, Statistical analysis**

I. Introduction

Obstructive sleep apnea (OSA) is characterized by frequent interruptions in breathing during sleep [1], with its prevalence estimated to range between 9 and 38% in the general adult population [2].

Elevated occurrences of OSA (ranging from 30 to 70%) are notably linked to advanced age (above 50 years), as well as conditions such as lung disease, heart disease, diabetes, substance use, and smoking [1]. Additionally, OSA shows strong associations with obesity and abnormalities in the upper airway.

Various treatment modalities have been proposed for OSA, encompassing behavioral interventions, surgical procedures, and medical devices. Among these, recent advancements in medical devices have exhibited a notable balance between efficacy and adherence. Medical devices primarily fall into two categories: positive airway pressure (PAP) devices and oral appliances [3, 4].

PAP devices function by delivering continuous or intermittent positive air pressure via a nasal or oral mask, effectively keeping the upper airways open during sleep. In contrast, oral appliances are customized devices designed to reposition the jaw or tongue, thereby preventing airway obstruction. While PAP devices have demonstrated higher efficacy rates (around 75% vs. 50% for oral appliances), patient adherence to oral appliances is higher (around 75% vs. 50% for PAP). Consequently, sleep physicians lean toward prescribing oral appliances; however, a significant barrier is the adjustment period required, typically lasting between 3 to 6 months, coupled with the cost of the device. As a result, physicians typically recommend oral appliances only when confident about the effectiveness of the treatment on the patient.

Numerous studies have illustrated that assessing upper airway collapsibility using various methods, such as multisensor-catheter awake nasendoscopy [5], drug-induced sleep endoscopy (DISE) [6], and craniofacial cephalometry [7, 8], can predict the response to oral appliance treatment with an accuracy ranging between 70% and 75%. Additionally, Vena et al. [9] introduced a polysomnography-based approach achieving a predictive efficacy of 77%. However, these techniques, while effective, disrupt normal sleep breathing patterns and necessitate specialized expertise.

Our recent research [10] demonstrated that employing time-frequency analysis of snoring vibrations and incorporating numerous acoustical features allowed prediction of efficacy with an accuracy of 88%. Nonetheless, the rationale behind the selection of these features and their interpretation remains unclear. Hence, the primary objective of this study is to assess the statistical significance of features employed in prior predictive models and ascertain the model's ability to predict treatment outcomes using spectrally significant features in a statistically robust manner.

II. Materials and Methods

A. Data acquisition

This study enrolled individuals referred to the Dental Sleep Apnea Clinic at the Faculty of Dentistry, University of British Columbia for oral appliance therapy. Participants underwent both baseline and follow-up sleep studies after five months of oral appliance use using a level-III portable home sleep apnea test device. This device captured physiological signals, including nasal airflow measured via a nasal cannula connected to a pressure transducer, thoracic and abdominal movements recorded through respiratory impedance plethysmography, and oxygen saturation levels monitored using pulse oximetry (SpO_2). The device also detected snore flow vibrations with a high-pass filter on the nasal pressure signal and captured snore flow audio through a forehead-connected microphone. Additionally, heart rate and body position were monitored.

Apnea was defined as the cessation of airflow lasting 10 seconds or more, and hypopnea as a 50% or greater reduction in airflow for at least 10 seconds, accompanied by a drop in SpO_2 of 3% or more. Snoring signals were sampled at 125 Hz and snoring segments were identified based on a 70% intensity threshold and verified audibly by a qualified sleep technician later. Participants were categorized as responders if their Apnea-Hypopnea Index (AHI), defined as the average number of apneas and hypopneas per hours of sleep after treatment ($AHI_{\text{post-titration}}$) was below 10 events per hour of sleep and displayed a minimum 50% reduction in AHI compared to their baseline (AHI_{baseline}).

B. Feature extraction

In prior work, we demonstrated the predictive capacity of spectral features in treatment success [10]. Building on this, our current study employs analogous methodology to extract spectral features. These include Mel-frequency cepstrum coefficients (MFCCs) alongside their first and second derivatives, chroma features, and spectral attributes such as spectral contrast, median of spectral centroid, bandwidth, flatness, and roll-off.

Furthermore, our analysis encompasses six band power levels and power spectral density (PSD) attributes like mean PSD, root-mean-square of PSD, skewness, range, and different summation moving average (DSMA) of PSD. The extraction of spectral features involved a 25-sample Hann window (200 milliseconds), with N_{fft} set to 2048. Additionally, bandpower features underwent normalization based on the segment's power within the 0.1 to 60 Hz range.

Given the inconsistent number of snoring segments across participants, we adopted an approach utilizing low-level descriptors—mean, variance, kurtosis, and skewness—to represent each feature across all snoring segments per participant. This methodology expanded the initial 73 features to a total of 292 features for each participant's data. The feature extraction was applied consistently across all participants, resulting in the feature matrix.

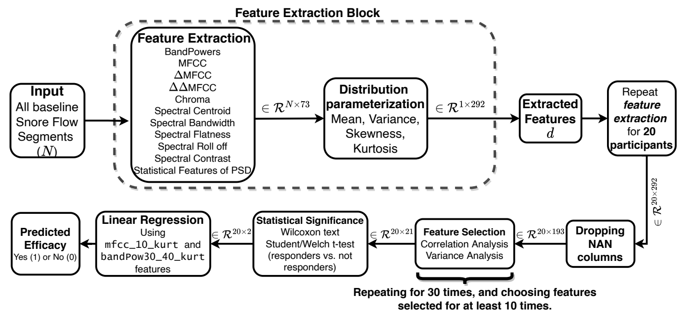

Fig. 1: The proposed study pipeline.

C. Feature preparation

After achieving the feature matrix, invalid features (designated as 'not a number' (NAN) values) were systematically removed, resulting in 192 remaining features. To refine feature selection, we employed the Searching for Uncorrelated List of Variables (SULOV) method [11], utilizing a correlation threshold of 0.6 to eliminate less significant features.

The feature matrix underwent random division, allocating 70% of the data to a training dataset for the application of the feature selection algorithm. This process was iterated 30 times, involving randomly shuffled combinations of participant data, and features that were selected at least 10 times out of the 30 repetitions were subsequently chosen for further analyses.

D. Statistical analyses and inference

In line with the primary objective of this study, our methodology proceeded as follows: initially, we employed the Wilcoxon statistical test to assess the normality of each feature's distribution within each group. Subsequently, if the data demonstrated normal distribution in both groups, we proceeded to employ the unpaired Student t-test to investigate differences between the two groups. Conversely, if the data did not exhibit normal distribution in both groups, we opted for the Welch t-test for the same purpose. Afterwards, significant features were identified for the subsequent stage of model development.

In the model development phase, we employed a linear regression model to predict treatment response, defined as 0 for non-responders and 1 for responders. As the linear model's output is not necessarily binary, a threshold of 0.5 was used to categorize individuals with a predicted output of ≥ 0.5 as responders and those with < 0.5 as non-responders. Subsequently, we computed the confusion matrix and determined the accuracy, sensitivity, and specificity metrics. Moreover, for all statistical analyses in this study, a significance level (p-value) of 0.05 was adopted. The summary of this study is illustrated in Fig. 1.

III. Results

This study comprised 20 individuals. More information about the participants' demographics can also be found in Table 1. Figure 2 also presents a scatter plot depicting the relationship between baseline and post-treatment AHI for both responders and non-responders. The feature reduction stage resulted in

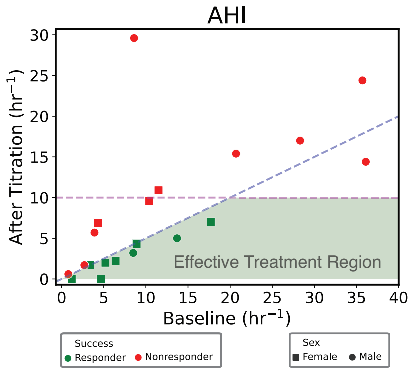

Fig. 2: Scatterplot of AHI with the region of effective treatment highlighted in green shading. The diagonal blue dashed line represents a 50% reduction post-titration, while the horizontal pink dashed line indicates an AHI of 10 events per hour after titration. Points are differentiated by sex using distinct shapes, while treatment success is indicated by different colours.

the selection of only 21 features, encompassing all the features detailed in typewriter font within Table 1, along with $\text{AHI}_{\text{baseline}}$. Features represented in typewriter font follow the format of "X_Y_Z," where "X" denotes the feature category (e.g., MFCC, chroma, spectral), "Y" indicates the feature's coefficient number or spectral characteristic (e.g., roll-off, contrast, bandwidth), and "Z" represents the utilized low-level descriptors such as kurtosis ("kurt"), skewness ("skew"), mean, and variance ("var").

However, among these features, only two features, namely `bandPow30_40_kurt` and `mfcc_10_kurt`, denoting the kurtosis of the band power between 30 to 40 Hz and the 10^{th} coefficient of the MFCC, displayed significant differences between the two groups, showcasing p-values of 0.006 and 0.031, respectively.

The linear model was constructed using `bandPow30_40_kurt` and `mfcc_10_kurt`, as detailed in Table 2, yielding an overall p-value of 0.007 and an R^2 value of 0.44. The performance of the predictive model is visually represented in Fig. 3, exhibiting an accuracy of 0.75, sensitivity of 0.67, and specificity of 0.82. Also, from Table 2, it can be inferred that the `bandPow30_40_kurt` feature is more significant when it comes to the prediction. This simple yet statistically meaningful predictive model, namely the

Table 1: Summary of the demographics and incorporated features in the 3 analyses. The columns corresponding to significantly different features are bolded.

Parameter	Overall ($n=20$)	Not Responder ($n=11$)	Responder ($n=9$)	p-value
Sex, Female	10	3	7	-
Age (years)	56 [49,56.25]	52.0 [50,56]	56 [49,57]	0.798
$\text{AHI}_{\text{baseline}}$	8.55 [4.20,14.70]	6.4 [4.7,8.9]	10.4[4.1,24.5]	0.127
$\text{AHI}_{\text{post-titration}}$	**5.35 [1.93,11.78]**	**2.2 [1.7,4.3]**	**10.9 [6.3,16.2]**	**0.006**
`bandPow0.1_10_kurt`	25.3 [15.65,38.93]	38.27 [22.23,47.03]	21.1 [12.76,29.71]	0.08
`bandPow20_30_kurt`	1.88 [1.02,3.23]	1.5 [0.36,3.2]	2.7 [1.62,3.18]	0.331
`bandPow30_40_kurt`	**4.45 [2.85,6.38]**	**6.27 [4.15,9.02]**	**3 [1.28,4.88]**	**0.006**
`bandPow50_60_var`	23.01 [13.83,30.85]	29.94 [19.35,31.22]	14.84 [12.72,28.24]	0.089
`chroma5_mean`	0.48 [0.45,0.49]	0.47 [0.47,0.49]	0.48 [0.45,0.49]	0.776
`mfcc_10_kurt`	**0.66 [0.23,0.9]**	**0.23 [0.16,0.81]**	**0.84 [0.51,1.08]**	**0.031**
`mfcc_12_kurt`	1.1 [0.79,1.68]	0.81 [0.58,1.78]	1.16 [0.92,1.65]	0.97
`mfcc_2_skew`	-0.09 [-0.19,0.02]	-0.08 [-0.28,-0.01]	-0.11 [-0.16,0.05]	0.619
`mfcc_3_skew`	0.09 [-0.07,0.32]	0.2 [-0.09,0.32]	0.08 [-0.04,0.29]	0.82
`mfcc_4_kurt`	0.51 [0.28,0.57]	0.53 [0.21,0.6]	0.5 [0.43,0.55]	0.457
`mfcc_6_kurt`	0.7 [0.26,0.81]	0.71 [0.6,0.79]	0.52 [0.21,0.8]	0.295
`mfcc_7_skew`	0.01 [-0.02,0.17]	0 [-0.01,0.09]	0.01 [-0.05,0.21]	0.882
`mfcc_9_kurt`	0.36 [0.11,1.03]	0.21 [0.08,0.55]	0.71 [0.21,1.14]	0.224
`mfcc_9_skew`	0.03 [-0.11,0.24]	0.04 [-0.1,0.22]	-0.02 [-0.1,0.25]	0.961
`spec_contrast2_kurt`	1.31 [1.1,1.91]	1.3 [1.07,1.8]	1.33 [1.12,1.93]	0.946
`spec_contrast3_kurt`	0.41 [0.09,0.58]	0.44 [0.39,0.65]	0.23 [-0.05,0.52]	0.373
`spec_contrast3_mean`	7.42 [7.23,7.74]	7.38 [7.15,7.78]	7.46 [7.26,7.69]	0.875
`spec_contrast4_kurt`	-0.26 [-0.45,-0.18]	-0.28 [-0.4,-0.24]	-0.23 [-0.48,-0.15]	0.932
`spec_contrast4_var`	14.67 [13.92,15.13]	14.25 [13.79,15.43]	14.67 [14.3,15.03]	0.851
`spec_rolloff_kurt`	0.95 [0.46,1.9]	0.67 [0.29,1.85]	1.57 [0.83,1.93]	0.561

following equation, demonstrates acceptable performance:

$$\text{Treatment}_{\text{Success}} \sim (-0.34 \times \texttt{mfcc\_10\_kurt} + 0.07 \times \texttt{bandPow30\_40\_kurt} + 0.29) \geq 0.5 \quad (1)$$

Table 2: The summary of the fitted linear regression model, where significant independent variables is shown in bold fonts.

Independent Var.	Coefficient	p-value
`mfcc_10_kurt`	-0.34	0.115
`bandPow30_40_kurt`	**0.07**	**0.023**

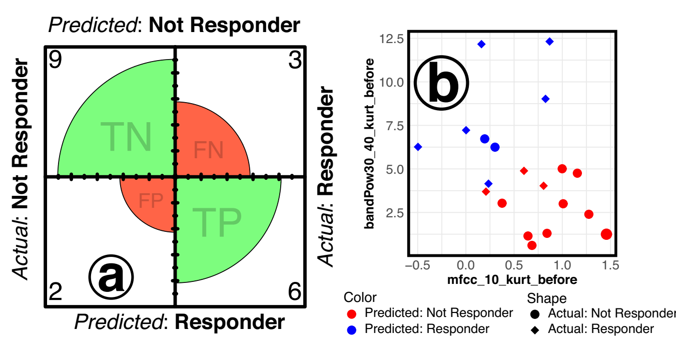

Fig. 3: Performance of the predictive model in a) confusion matrix, and b) the pair-wise illustration of the selected features and their corresponding predicted label (shown in two different colors) and their actual values (decoded in distinct shapes).

IV. Discussion

The primary objective of this study was to delve deeper into spectral analyses of snoring vibrations as a potential tool for predicting and distinguishing between responders and non-responders to oral appliance therapy.

Our findings revealed that a simple linear model leveraging only two features—specifically, `bandPow30_40_kurt` and `mfcc_10_kurt`—achieved a notable accuracy of 75% in predicting the efficacy of oral appliance therapy.

Moreover, our investigation highlighted `bandPow30_40_kurt` as the most pertinent feature, reflecting the tailedness of spectral power distribution within snoring segments between 30 to 40 Hz. This frequency range aligns with high-frequency resonances during snoring, often indicative of upper airway characteristics. Previous studies have suggested that individuals with apnea exhibit pronounced high-frequency formants in their power spectrum (compared to control groups [12]). This observation indicates heightened spectral activity in higher frequencies among individuals with OSA. Therefore, the efficacy of oral appliance therapy, potentially attributed to upper airway modifications, appears to be reflected in the upper-frequency spectrum, which is a noteworthy finding of this study.

Moreover, in contrast to previous studies highlighted in the literature, such as the work by Huntley et al. [13], which utilized DISE on 40 participants achieving a sensitivity of 75% and specificity of 50%, Vena et al. [9] achieved a sensitivity of 70% and specificity of 78% on 108 individuals using airflow recorded during full PSG, and our prior research [10] demonstrated a sensitivity and specificity of 88% based on a cohort of 50 participants using the time-frequency analysis of snoring vibrations. However, our approach in this study differed significantly; it centred more on a statistical analysis aimed at constructing a straightforward and easily interpretable predictive model.

While this research offers valuable insights within the field, it is essential to acknowledge several limitations that necessitate attention in forthcoming studies. This study is based on a relatively small sample size and a limited age group, which can question the generalizability of the findings. Furthermore, the proposed approach is sensitive to the feature selection method. Also, this study primarily focused on spectral features due to their proven efficacy in snoring analysis, while incorporating temporal features could further enhance the analysis. Therefore, future studies could explore alternatives to address these limitations.

In summary, our study delved into the viability of employing statistical analyses to identify the most effective features for straightforward prediction of OSA treatment through snoring analysis. Our findings underscore that utilizing just two simple features can yield a prediction accuracy as high as 75%.

Acknowledgements

B. Taghibeyglou received partial support through grants from the Canadian Institutes of Health Research (CIHR) and doctoral scholarships from AI4PH, a CIHR-funded training program, CaRDM Eq, a research and training program funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Institute of Biomedical Engineering at the University of Toronto.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. D. J. Gottlieb and N. M. Punjabi, "Diagnosis and management of obstructive sleep apnea: a review," *Jama*, vol. 323, no. 14, pp. 1389–1400, 2020.
2. C. V. Senaratna, J. L. Perret, C. J. Lodge, A. J. Lowe, B. E. Campbell, M. C. Matheson, G. S. Hamilton, and S. C. Dharmage, "Prevalence of obstructive sleep apnea in the general population: a systematic review," *Sleep medicine reviews*, vol. 34, pp. 70–81, 2017.
3. P. A. Cistulli, J. Armitstead, J.-L. Pepin, H. Woehrle, C. M. Nunez, A. Benjafield, and A. Malhotra, "Short-term cpap adherence in obstructive sleep apnea: a big data analysis using real world data," *Sleep medicine*, vol. 59, p. 114, 2019.
4. F. R. Almeida and A. A. Lowe, "Principles of oral appliance therapy for the management of snoring and sleep disordered breathing," *Oral and Maxillofacial Surgery Clinics*, vol. 21, no. 4, pp. 413–420, 2009.
5. K. Okuno, Y. Sasao, K. Nohara, T. Sakai, B. T. Pliska, A. A. Lowe, C. F. Ryan, and F. R. Almeida, "Endoscopy evaluation to predict oral appliance outcomes in obstructive sleep apnoea," *European Respiratory Journal*, vol. 47, no. 5, pp. 1410–1419, 2016.
6. A. V. Vroegop, O. M. Vanderveken, M. Dieltjens, K. Wouters, V. Saldien, M. J. Braem, and P. H. Van De Heyning, "Sleep endoscopy with simulation bite for prediction of oral appliance treatment outcome," *Journal of sleep research*, vol. 22, no. 3, pp. 348–355, 2013.
7. A. T. M. Ng, M. A. Darendeliler, P. Petocz, and P. A. Cistulli, "Cephalometry and prediction of oral appliance treatment outcome," *Sleep and Breathing*, vol. 16, pp. 47–58, 2012.
8. H.-L. Shen, Y.-W. Wen, N.-H. Chen, and Y.-F. Liao, "Craniofacial morphologic predictors of oral appliance outcomes in patients with obstructive sleep apnea," *The Journal of the American Dental Association*, vol. 143, no. 11, pp. 1209–1217, 2012.
9. D. Vena, A. Azarbarzin, M. Marques, S. Op de Beeck, O. M. Vanderveken, B. A. Edwards, N. Calianese, L. B. Hess, R. Radmand, G. S. Hamilton *et al.*, "Predicting sleep apnea responses to oral appliance therapy using polysomnographic airflow," *Sleep*, vol. 43, no. 7, p. zsaa004, 2020.
10. B. TaghiBeyglou, T. Kamal, F. Almeida, and A. Yadollahi, "Predicting Treatment Outcomes in Obstructive Sleep Apnea: A Distribution-based Spectral Analysis of Low-Sampled Snoring Vibrations," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2023.
11. S. Ramírez-Gallego, I. Lastra, D. Martínez-Rego, V. Bolón-Canedo, J. M. Benítez, F. Herrera, and A. Alonso-Betanzos, "Fast-mrmr: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," *International Journal of Intelligent Systems*, vol. 32, no. 2, pp. 134–152, 2017.
12. A. K. Ng, T. San Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?" *Sleep medicine*, vol. 9, no. 8, pp. 894–898, 2008.
13. C. Huntley, J. Cooper, M. Stiles, R. Grewal, and M. Boon, "Predicting success of oral appliance therapy in treating obstructive sleep apnea using drug-induced sleep endoscopy," *Journal of Clinical Sleep Medicine*, vol. 14, no. 8, pp. 1333–1337, 2018.