2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Unsupervised Approach for the Identification of the Predominant Site of Upper Airway Collapse in Obstructive Sleep Apnoea Patients Using Snore Signals

Arun Sebastian *Student Member, IEEE*, Peter A. Cistulli, Gary Cohen and Philip de Chazal *Senior Member, IEEE*

Abstract— **Knowledge regarding the site-of-collapse in the upper airway in obstructive sleep apnoea (OSA) has implications for treatment options and their outcomes. However, current methods to identify the site-of-collapse are not suitable for clinical practice due to the invasive nature, the time/cost of the tests and the inconsistency of the obstruction site identified with natural and drug-induced sleep. In this study, we adopted an unsupervised algorithm to identify the predominant site-of-collapse of the upper airway during natural sleep using nocturnal audio recordings. Nocturnal audio was recorded together with full-night polysomnography using a ceiling microphone. Various acoustic features of the snore signal during hypopnoea events were extracted. We developed a feature selection algorithm combining silhouette analysis with the Laplacian score algorithm to select the high performing features. A k-means clustering model was developed to form clusters using the features extracted from snore data and analyse the correlation between the clusters generated and the predominant site-of-collapse. Cluster analysis showed that the data tends to fit well in two clusters with a mean silhouette coefficient of 0.79 and with an accuracy of 68% for classifying tongue/non-tongue collapse. The results indicate a correlation between snoring and the predominant site-of-collapse. Therefore, it could potentially be used as a practical, non-invasive, low-cost diagnosis tool for improving the selection of appropriate therapy for OSA patients without any additional burden to the patients undergoing a sleep test.**

I. INTRODUCTION

Conventional treatments for OSA include positive airway pressure devices, oral appliance (OA) therapy, and surgery [1, 2]. The choice of therapy is guided by the severity of OSA and patient preference. Even though different treatments are available, efficacy is highly variable between patients, creating the need to predict treatment outcome [3]. Studies have shown that information regarding airway obstruction site plays a role in predicting treatment outcome and, therefore, helps clinicians choose the most appropriate treatment. This is especially true for patients who have “tongue-base” airway collapse, as they appear to be more likely to gain a large therapeutic benefit from OAs [4, 5].

Conventional methods in determining the site-of-collapse involve using an endoscope or a pressure catheter during drug-induced or natural sleep [4, 5]. Unfortunately, these methods are not well tolerated by patients due to their invasive nature, time/expense of the tests, and the inconsistency of the obstruction site identified with natural sleep and drug-induced sleep, limiting their clinical application [6, 7].

Previous studies have shown that acoustic analysis of snoring has been successfully implemented in the diagnosis and estimation of the severity of OSA [8, 9]. However, a limited number of studies have been conducted to determine the relationship between snoring and obstruction sites. The common method, VOTE classification (velum (V), oropharyngeal area (O), tongue (T), and epiglottis (E)) has been explored by researchers utilising different classification algorithms and signal processing techniques to categorise the snore signal [10, 11]. However, studies have demonstrated inconsistencies regarding the obstruction site identified in terms of natural and drug-induced sleep [6], and acoustic properties of snoring differ significantly from natural sleep [7]. Another drawback was that all the studies were based on analysing a single snore episode and thus were unable to determine the predominant site-of-collapse over an entire night.

Our previous study has demonstrated that snore during hypopnoea can be used to identify the predominant site-of-collapse for a patient [12-14]. A classification model was adopted to predict the predominant site-of-collapse in the upper airway and achieved an accuracy of 81% for discriminating tongue/non-tongue collapse [12]. One of the limitations of this study was that a gold standard method was not used to identify the site-of-collapse, and the ground truth for the labelling employed was based on an indirect method that manually identifies the site-of-collapse based on airflow signal. This bypassed the clinical limitations of more invasive methods but at the expense of potentially introducing some labelling errors. We conducted another study to find the correlation between clusters generated using snore features and the site-of-collapse [15]. The result indicates that acoustic features of snore data tend to form clusters based on the site-of-collapse. This supports the role of the snore signal as a correlate of the site-of-collapse.

Research supported by University of Sydney
Arun Sebastian and Philip de Chazal are with the Sleep Research Group, Charles Perkins Centre, School of Biomedical Engineering, University of Sydney, Sydney, NSW, 2006, Australia {arun.sebastian,philip.dechazal}@sydney.edu.au
Peter A. Cistulli is with the Charles Perkins Centre and Northern Clinical School, University of Sydney, Sydney, NSW, 2006, Australia peter.cistulli@sydney.edu.au
Gary Cohen is with the Sleep Investigation Laboratory, Center for Sleep Health and Research, Royal North Shore Hospital, Sydney, NSW, 2065, Australia Gary.Cohen@health.nsw.gov.au

Motivated by the possibilities of an unsupervised method in resolving the challenges in labelling, the objective in the current study was to introduce a clustering algorithm to identify predominant site-of-collapse for an OSA patient and compare its performance with the classification model.

II. METHOD

A. Data Collection

Nocturnal audio data from 58 patients, who had attended a full night sleep study at the Sleep Investigation Unit, Royal North Shore Hospital Sydney, and received a diagnosis of OSA (AHI$\geq$5) were used for this study. The audio signal was recorded with a Condenser Microphone, which is placed on the ceiling approximately $1.75m$ above the patient's bed. The audio signal was sampled at a frequency of 32 kHz with a resolution of 16-bits per sample. This study received ethical approval by the Northern Sydney Local Health District Human Research Ethical Committee as application RESP/18/184.

B. Data Labelling

In this study, we adopted an indirect process for labelling the site-of-collapse using the airflow signal shape (flattening or scooping of airflow contour) as shown in Fig. 1 [16], similar to our previous studies [12-15]. Based on this, we generated a database of 1807 hypopnoea events (881 non-tongue related and 926 tongue related collapse). In the next step, the predominant site-of-collapse for each patient was identified by adopting a similar rule implemented by Xu et al. (if more than 60% of the total events were the same event type, then the predominant site-of-collapse for a patient was set to the majority type) [17]. This process resulted in 26 (45%) patients labelled as non-tongue and 32 (55%) patients as tongue-based collapse. This labelling information was used to evaluate the model performance (external cluster validation).

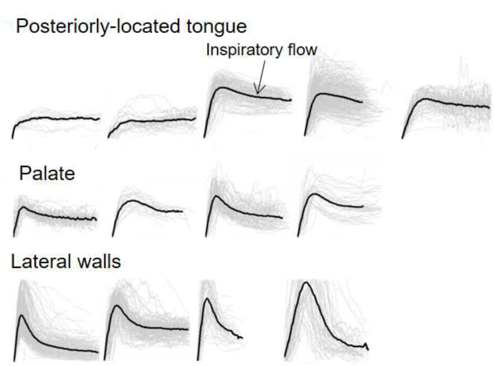

Fig. 1. Identifying site-of-collapse using thirteen reference nasal pressure signals [16]. The darker tracing represents the ensemble average inspiratory flow.

C. Signal Processing

1) Preprocessing: As the audio signal was recorded using a general-purpose microphone, a spectral subtraction method was deployed to remove background noise from the audio recordings to improve the signal-to-noise ratio.

2) Feature Extraction: Fifty identical features were derived from each hypopnoea event. The time-domain features were: (1) energy, (2) entropy, and (3) ZCR. The frequency-domain feature consisted of (1) first three formant frequencies, (2) thirteen MFCC and its first derivatives, (3) twelve spectral chroma features, (4) spectral entropy, (5) spectral flux, (6) spectral centroids, (7) spectral roll-off, and (8) fundamental frequency and harmonic frequency. Further details on the signal processing methods can be found in [12].

D. Cluster Analysis

1) K-means clustering: K-means clustering is a simple, fast and efficient data clustering algorithm that works iteratively to allocate each data point to one of k subgroups (to the nearest cluster) based on the similarity of features provided [18].

2) Silhouette analysis: Silhouette analysis can be used to analyse the separation distance between the clusters generated. It can also be used to validate the performance of the clusters by measuring how well a data point fits into its own cluster (cohesion) compared with other clusters (separation) [19].

3) Laplacian Score: Laplacian Score is a widely used feature ranking algorithm used in unsupervised learning to select the most important features to build the model [20]. Laplacian Score evaluates the importance of features according to their locality preserving power and ranks the features based on the score.

4) Optimal Cluster Number: To find the optimal cluster number, the silhouette values for each feature and all features together were determined by varying the number of clusters, ranging from two to six. The optimal cluster number for a particular feature is the number with the maximum silhouette value. Finally, the optimal cluster number of the model was identified using the majority principle-that is, selecting the most common cluster number from individual feature results.

5) Feature Selection: We developed a feature selection algorithm combining silhouette analysis and the Laplacian score algorithm. The algorithm consisted of the following steps

- Determine the optimal number of clusters for the model using all of the features separately, as described in the previous section.
- Using the optimal cluster number, evaluate the mean silhouette value and silhouette plot for each feature.
- Select the features with a uniform cluster thickness using the silhouette plot. The thickness of the silhouette plot quantifies the cluster size and it is calculated by the identifying the number of data points in a cluster. For our purposes, we selected features where the number of data points in a cluster was less than 60

- Rank the selected features from the previous step based on the importance using Laplacian scores.
- Evaluate the performance of feature combinations by sequentially adding features based on the rank of the features from the previous step and calculating the mean silhouette value.
- Repeat the previous step until the mean silhouette value reaches the first maximum. The feature set associated with this maximum becomes the selected feature set.

III. RESULTS AND DISCUSSION

A k-means clustering algorithm was developed using the relevant features to assign snore data into clusters and investigate the correlation between the predominant site-of-collapse and the clusters generated.

1) Optimal Cluster Number: The results revealed that the most common cluster was two, as the optimal cluster number for 36 out of 50 features (72%) was two. Also, the optimal cluster was two when all of the features were combined. Based on the results, the optimum cluster for the model was determined to be two.

2) Feature Selection: Using the optimal cluster number, the mean silhouette value and silhouette plot for each feature were evaluated using a 2-means clustering. The "best" features were selected based on the thickness (uniform thickness) of the silhouette plot , where a similar thickness indicates an approximately equal number of data in the cluster. Further details on the silhouette plot can be found in [15]. As the optimal cluster number was two for the model, the best features comprised of clusters with less than 60% of the total data points. Based on the number of data points in a cluster, the algorithm selected 27 features by discarding the features with the number of data points in a cluster higher than 60% of the total data points. The feature set consists of 7 MFCC coefficients, the 7 first derivatives of MFCC coefficients, 5 chroma features, the first 3 formant frequencies, energy, ZCR, spectral entropy, flux, and fundamental frequency.

In the next phase, the Laplace score was identified for the 27 selected features and the selected features were ranked based on the Laplacian score. Despite the fact that the ranking was based on the score, there was only a slight difference between the Laplace scores for all features. As a result, we assess performance by calculating the mean silhouette value by sequentially adding features based on their rank and then selecting the features until the mean silhouette value reaches the first maximum. The results showed that the maximum mean silhouette value was achieved (0.79) when the cluster generated with the first 17 features, and therefore, the 17 features were selected to build the final model. Table I shows the features that were selected.

For the final system, a 2-means cluster model was developed using the selected features. Internal cluster validation showed that the data had a strong tendency to form two clusters, with a mean silhouette value of 0.79. Additionally, 1442 (80%) data points had silhouette values higher than the mean silhouette value (0.79).

TABLE I
FINAL FEATURES SELECTED FOR THE MODEL

No	Feature
1	1^{st}, 2^{nd}, 4^{th}, 6^{th}, 9^{th} and 12^{th} Coefficient of MFCC
2	2^{nd}, 4^{th}, 5^{th}, 7^{th} and 9^{th} Coefficient of First derivative MFCC
3	1^{st}, 4^{th}, 5^{th} and 7^{th} Coefficient of spectral chroma
4	1^{st} and 2^{nd} Formant frequency

3) External Cluster Validation: External cluster validation was performed to evaluate the performance of the model using the manual labels and to analyse the correlation between the cluster formed and the predominant site-of-collapse in the upper airway. As the optimal cluster number was identified as two for the model, we related the clusters with tongue and non-tongue related collapse. Furthermore, studies have demonstrated that the information regarding tongue/non-tongue is the most important measure as the best predictor of OA success, and patients with tongue-base collapse gain a significant therapeutic benefit from OAs [4, 5].

For performance validation, cluster 1 and 2 were labelled as tongue and non-tongue base collapse, respectively, and then identifying the predominant site-of-collapse. The predominant site-of-collapse for a patient was identified using the same rule in the manual labelling of the patient (i.e., if more than 60% of total events comes in one cluster, while clusters were labelled as tongue and non-tongue base collapse). A representative example of k means clustering using the first two relevant features is given in Fig. 2. This model achieved an overall accuracy of 68% (39/58) for categorising tongue/non-tongue related collapse based on the predominant site-of-collapse. Detailed results are given in Table II.

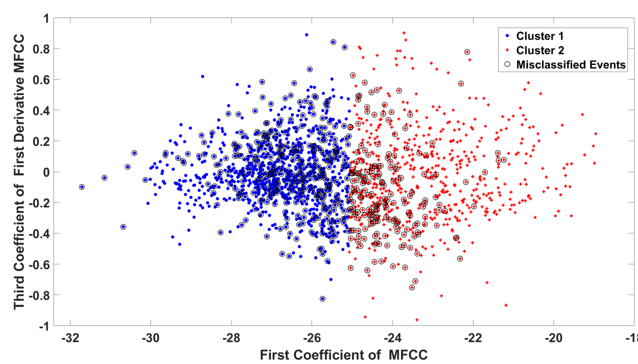

Fig. 2. 2D representation of $k-$means clustering. Clustering was done with first two features identified using Laplacian scores for the simple visualisation. For the external cluster validation, cluster 1 and 2 were labelled as tongue and non-tongue respectively. ○ represents the misclassified event based on the manual labelling.

Comparing the classification [12] and clustering model performance, the classification algorithm achieved an accuracy of 81% while the clustering model achieved an accuracy of 68%. Although the classification model outperformed the

TABLE II

CONFUSION MATRIX FOR THE PREDOMINANT SITE-OF-COLLAPSE (TONGUE/NON-TONGUE) CLASSIFICATION.

		Cluster Based Labelling	
		Non-Tongue	**Tongue**
Manual Labelling	**Non-Tongue**	**16**	10
	Tongue	9	**23**

clustering model, the results indicate that the snore signal had intrinsic properties relevant for the collapse site analysis. Moreover, there are some advantages to the clustering algorithm. Firstly, the clustering algorithm does not require labels or ground truth information, whereas the classification methods require ground truth information. Another advantage of cluster analysis is that the computation complexity is less compared with the classification method, as the classification algorithm requires a training set to tune the model.

Comparing the optimal features selected for the clustering and classification models indicates that frequency related features provide the most discriminating information regarding the site-of-collapse. As the optimal features selected predominantly comprise MFCC features (five out of seven features in classification [12] and 11 out of 17 features in clustering), MFCCs tend to play an important role in identifying the obstruction site, which has successfully been used in speech and snore analysis.

There are several limitations to this study. The main limitation arose due to the implementation of $k-$means clustering, as it is highly sensitive to initial values, and there is difficulty in finding the optimal cluster number. Another disadvantage was that our method currently requires hypopnoea event scoring, limiting its application only to PSG studies. Furthermore, the snores were chosen without regard for body posture, affecting the recording signal quality and audio frequencies. Further studies are needed to investigate the correlation between clusters generated and the snore signal, including more patient characteristics and predictions regarding treatment outcomes.

IV. CONCLUSION

This study presents an analysis of the correlation between the clusters generated using the features extracted from snore data during hypopnoea events and the site-of-collapse. The unsupervised approach used a k-means clustering to assign data into clusters and investigated the correlation between the predominant site-of-collapse and the clusters generated. A feature selection algorithm was also developed to select the most relevant features by combining silhouette analysis with the Laplacian score algorithm. The performance of the model was evaluated by comparing the automatic and manually labelled data based on the predominant site-of-collapse. Results showed that the model achieved an accuracy of 68% in labelling the participants based on the predominant site-of-collapse. The results indicate there exists a correlation between the clusters and the predominant site-of-collapse, which supports the evidence that the snore signal is useful in determining the site-of-collapse, as discussed in our previous studies.

REFERENCES

[1] T. Kirby, "Colin Sullivan: inventive pioneer of sleep medicine", *Lancet,* vol. 377, no. 9776, pp. 1485, 2011.

[2] K. Sutherland, O. M. Vanderveken, H. Tsuda, M. Marklund, et al., "Oral Appliance Treatment for Obstructive Sleep Apnea: An Update", *J Clin Sleep Med.,* vol. 10, no. 2, pp. 215-227, 2014.

[3] K. Okuno, B. T. Pliska, M. Hamoda, A. A. Lowe, and F. R. Almeida, "Prediction of oral appliance treatment outcomes in obstructive sleep apnea: A systematic review", *Sleep Med Rev.,* vol. 30, pp. 25-33, 2016.

[4] Vena D, Azarbarzin A, Marques M et al., "Predicting sleep apnea responses to oral appliance therapy using polysomnographic airflow.", *Sleep,* vol. 43, no. 7, pp. 887-895, 2020.

[5] Marques M, Genta PR, Azarbarzin A et al., "Structure and Severity of Pharyngeal Obstruction Determine Oral Appliance Efficacy in Sleep Apnoea", *J Physiol.,* vol. 597, no. 22, pp. 5399-5410, 2019.

[6] M. J. Ravesloot, and N. de Vries, "One hundred consecutive patients undergoing drug-induced sleep endoscopy: results and evaluation ", *Laryngoscope,* vol. 121, no. 12, pp. 2710-2716, 2011.

[7] S. M. Jones, M. S. Ho, J. E. Earis, A. C. Swift, and P. Charters, "Acoustic parameters of snoring sound to compare natural snores with snores during steady-state propofol sedation", *Clin Otolaryngol Allied Sci.,* vol. 31, no. 1, pp. 46-52, 2006.

[8] A. Azarbarzin, and Z. Moussavi, "Snoring sounds variability as a signature of obstructive sleep apnea", *Med Eng Phys.,* vol. 35, no. 4, pp. 479-485, 2013.

[9] J. Kim, T. Kim, D. Lee, J. W. Kim, K. Lee, "Exploiting temporal and nonstationary features in breathing sound analysis for multiple obstructive sleep apnea severity classification ", *Biomed Eng Online.,* vol. 16, no. 1, 2017.

[10] Qian K, Janott C, Pandit V et al., "Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis", *IEEE Trans Biomed Eng.,* vol. 64, no. 8, pp. 1731-1741, 2017.

[11] F. Demir, A. Sengur, N. Cummins, S. Amiriparian, B. Schuller, "Low Level Texture Features for Snore Sound Discrimination," *Conf Proc IEEE Eng Med Biol Soc.*, pp. 413-416, 2018.

[12] A. Sebastian, P. A. Cistulli, G. Cohen, P. De Chazal, Automated identification of the predominant site of upper airway collapse in obstructive sleep apnoea patients using snore signal, Physiol Meas., vol. 41, no. 9, 2020. doi: 10.1088/1361-6579/abaa33

[13] A. Sebastian, P. A. Cistulli, G. Cohen, P. De Chazal, A Preliminary Study of the Automatic Classification of the Site of Airway Collapse in OSA patients Using Snoring Signals, in Conf. Proc. IEEE Eng. Med. Biol. Soc., 2019.

[14] A. Sebastian, P. A. Cistulli, G. Cohen, P. De Chazal, Identifying the Predominant Site of Upper Airway Collapse in Obstructive Sleep Apnoea Patients Using Snore Signals, in Conf. Proc. IEEE Eng. Med. Biol. Soc., 2020.

[15] A. Sebastian, P. A. Cistulli, G. Cohen, P. De Chazal, Characterisation of Upper Airway Collapse in OSA Patients Using Snore Signals: A Cluster Analysis Approach, in Conf. Proc. IEEE Eng. Med. Biol. Soc., 2020.

[16] P. R. Genta, S. A. Sands, J. P Butler, and S. H. Loring, Airflow shape is associated with the pharyngeal structure causing obstructive sleep apnea, in J. Chest., vol. 152, no. 3, 2017.

[17] H. Xu, Sound spectral analysis of snoring sound and site of obstruction in obstructive sleep apnea syndrome, in J. Acta. Otolaryngol., vol.130, no.10, pp. 1175-1184, 2010.

[18] A. Azarbarzin, Z. Moussavi, Unsupervised classification of respiratory sound signal into snore/no-snore classes, in Conf. Proc. IEEE Eng. Med. Biol. Soc., 2010.

[19] S. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, in J. Comput. Appl. Math., vol. 20, pp. 53-65, 1987.

[20] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," *Adv Neural Inf Process Syst*, pp. 507-514, 2006.