



A Snoring Sound Dataset for Body Position Recognition: Collection, Annotation, and Analysis

Li Xiao¹, Xiuping Yang^{2,†}, Xinhong Li¹, Weiping Tu^{1,✉},
Xiong Chen^{2,✉}, Weiyang Yi¹, Jie Lin¹, Yuhong Yang¹, Yanzhen Ren³

¹NERCMS, School of Computer Science, Wuhan University

²Sleep Medicine Centre, Zhongnan Hospital of Wuhan University

³School of Cyber Science and Engineering, Wuhan University

tuweiping@whu.edu.cn, zn_chenxiong@whu.edu.cn

Abstract

Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is a chronic breathing disorder caused by a blockage in the upper airways. Snoring is a prominent symptom of OSAHS, and previous studies have attempted to identify the obstruction site of the upper airways by snoring sounds. Despite some progress, the classification of the obstruction site remains challenging in real-world clinical settings due to the influence of sleep body position on upper airways. To address this challenge, this paper proposes a snore-based sleep body position recognition dataset (SSBPR) consisting of 7570 snoring recordings, which comprises six distinct labels for sleep body position: supine, supine but left lateral head, supine but right lateral head, left-side lying, right-side lying and prone. Experimental results show that snoring sounds exhibit certain acoustic features that enable their effective utilization for identifying body posture during sleep in real-world scenarios.

Index Terms: OSAHS, Snoring sound, Sleep body position recognition

1. Introduction

Obstructive sleep apnea-hypopnea syndrome (OSAHS) is a severe chronic breathing disorder and is caused due to a blockage or collapse of the upper airways [1, 2]. The gold standard approach for diagnosing OSAHS is attended overnight polysomnography (PSG) in a sleep laboratory [3, 4]. Effective management of moderate and severe patients with OSAHS requires accurate identification of the site of upper airway obstruction. The standard diagnostic approach to determine the obstructive site is through the Drug-Induced Sleep Endoscopy (DISE) procedure [5]. This tool assesses the upper airway of OSAHS patients while simulating the conditions of natural sleep. However, this method of diagnosis is associated with several limitations, such as extended examination time and elevated costs.

Snoring is one of the most prominent symptoms of OSAHS and can be used to identify the obstructive site, including the velum, oropharyngeal lateral walls, tongue, and epiglottis [6]. Numerous studies have been conducted in this field by Munich-Passau snore sound corpus (MPSSC) [6], with a focus on the automatic classification of snoring sound excitation location [7, 8, 9, 10]. For example, Zhang et al.[8] proposed a novel data augmentation approach utilizing semi-supervised conditional generative adversarial networks. Their study demonstrates significant competitiveness compared to previous work on the same dataset, achieving an unweighted average recall

(UAR) of 67.4%. Ding et al. [10] demonstrated the efficacy of a convolutional neural network and complement-cross-entropy loss function in solving the problem of imbalance distribution of the dataset, yielding a UAR of 77.13%.

Nevertheless, automatic snoring sound excitation location classification remains challenging in a real-world clinical setting mainly because of changes in the sleep positions of patients. The patients usually lie down supine on the operating table during DISE [11], and the snoring sounds of MPSSC datasets are classified based on the simultaneous endoscopic video recordings of DISE. However, sleep body position could influence the site, direction, and severity of upper airway obstruction in patients with OSAHS [12, 13]. In the natural sleep environment, people change their sleeping position several times nightly, affecting the corresponding snoring sounds generated and their excitation location in the upper airway. Therefore, an important relationship exists between sleep posture and the snoring sounds generated. The lack of information on sleep body position could become one of the significant barriers to influencing the result of predicting obstruction sites in the upper airway only based on snoring sounds. Most of the existing research about sleep posture recognition relies upon using camera or pressure sensors [14, 15, 16]. However, they engender apprehensions about privacy and impose a financial burden associated with the acquisition and maintenance of hardware.

In this paper, a novel snore-based sleep body position recognition dataset (SSBPR) is presented, and the snoring sounds of SSBPR dataset are annotated based on simultaneous PSG signals. The aim of this study is to enhance the accuracy of predicting the site of upper airway obstruction in a clinical setting by body position information only derived from snoring sounds. Specifically, a two-stage method could be used: 1) sleep position recognition by SSBPR and 2) obstructive site identification by other datasets, such as MPSSC. Therefore, human sleep body positions which influence the upper airway are considered as much as possible in this paper, including supine, supine but left lateral head, supine but right lateral head, left-side lying, right-side lying and prone. Furthermore, motivated by the recent advancements in attention-based models in the field of computer vision and natural language processing [17, 18], the present study employs AST [19], a model based on Transformer, as the baseline method for sleep body position classification. Experimental results show that our work holds significant potential in terms of offering practical solutions that could enhance the diagnosis of OSAHS and the management of sleep posture for special populations.

To be more specific, the key contributions of our work are as follows:

- 1) We construct a snoring sound dataset named SSBPR for sleep body position recognition. The dataset can serve as a valu-

† Equal contribution

✉Corresponding author

able supplementary resource to the MPSSC and future snoring dataset of upper airway obstruction, which can improve the precision and reliability of identifying the location of upper airway obstruction.

- 2) The SSBPR dataset has potential applications beyond helping to identify upper airway obstruction sites. Specifically, SSBPR can aid in the monitoring and management of sleep body position in specific populations with snoring, such as pregnant women and patients with laryngopharyngeal reflux disease.

In the remainder of this paper, we first describe our dataset in more detail in section 2, then discuss our baseline experimental results and finally provide conclusions in section 3 and 4.

2. Datasets

The purpose of the SSBPR dataset is to provide a new benchmark and further advance the research on the automatic snoring sound excitation location classification. In this section, we describe the details of the SSBPR dataset.

2.1. Data Collection

In this study, data were collected from 20 adult patients who underwent overnight PSG at a local Sleep Medicine Research Center within the hospital. The study was conducted with the approval of the local medical ethics committee, and patients provided signed consent for their participation, including audio and video recordings during sleep. The personal information of the study subjects was collected and stored anonymously to ensure privacy protection.

The snoring audio recording was obtained using a sub-miniature lavalier microphone (Shure MX150/C, USA) driven by an audio Interface (Antelope Zen Go Synergy Core, Bulgaria), with a sampling frequency of 32 kHz and a sampling resolution of 16 bits (bit depth). The microphone was positioned on the patient's face, facing the mouth, with an approximate distance of 3 cm, as shown in Figure 1. The microphone was secured in place using medical tape, ensuring stability during the entire night. Recording snoring sounds at a distance of 3 cm offers a range of benefits in terms of capturing precise and detailed information about the sound. The proximity of the recording device enables it to capture subtle nuances in the snoring sounds, such as variations in volume and pitch, which may not be evident when recording at greater distances. This high level of detail can be beneficial for analyzing and diagnosing health issues such as sleep apnea. In addition, recording at a close distance reduces background noise, like traffic or other sounds in the house, making it result in a recording of the snoring sound that is more focused and clearer.

2.2. Data Annotation

A standard full PSG simultaneously records more than 10 physiologic signals during sleep, including electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), body position, video recording and et al. For recorded PSG data of each patient in the SSBPR dataset, sleep stages were scored by three experienced experts according to the AASM Manual V2.6 [20]. A certified technician performed the first level of scoring, and the final level was conducted by two certified doctors who verified the scoring of PSG. Both of them got RPSGT (Registered Polysomnographic Technologist) certified. Compumedics Profusion Sleeps was used to record the



Figure 1: Diagram of the patient's face frame of PSG and location of the microphone in the sleep medicine research center. Sensors attached to the patient's face and body are part of the PSG acquisition system.

data, display multiple channel recordings, manually score the data, report and export the recorded data. In half of OSAHS patients, disease severity depends on trunk position and head position, in which the supine head and trunk position is usually the worst sleeping position [21, 22]. In positional OSAHS patients, lateral head rotation alone significantly differed at all levels observed during DISE compared to lateral head and trunk rotation [12, 13]. The snoring sounds are usually influenced by changes in the structure of the upper airway. These findings underscore the importance of considering both trunk and head position in accurately monitoring the upper airway of OSAHS patients. Therefore, to effectively monitor the upper airway of patients with OSAHS, it is necessary to label snoring sounds using trunk and head position information.

In this paper, the annotation of snoring data (duration of 0.29 - 8.39 s) in natural sleep requires synchronized PSG signals: sleep stages, body position and video. The human sleep stages include wake, non-rapid eye movement (NREM) and rapid eye movement (REM) sleep, where NREM sleep encompasses three sleep stages, referred to as stage 1 (N1), stage 2 (N2), and stage 3 (N3) NREM sleep [23]. As shown in Figure 2, the watermark "N2" indicates that the patient is in non-rapid eye movement stage N2 by EEG, EOG and EMG, which is used to determine whether the patient is asleep. The position and video signal together to identify the trunk and head position of the patient when snoring. In summary, the snoring sounds in the SSBPR dataset are labeled with 6 types: supine, supine but left lateral head, supine but right lateral head, left-side lying, right-side lying and prone.

2.3. Data Distribution

Research studies have consistently demonstrated that OSAHS is more prevalent among men compared to women [24]. To ensure equitable representation of genders in the study sample, a deliberate effort was made to include an equal number of male and female patients in the dataset, with a total of 10 patients of each gender being selected. Given that age and obesity are recognized risk factors for OSAHS [25], older individuals and those with higher body mass index may exhibit a more collapsi-

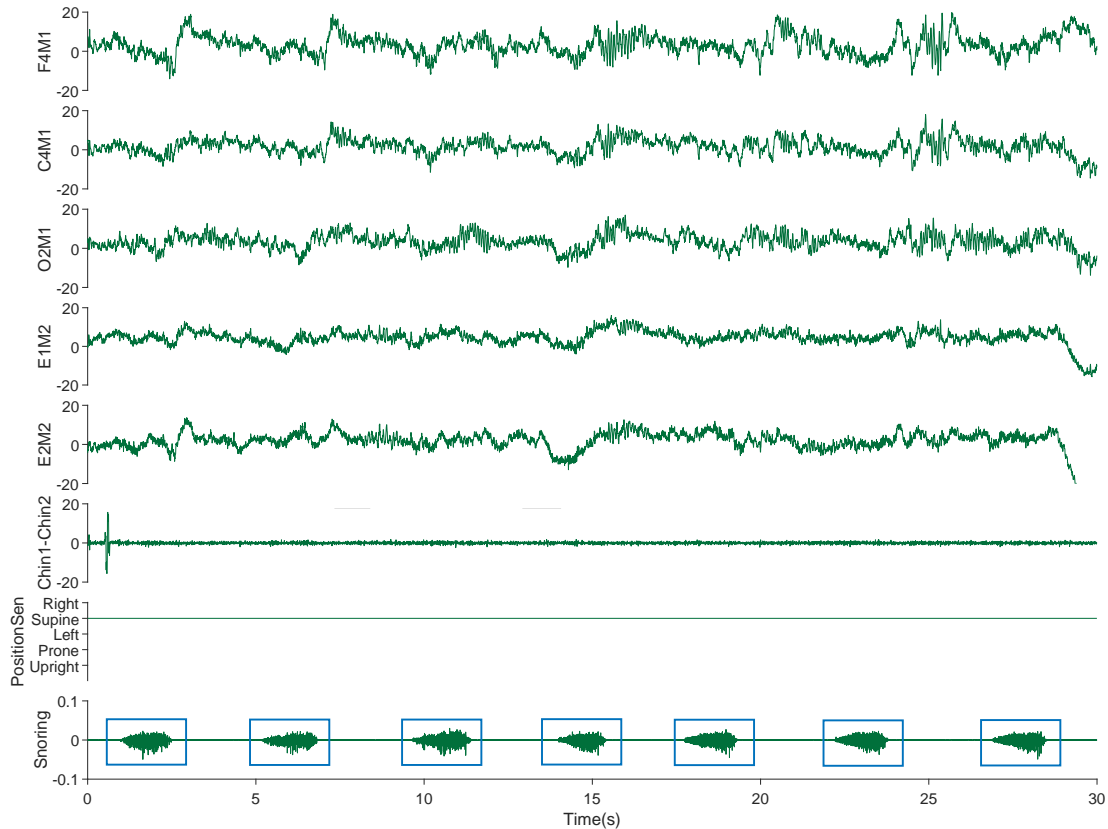


Figure 2: An example of snoring data annotation using PSG signals. EEG: F4M1, C4M1 and O2M1; EOG: E1M2 and E2M2; EMG: Chin1-Chin2; Trunk position: PositionSen. EEG, EOG and EMG are employed to determine the sleep state of the individual, while a body position sensor is utilized to ascertain the positioning of the trunk. Additionally, video monitoring is employed to evaluate the position of the patient’s head (In order to protect the patient’s privacy, the video signals are not given in this paper). Blue boxes indicate the snoring sounds during sleep while the patient is supine, and the watermark “N2” in the figure indicates that the patient is in non-rapid eye movement stage N2.

Table 1: The number of classed instances in each set of the SSBPR. The labels 0, 1, 2, 3, 4 and 5 indicate different body positions: supine, supine but left lateral head, supine but right lateral head, left-side lying, right-side lying and prone, respectively.

Labels	Train		Validation		Test		Σ
	Male	Female	Male	Female	Male	Female	
0	491	491	164	164	164	164	1638
1	436	436	146	146	145	145	1454
2	381	381	127	127	127	127	1270
3	500	500	167	167	167	167	1668
4	431	431	144	144	144	144	1438
5	61	0	20	0	21	0	102
Σ	2300	2239	768	748	768	747	7570

ble airway during sleep. Consistent with this notion, the sample population comprises 20 individuals aged 26 to 57 years, with a mean age of 43.1. The patients in this study exhibit a mean body mass index (BMI) of 26.57 kg/m^2 . The SSBPR dataset contains a total of 7570 snoring samples. Both genders had equal snoring sound samples except for prone labeling because female patients did not appear prone during PSG monitoring.

Detailed data split information can be found in Table 1.

3. Baseline Experiments

3.1. Baseline Method

The Transformer architecture has gained widespread usage in many areas, including natural language processing, image processing and audio processing [17, 18, 19]. This popularity can be attributed to the Transformer’s capability to effectively model long-term dependencies, which are crucial in numerous tasks that involve sequences. Snoring signals possess complex dependencies between various temporal steps due to changes in the upper airway. The self-attention mechanism of the Transformer facilitates the weighing of the contribution of each temporal step in predicting the current step, thereby capturing the inter-temporal relationships effectively of snoring sounds.

For baseline experiments, we use an AST [19] audio classifier, which has the same Transformer architecture as Vision Transformer (ViT) [26]. AST is the first convolution-free, a purely attention-based model for audio classification, which supports variable length input and can be applied to various tasks. In this paper, AST partitions the 2D snoring sound spectrogram into a sequence of 16×16 patches with overlapping areas. Each patch is then linearly transformed into a sequence

Table 2: Five-classes and six-classes sleep body position classification results.

#	Gender	Accuracy (%)
Five-classes	Male	82.7
	Female	94.6
Six-classes	/	85.8

of 1-D patch embeddings, augmented with learnable positional embeddings. The resulting sequence is then prepended with a classification token before input into a Transformer. The classification token output is obtained by applying a linear layer to the output embedding, which is used for the final classification task.

3.2. Experiment Settings

Model inputs. Snoring sound and speech exhibit significant acoustic similarities in their generation and emission [27]. Both are produced in the vocal tract through airflow vibration and are acoustically shaped by the frequency transfer function of the upper airway, before being emitted through the mouth and nose. Given that certain acoustic features have proven to be effective in speech-related machine learning tasks, it is plausible that they would be equally applicable to the analysis of snoring. With this in mind, we follow the AST [19] to extract log Mel spectrograms, with a frequency dimension of 128, using a 25 ms Hamming window and a hop length of 10 ms. This gives us a resultant input size of $128 \times 100t$ for t seconds of audio.

Baseline experiments. The present study involved the implementation of the following baseline experiments using the proposed (SSBPR) dataset.

- 1) Since gender is a significant OSAHS risk factor and female patients in our dataset did not snore in a prone position, we conduct a 5-class classification experiment on the SSBPR dataset. We aim to prove the ability of the model to identify sleep positions based on snoring for different genders.
- 2) We evaluate the various sleep positions of the SSBPR by a six-classes classification experiment on the SSBPR dataset. This experiment aims to assess the ability of the dataset to accurately classify sleep body positions among patients with OSAHS and special patient populations (pregnant women and patients with laryngopharyngeal reflux disease).

Evaluation metric. The aim of the SSBPR dataset is to identify the sleeping posture of patients during snoring, so we employ *Accuracy* as the main metric to evaluate the performance of method.

Training details. The model has been trained on two NVIDIA GeForce RTX 3090 GPUs with batch size 12 for 30 epochs. The experimental setup mostly follows the AST. The model utilizes the initial learning rate of $1e-5$ with Adam optimization [28] and binary cross-entropy loss. The warm-up step is set to 5 epochs and the learning rate is scheduled with a factor of 0.85 for every epoch. We train both models by using frequency/time masking data augmentation [29] with max time mask length of 48 frames and max frequency mask length of 48 bins.

3.3. Results

As shown in Table 2, the model achieves an accuracy of 82.7%, 94.6% and 85.8% on male, female subjects and all subjects, respectively. The results show that SSBPR can be used to identify

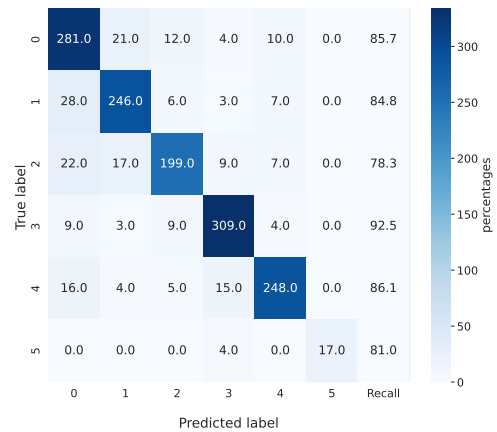


Figure 3: Confusion matrix of six-classes classification. The numbers 0, 1, 2, 3, 4 and 5 indicate different body positions: supine, supine but left lateral head, supine but right lateral head, left-side lying, right-side lying and prone, respectively.

the specific body position of patients and help to diagnose the obstruction site in the upper airway accurately. However, the accuracy for identifying sleeping positions is higher for female patients than for male patients in the five-classes classification experiment. Our inference is that one of the reasons is some pathophysiological differences between men and women [30]. Therefore, the accuracy of the methods can be improved by developing specific algorithms for male patients.

Furthermore, the model achieves an accuracy of 85.8% in six-class classification experiments. The corresponding confusion matrix in Figure 3 reveals significant differences between the classes of supine but right lateral head and left-side lying, with a range of 14.2% in the recall. This disparity may be attributed to data imbalance and limitations in the selected model architecture or model input. Future research should focus on addressing these challenges to enhance the accuracy of algorithms by accounting for the specific differences between the classes.

4. Conclusions

The phenomenon of snoring is influenced by the physical position of an individual's body during sleep. Despite this recognition, a lack of comprehensive datasets that consider the role of body position in snoring analysis has been observed. In order to address this shortfall, this paper endeavors to create a snoring dataset called SSBPR that incorporates the aspect of body position, which is expected to provide valuable insights into the field of snoring research. Furthermore, baseline experiments demonstrate that automatic classification models based on acoustic properties can differentiate between snoring at different sleep body positions. Adding more participants to the database, refining the snoring classes, and developing novel methods for snoring sound classification are areas for future work to further enhance the classification performance of different sleep body positions of snoring. The dataset will be released at <https://github.com/xiaoli1996/SSBPR>.

5. Acknowledge

This work was supported by the Hubei Province Technological Innovation Major Project (No. 2022BCA041).

6. References

- [1] P. J. Strollo Jr and R. M. Rogers, "Obstructive sleep apnea," *New England Journal of Medicine*, vol. 334, no. 2, pp. 99–104, 1996.
- [2] J. Fiz, J. Abad, R. Jane, M. Riera, M. Mananas, P. Caminal, D. Rodenstein, and J. Morera, "Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea," *European Respiratory Journal*, vol. 9, no. 11, pp. 2365–2370, 1996.
- [3] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 136–143, 2008.
- [4] J. A. Dempsey, S. C. Veasey, B. J. Morgan, and C. P. O'Donnell, "Pathophysiology of sleep apnea," *Physiological reviews*, vol. 90, no. 1, pp. 47–112, 2010.
- [5] C. Croft and M. Pringle, "Sleep nasendoscopy: a technique of assessment in snoring and obstructive sleep apnoea," *Clinical Otolaryngology & Allied Sciences*, vol. 16, no. 5, pp. 504–509, 1991.
- [6] C. Janott, M. Schmitt, Y. Zhang, K. Qian, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert *et al.*, "Snoring classified: the munich-passau snore sound corpus," *Computers in Biology and Medicine*, vol. 94, pp. 106–118, 2018.
- [7] M. Schmitt and B. Schuller, "End-to-end audio classification with small datasets—making it work," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.
- [8] Z. Zhang, J. Han, K. Qian, C. Janott, Y. Guo, and B. Schuller, "Snore-gans: Improving automatic snore sound classification with synthesized data," *IEEE journal of biomedical and health informatics*, vol. 24, no. 1, pp. 300–310, 2019.
- [9] K. Qian, M. Schmitt, C. Janott, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "A bag of wavelet features for snore sound classification," *Annals of Biomedical Engineering*, vol. 47, pp. 1000–1011, 2019.
- [10] L. Ding and J. Peng, "Automatic classification of snoring sounds from excitation locations based on prototypical network," *Applied Acoustics*, vol. 195, p. 108799, 2022.
- [11] M. Carrasco-Llatas, S. Matarredona-Quiles, A. De Vito, K. B. Chong, and C. Vicini, "Drug-induced sleep endoscopy: technique, indications, tips and pitfalls," in *Healthcare*, vol. 7, no. 3. MDPI, 2019, p. 93.
- [12] F. Safiruddin, I. Koutsourelakis, and N. de Vries, "Analysis of the influence of head rotation during drug-induced sleep endoscopy in obstructive sleep apnea," *The Laryngoscope*, vol. 124, no. 9, pp. 2195–2199, 2014.
- [13] P. E. Vonk, M. J. van de Beek, M. J. Ravesloot, and N. De Vries, "Drug-induced sleep endoscopy: new insights in lateral head rotation compared to lateral head and trunk rotation in (non) positional obstructive sleep apnea patients," *The Laryngoscope*, vol. 129, no. 10, pp. 2430–2435, 2019.
- [14] S. M. Mohammadi, M. Alnowami, S. Khan, D.-J. Dijk, A. Hilton, and K. Wells, "Sleep posture classification using a convolutional neural network," in *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 1–4.
- [15] M. Jesmeen, T. Bhuvanewari, A. H. Mazbah, Y. B. Chin, L. H. Siong, and N. H. A. Aziz, "Sleepcon: Sleeping posture recognition model using convolutional neural network," *Emerging Science Journal*, vol. 7, no. 1, pp. 50–59, 2022.
- [16] A. Breuss, N. Vonau, C. Ungricht, E. Schwarz, M. Irion, M. Bradich, F. A. Grewe, S. Liechti, S. Thiel, M. Kohler *et al.*, "Sleep position detection for closed-loop treatment of sleep-related breathing disorders," in *2022 International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2022, pp. 1–6.
- [17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.
- [18] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *arXiv preprint arXiv:2203.02155*, 2022.
- [19] Y. Gong, Y.-A. Chung, and J. Glass, "Ast: Audio spectrogram transformer," *arXiv preprint arXiv:2104.01778*, 2021.
- [20] R. B. Berry, R. Brooks, C. E. Gamaldo *et al.*, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6.0. <https://aasm.org/>," *American Academy of Sleep Medicine, Darien, Illinois (AASM)*, 2020.
- [21] M. Ravesloot, J. Van Maanen, L. Dun, and N. De Vries, "The undervalued potential of positional therapy in position-dependent snoring and obstructive sleep apnea—a review of the literature," *Sleep and Breathing*, vol. 17, no. 1, pp. 39–49, 2013.
- [22] E. R. Van Kesteren, J. P. van Maanen, A. A. Hilgevoord, D. M. Laman, and N. de Vries, "Quantitative effects of trunk and head position on the apnea hypopnea index in obstructive sleep apnea," *Sleep*, vol. 34, no. 8, pp. 1075–1081, 2011.
- [23] M. A. Carskadon, W. C. Dement *et al.*, "Normal human sleep: an overview," *Principles and practice of sleep medicine*, vol. 4, no. 1, pp. 13–23, 2005.
- [24] W. De Backer, "Obstructive sleep apnea/hypopnea syndrome," *Panminerva medica*, vol. 55, no. 2, pp. 191–195, 2013.
- [25] C. H. Daltro, F. H. d. O. Fontes, R. Santos-Jesus, P. B. Gregorio, and L. M. B. Araújo, "Obstructive sleep apnea and hypopnea syndrome (osahs): association with obesity, gender and age," *Arquivos Brasileiros de Endocrinologia & Metabologia*, vol. 50, pp. 74–81, 2006.
- [26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [27] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita, "Acoustic analysis of snoring sounds by a multidimensional voice program," *The Laryngoscope*, vol. 116, no. 3, pp. 379–381, 2006.
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [29] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," *arXiv preprint arXiv:1904.08779*, 2019.
- [30] C. M. Lin, T. M. Davidson, and S. Ancoli-Israel, "Gender differences in obstructive sleep apnea and treatment implications," *Sleep medicine reviews*, vol. 12, no. 6, pp. 481–496, 2008.