



Article

Auditory Property-Based Features and Artificial Neural Network Classifiers for the Automatic Detection of Low-Intensity Snoring/Breathing Episodes

Kenji Hamabe ¹, Takahiro Emoto ^{2,*}, Osamu Jinnouchi ³, Naoki Toda ⁴ and Ikuji Kawata ⁵

¹ Graduate School of Advanced Technology and Science, Tokushima University, Tokushima 770-8506, Japan; c612034034@tokushima-u.ac.jp

² Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima 770-8506, Japan

³ Research Laboratory, Imai Otorhinolaryngology Clinic, Tokushima 770-0026, Japan; jichy036@gmail.com

⁴ Department of Otolaryngology, Anan Medical Center, Tokushima 774-0045, Japan; naoki527@outlook.jp

⁵ Department of Otolaryngology, Yoshinogawa Medical Center, Tokushima 776-8511, Japan; ymc.jibika@ja-ymc.jp

* Correspondence: emoto@tokushima-u.ac.jp

Abstract: The definitive diagnosis of obstructive sleep apnea syndrome (OSAS) is made using an overnight polysomnography (PSG) test. This test requires that a patient wears multiple measurement sensors during an overnight hospitalization. However, this setup imposes physical constraints and a heavy burden on the patient. Recent studies have reported on another technique for conducting OSAS screening based on snoring/breathing episodes (SBEs) extracted from recorded data acquired by a noncontact microphone. However, SBEs have a high dynamic range and are barely audible at intensities >90 dB. A method is needed to detect SBEs even in low-signal-to-noise-ratio (SNR) environments. Therefore, we developed a method for the automatic detection of low-intensity SBEs using an artificial neural network (ANN). However, when considering its practical use, this method required further improvement in terms of detection accuracy and speed. To accomplish this, we propose in this study a new method to detect low SBEs based on neural activity pattern (NAP)-based cepstral coefficients (NAPCC) and ANN classifiers. Comparison results of the leave-one-out cross-validation demonstrated that our proposed method is superior to previous methods for the classification of SBEs and non-SBEs, even in low-SNR conditions (accuracy: $85.99 \pm 5.69\%$ vs. $75.64 \pm 18.8\%$).

Keywords: obstructive sleep apnea syndrome; auditory property; polysomnography; artificial neural network; snoring/breathing episode



Citation: Hamabe, K.; Emoto, T.; Jinnouchi, O.; Toda, N.; Kawata, I. Auditory Property-Based Features and Artificial Neural Network Classifiers for the Automatic Detection of Low-Intensity Snoring/Breathing Episodes. *Appl. Sci.* **2022**, *12*, 2242. <https://doi.org/10.3390/app12042242>

Academic Editors: Keun Ho Ryu and Nipon Theera-Umporn

Received: 13 January 2022

Accepted: 11 February 2022

Published: 21 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Obstructive sleep apnea syndrome (OSAS) is characterized by complete or incomplete obstruction of the upper airway during sleep. The main symptoms of OSAS are light sleep, excessive daytime sleepiness, and snoring; these are said to increase the risk of developing serious illnesses, such as ischemic heart disease, hypertension, stroke, and cognitive dysfunction [1]. Furthermore, it is said that 6–19% of females and 13–33% of males have OSAS, with the prevalence rate increasing with age [2,3]. A definitive diagnosis of OSAS is currently made using polysomnography (PSG) tests. However, this test requires multiple measurement sensors (e.g., oral thermistor, nasal pressure cannula, chest belt) to be worn directly on the body all night, which imposes a heavy burden on the patient. Previous studies suggested that the discomfort of wearing multiple sensors during PSG and restricted movements affect sleep efficiency, electrocardiographic (EEG) spectral power, and rapid-eye movements [4–7].

To resolve these problems, research is being conducted with the aim to establish an OSAS screening method that is based on noncontact microphones. These studies include (i) technological development to detect snoring/breathing episodes (SBEs) [8–13], (ii) studies of snoring that characterize OSAS [14–18], and (iii) snoring-based OSAS screening methods and evaluations [19–21].

In [8–13], snoring characteristics were extracted using ZCR, MFCC, and other statistical processing from the respiratory sounds during sleep which were obtained from patients; it was then shown that SBE sections could be classified using deep learning with accuracies in the range of 75.1–96.8% in various environments, including noise.

In [14–18], the effectiveness of various characteristics in OSAS and non-OSAS patients in terms of temporal, frequency, intensity, and clinical features was evaluated to characterize OSAS-related upper airway obstruction.

In [19–21], snoring sounds obtained from noncontact microphones were segmented, features were extracted by statistical processing and formant acoustic analysis, and machine learning tools, such as logistic regression and AdaBoost were used to classify OSAS/non-OSAS and sleep/waking states at sensitivities in the range of 80–90%, while doing so at a low cost.

As shown in [22–25], it has recently been suggested that sleep–awake activity and sleep quality could be estimated based on the analysis of respiratory sounds obtained during sleep. These results emphasized the importance of detecting SBEs during sleep. The automatic detection of SBEs from sleep sounds is the first step for automatic OSAS screening based on snoring. However, SBEs have a high dynamic range and are barely audible at intensities >90 dB. Specifically, there is a need for a method to automatically detect low-intensity SBEs without any contact, even in a low-SNR environment.

Therefore, our research group has been developing a system that automatically detects low-intensity SBEs from sleep sounds obtained by noncontact recording [10,12]. It has been suggested that the automatic detection of low-intensity SBEs has a high performance compared with other methods proposed in recent studies. However, the calculation speed and performance must be improved further for practical use.

The purpose of our study was to develop a more efficient method to detect low-intensity SBEs in sleep sound recordings.

Even if low-intensity SBEs are present in sleep sounds, human hearing can distinguish them from sleep sounds by careful listening. This is because the human auditory pathway has an innate function which is used to analyze the fine temporal characteristics of sound. The auditory image model (AIM) [26–28], which simulates a human auditory mechanism from an engineering perspective, was developed by Patterson in 1995 [26].

To generate a stabilized auditory image (SAI), this AIM describes a process of strobed temporal integration which transforms the signal flow from the cochlea up the auditory nerve to the brain. For sound event classification [1,13,29], front-end, ear-like audio analysis has been conducted by generating features extracted from an SAI. However, the calculation of SAI requires large computational and memory costs. Conversely, sound event detection performed based on the peaks corresponding to glottal pulses was apparent in the neural activity pattern (NAP) which was converted into an SAI [30–32]. Furthermore, the NAP which produces spectral profiles from AIM were used for the communication of sound recognition and the analysis of cochlear implant representations [33,34]. From these reports, we hypothesize that NAP carries information on the presence or absence of sound events even before SAI modeling.

A novel aspect of this study is that we propose the new feature, NAP-based cepstral coefficients (NAPCC), for the automatic, accurate, and faster detection of low-intensity SBEs in sleep sound recordings.

Based on leave-one-out cross validation of sleep sound data stored in a database, the performance of the proposed method was investigated and compared with that of the low-intensity SBEs detection method developed in our previous study in 2018 [10,12].

To date, sleep–awake evaluation methods and OSAS screening methods have been developed using SBEs obtained based on the noncontact approach [19–21]. High-intensity SBEs can be detected by the energy-based approach; however, if low-intensity SBEs can be detected efficiently and automatically by this study, then the presence or absence of patient’s breathing can be estimated from the recorded data, regardless of SBE intensity.

A noncontact approach based on sleep sound analysis was developed with the objective of a cost-effective alternative approach to OSAS diagnosis. Incorporating the proposed method in these approaches may enable more accurate OSAS screening and sleep stage evaluations.

2. Materials and Methods

2.1. Snoring/Breathing Episodes

This study was conducted after obtaining approval from the ethics review boards of the Division of Science and Technology, Graduate School of Technology, Industrial and Social Sciences, Tokushima University, and Anan Kyoei Hospital. Sleep sounds were recorded during a PSG test conducted at the Anan Kyoei Hospital. A microphone (Model NT3, RODE, Sydney, Australia) was placed approximately 50 cm away from the patient, and its distance could vary from 40 to 70 cm depending on the patient’s movements. Sleep sounds were recorded using a preamplifier (Mobile pre USB, M-Audio, CA, USA), with a sampling rate of 44.1 kHz and digital resolution of 16 bits/sample. However, in this study, the recorded data were downsampled to 11.025 kHz at the time of analysis in consideration of the main SBE components [35,36].

The SBEs and non-SBEs used in this study were identified by three annotators who carefully listened to the recorded data. The SBE/non-SBE sections were finally determined from the average values of the start and end points of the SBEs/non-SBE sections identified by the three annotators after they carefully listened to the recordings. The degree of matching of the annotations of the three annotators was calculated using Cohen’s kappa [37,38] to guarantee the reliability of annotations. The SNRs of the SBEs included in the recorded data were calculated from the annotation results using the following equation:

$$\text{SNR} = 10 \log_{10} \frac{P_S - P_N}{P_N}, \quad (1)$$

Herein, P_S and P_N denote the SBE and noise power, respectively. In this study, the recorded data were selected to satisfy the following conditions to evaluate the performance of the proposed method in low-SNR conditions. Furthermore, the recorded data used in this study included a 120 s section extracted from the 6 h sleep sound data, wherein multiple low-intensity SBEs existed which were composed of SBE and non-SBE sections.

1. The amplitude of SBE within the 120 s interval did not change considerably across all the recorded data
2. SBEs with low SNR were repeated in the 120 s interval of recorded data

(1) Exp-1: SBE detection of the recorded data from 25 individuals, wherein as SBEs and non-SBEs only included silence periods and (2) Exp-2: SBE detection of recorded data from 15 individuals, wherein SBEs and non-SBEs included talking, alarm sounds, footsteps, and fan noise which may have occurred during actual sleep. Table 1 shows the subject record databases used in Exp-1 and Exp-2. It can be observed from Table 1 that in Exp-1, the range of average SNR of the SBEs recorded from the OSAS (AHI > 10) and non-OSAS (AHI < 10) subjects ranged from -8.34 ± 1.40 to 0.88 ± 3.24 . In Exp-2, the range of average SNRs of the SBEs recorded from the OSAS and non-OSAS subjects ranged from -13.84 ± 4.02 to 0.05 ± 3.22 .

Table 1. Subject record database used in Exp-1.

	Exp-1	Exp-2
No. of patients	25	15
AHI	26.8 ± 22.9	24.1 ± 22.8
BMI	26.2 ± 6.2	25.5 ± 2.2
Age	56.6 ± 22.0	60.3 ± 12.3
Gender	19 males/6 females	9 males/6 females
SNR [dB]	$-8.34 \pm 1.40 \sim 0.88 \pm 3.24$	$-13.84 \pm 4.02 \sim -0.05 \pm 3.22$
Class	18 OSAS/7 non-OSAS	8 OSAS/7 non-OSAS
No. of SBEs	52.7 ± 13.2	54.1 ± 12.4

The average number of segments with SBEs/silence used in Exp-1 was 54.7 ± 13.2 and 54.1 ± 12.4 , respectively. The number of segments with SBEs/non-SBEs used in Exp-2 is described in detail in Table 2. It can be observed from Table 2 that non-SBEs that are expected to occur during actual sleep were used in the experiment.

Table 2. Number of segments with snoring/breathing episodes (SBEs)/non-SBEs used in Exp-2.

	No. of Segments
SBEs	963.4 ± 290.2
Silence	502.7 ± 651.2
Music	64.3 ± 193.2
Alarm	93.3 ± 257.9
Speech	52.9 ± 172.5
Non-SBEs	15.7 ± 30.0
Footsteps	2.5 ± 4.4
Mouth movement	22.0 ± 59.9
Duvet noise	718.9 ± 785.9
Fan	46.0 ± 71.6
Tapping	

2.2. Auditory Property-Based Features and Artificial Neural Network Classifiers

We describe herein a new method based on the use of auditory model-based features wherein artificial neural network (ANN) classifiers were used to detect quickly low-intensity SBEs in the sleep sound records. Humans can distinguish small sound events and silence from sleep sounds. Therefore, in this study, we used the AIM of Patterson et al. [26], which simulated the processing mechanism of the auditory system. AIM consists of precochlear processing (PCP), basilar membrane motion (BMM), and the NAP and stabilized auditory image (SAI), which is converted using strobe temporal integration modules. However, given that (i) the calculation of SAI requires large computational and memory costs and (ii) the information on the detection of sound event is included prior to the strobe temporal integration processing [39], we used the PCP, BMM, and NAP modules of the AIM. Figures 1 and 2 show the flow charts of the automatic SBE detection system proposed in this study.

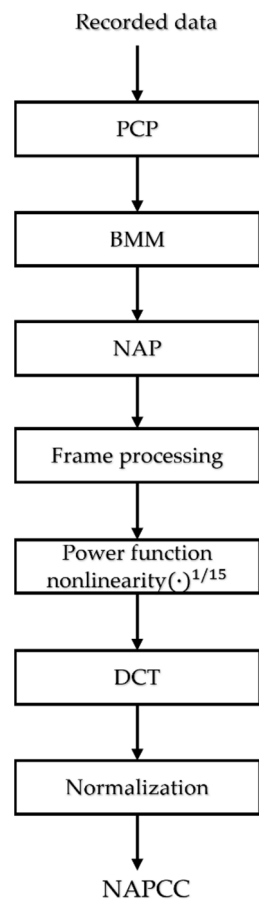
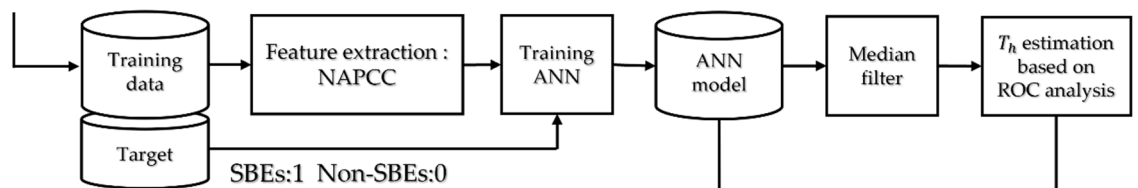


Figure 1. Flow chart of feature extraction based on the adopted auditory model.

Training phase

Recorded data of N-1 subjects



Testing phase

Recorded data of 1 subject

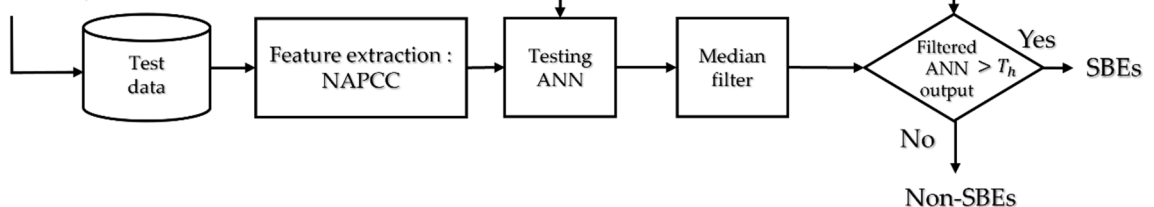


Figure 2. Flow chart of the automatic snoring/breathing event (SBE) detection system proposed in this study.

As shown in Figure 1, the 120 s recorded data segment first underwent preprocessing with a bandpass filter that simulated the characteristics of the outer and middle ears at the PCP stage of AIM (lower cutoff frequency: 1000 Hz, and upper cutoff frequency: 6000 Hz) [40]. After the recorded data passed through the PCP stage, the recorded data

were divided in windows with widths equal to 1024 samples and a shift width equal to 1 sample. Subsequently, at the BMM stage, filtering was conducted with an auditory filter bank that simulated the cochlear frequency analysis mechanism. In this study, we used a gamma chirp filter bank composed of 50 channel filters between the asymptotic frequencies of 100 Hz and 5000 Hz as an auditory filter bank. Given that the NAP stage simulates the phase fixing characteristics when converting BMM physical information into acoustic nerve information, the filter output obtained from the BMM stage was filtered by half-wave rectification and a low-pass filter. The output obtained from the NAP stage was framed with a frame size of 1024 samples and a shift size of 512 samples. Through this frame processing, a total number of 2582 frames for the 120 s recorded data were obtained.

We applied power-law nonlinearity with an exponent of $1/15$ on each NAP frame to derive cepstral features. Furthermore, DCT which has a property of energy compaction was also applied.

The output extracted from the DCT generally needed to be normalized before the classification process. However, there is no optimal way of normalization or formal correction, as described in [41]. Thus, the output extracted from the DCT was normalized using mean normalization [42] and sigmoid normalization [43], which are typically performed on the analysis data. Herein, we compared the effect of mean normalization with that of sigmoid normalization on the output from the DCT.

In this study, we used a new feature extraction algorithm called NAPCC that was based on auditory processing which corresponded to the above procedure. Figure 2 shows the structure of the new NAPCC approach that we introduce in this study.

We used an ANN based on NAPCC as a discriminator for the classification of SBE and non-SBE sections from the recorded data.

We used multilayer perceptron (MLP)-ANN and radial basis function (RBF)-ANN as a classifier for the classification of SBE and non-SBE sections from the recorded data. The MLP-ANN consists of three layers: input layer, hidden layer, and output layer [44]. Herein, the output function of the hidden layer unit was a hyperbolic tangent function, and the transfer function of the output layer unit was a linear function. The number of units in the hidden and output layers of the MLP-ANN were 10 and 1, respectively.

RBF-ANN is also composed of three layers: an input layer, a hidden radial basis function layer, and an output layer. The number of units in the hidden and output layers of the RBF-ANN were also 10 and 1, respectively. The weighted input of the RBF hidden layer is computed by the ratio of the Euclidean distance between the weight vector and the input vector to the spread parameter (σ) which allows the sensitivity of the radial basis neuron. In this work, the spread parameter (σ) was set to 1. This network is known to have strong tolerance to input noise and fast and comprehensive training and responds well to test patterns [45,46].

The NAPCC extracted from each of the frames of the recorded data were given as input to the input layer of both ANN. The target signals of 1 and 0 were provided to the SBE and non-SBE sections, respectively, and to prevent overfitting, both ANN were trained by the error back propagation method based on the Levenberg–Marquardt method with early stopping [43,44]. The output result after learning was subjected to 4th-order median filtering to eliminate the influences of sudden changes.

2.3. Evaluation of the Performance of the Proposed Detection Method

The effectiveness of the SBE/non-SBE classification of the proposed method was validated by dividing the recording data of N individuals into the recording data of one individual for test purposes and the recorded data of $N-1$ individuals for training purposes; the leave-one-out cross-validation (which repeats the validation process N times so that the recording data of each individual were selected once as test data) was used. The training data were composed of NAPCC patterns extracted frame by frame from the recorded data of $N-1$ individuals. The testing data were composed of NAPCC patterns extracted frame by frame from the recorded data of one individual.

Receiver operating characteristic (ROC) analysis was conducted from the output of the ANN after learning was achieved from the training data, and the test data were used to estimate the optimal threshold T_h for use when the SBEs and non-SBEs were classified. The optimal threshold value at this time is the threshold T_h that minimizes the Euclidean distance from the position of sensitivity of 1 and specificity of 1 on the ROC curve [47]. Based on the threshold value that was estimated in this way, the SBE/non-SBE classification accuracies that used the test data were calculated. Specifically, the sensitivity, specificity, PPV, NPV, accuracy, and F1 score were estimated. As this study used the leave-one-out cross-validation method, validation was conducted N times in total, and the means and standard deviations of the classification accuracies were calculated for a total number of N.

3. Results

Fifty dimensional features extracted from the DCT were normalized using mean or sigmoid normalizations. In consideration of the initial value dependence of MLP-ANN, leave-one-out cross-validation was used, and a mean F1 score was obtained. Subsequently, the initial value of the MLP-ANN was changed, and validation was repeated 10 times; a performance evaluation of the proposed method was then conducted based on the trial results that maximized the F1 score.

Table 3 shows the results of the inter-annotator agreement of the three human judges using Cohen's kappa for our careful listening process of the sleep sound recordings. From this table, we can confirm that our annotators achieved a kappa coefficient > 0.9 , which indicates an almost perfect agreement.

Table 3. Inter-annotator agreement computed based on Cohen's kappa.

	Annotator 2	Annotator 3
Annotator 1	0.97	0.98
Annotator 2		0.98

3.1. Normalization Used for NAPCC and Optimum Number of NAPCC

In this section, the performance of the proposed method was evaluated by changing the number of dimensions of NAPCC presented to MLP-ANN to investigate how the number of dimensions of NAPCC obtained based on the use of these normalization methods influenced the automatic extraction performance of SBEs. Figure 3 shows the relationship between the mean F1 score and the number of NAPCC dimensions when sigmoid or mean normalizations were used in Exp-1. Figure 4 shows the respective results for Exp-2.

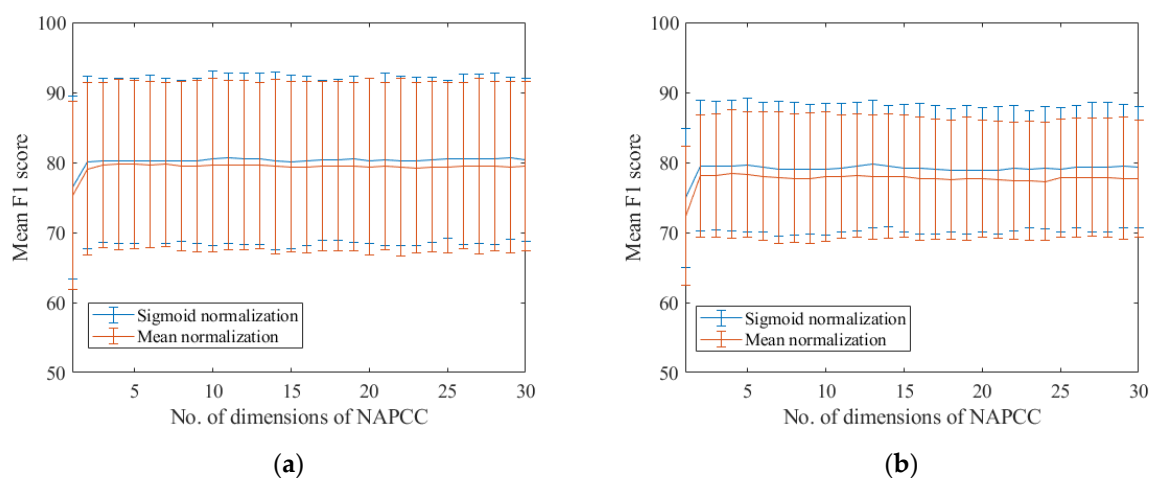


Figure 3. Relationship between mean F1 score and number of dimensions of neural activity pattern-based cepstral coefficients (NAPCC) when sigmoid or mean normalizations were used: (a) Exp-1 and (b) Exp-2 results.

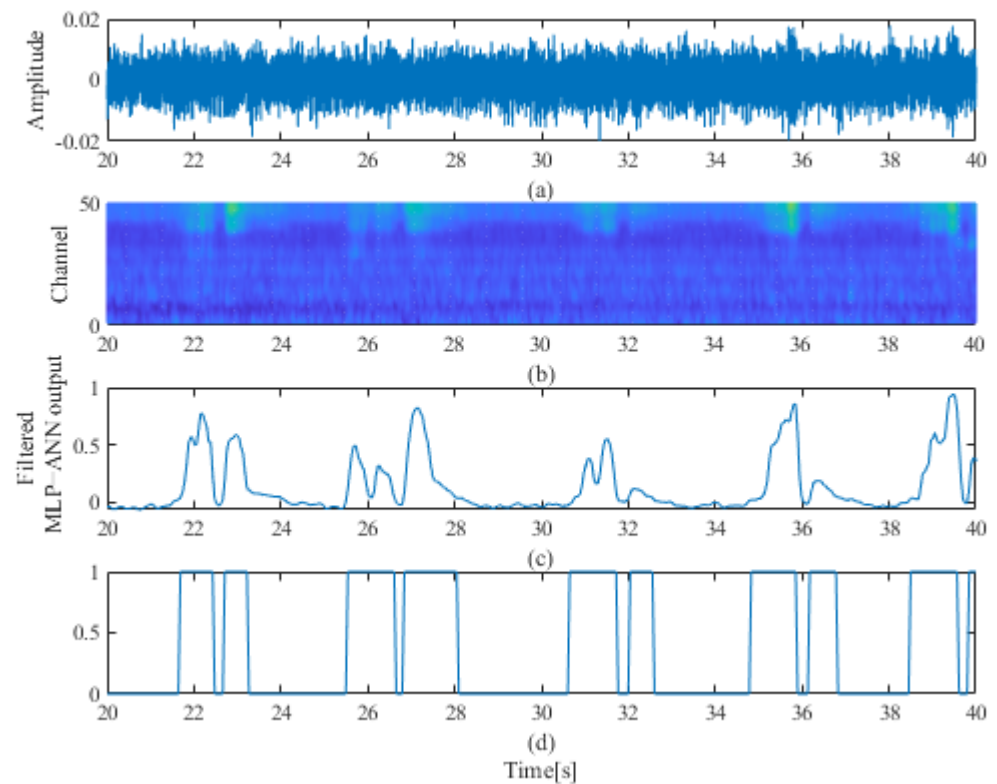


Figure 4. SBE detection results based on the proposed method: (a) 20 s of recorded data, (b) NAP output obtained by analyzing the recorded data, (c) trained MLP-ANN output results, (d) results of labeling by three annotators (1 for SBE sections, 0 for non-SBE sections).

It can be observed from Figure 3 that the F1 score of sigmoid normalization was approximately 0.5% and 1% higher than that of mean normalization in Exp-1 and Exp-2, respectively. There were no large differences in the standard deviation between the sigmoid and the mean normalizations. Given that sigmoid normalization seems to perform better based on the results obtained in these experiments, we employed the sigmoid normalization in the NAPCC extraction process in this study.

According to the results associated with the use of sigmoid normalization in Figure 3, it was confirmed that the maximum score in 13 dimensions was obtained when the mean F1 scores in each dimension obtained based on Exp-1 and Exp-2 were multiplied. The 13-dimensional NAPCC was determined as the optimal characteristic vector used in the proposed method in this study. In the following sections, this characteristic vector was used for the evaluation of the performance of the proposed method and for the comparison of the proposed method with the previous method.

Figure 4 shows (as examples) (a) 20 s of recorded data, (b) NAP output obtained by analyzing the recorded data, (c) trained MLP-ANN output results, and (d) labeling results by three annotators (1 for SBE sections, 0 for non-SBE sections). It can be confirmed from these figures that the NAP output and the trained MLP-ANN output reflected even the information of SBEs that appeared to be buried in background noise.

3.2. Evaluation of the Performance of the Proposed Method and Comparison of the Proposed Method with Our Previous Method

We developed an SBE/non-SBE classification technique based on MLP-ANN, which was used as a time-series classifier [12]. When this MLP-ANN was used as a subject-independent classifier, we showed that our previous method can classify low-intensity SBEs and low-intensity non-SBEs from sleep sounds recorded in noisy environments with an average accuracy of 75.10%. The performance of our previous technique was compared with that of other recent techniques. Even though we focused on the detection of low-

intensity SBEs in sleep sounds, the classification accuracy was as good as that attained with recent techniques. In this study, we compared the performance of the proposed method with that of this method. For this purpose, Exp-1 and Exp-2 were performed using the same parameters adopted in our previous work.

Tables 3 and 4 show the performance evaluation results of the proposed and the conventional methods for Exp-1 and Exp-2, respectively. As mentioned in Section 3.1, in consideration of the initial value dependence of MLP-ANN, leave-one-out cross-validation was conducted 10 times, and performance evaluations of the proposed method and previous method were conducted based on trial results that maximized the F1 score.

Table 4. Performance evaluation results of the proposed and conventional methods in Exp-1.

	Accuracy [%]	Sensitivity [%]	Specificity [%]	PPV [%]	NPV [%]	F1 Score [%]
Proposed method	85.83 ± 7.90	81.75 ± 11.98	91.95 ± 7.72	83.81 ± 18.23	85.45 ± 14.99	80.48 ± 12.25
Previous method	82.59 ± 10.14	82.84 ± 12.91	87.86 ± 13.07	79.56 ± 22.29	85.45 ± 15.70	77.57 ± 14.48

According to Table 4, in the case of Exp-1, we found that the mean value was larger than that of our previous method, and the standard deviation of the proposed method was smaller than that of our previous method. As shown in Table 5, the results of Exp-2 yield the same trend as that of Exp-1. These results clearly indicate that the performance of the proposed method was superior to that of our previous method even in noisy environments.

Table 5. Performance evaluation results of the proposed and conventional methods in Exp-2.

	Accuracy [%]	Sensitivity [%]	Specificity [%]	PPV [%]	NPV [%]	F1 Score [%]
Proposed method	85.99 ± 5.69	79.64 ± 9.50	91.34 ± 7.18	82.87 ± 15.80	87.31 ± 8.63	79.81 ± 9.14
Previous method	75.64 ± 18.80	73.64 ± 17.34	81.97 ± 26.06	75.65 ± 15.80	83.81 ± 11.42	69.40 ± 15.85

Figure 5a,b shows the F1 scores of the subjects obtained via the proposed method or our previous method in Exp-1 and Exp-2, respectively. Please note that herein, the mean F1 score value was used after leave-one-out cross-validation was conducted 10 times. It can be observed from these figures that the F1 score of the proposed method in each subject exceeded the F1 score of the conventional method in most cases. In particular, the results of Exp-2 suggest that the use of the proposed method improved the detection performance of subjects for whom detection was difficult with the conventional method. Exp-1 demonstrates that in most cases, the proposed method worked better than the previous method. In particular, for the recorded data (No. 9 and 10) of subjects who were not detected, the F1 score of the proposed method was improved by about 20–30% compared with that of the previous method.

These findings suggest that the proposed method was more effective than the conventional method even when SBEs were detected from the recorded data which contained low-SNR non-SBEs and SBEs.

Regarding subject No. 10 in Exp-1 and subject No. 5 in Exp-2, a degradation in performance was found compared to the results obtained from the other subjects. This could be due to the shorter duration of low-intensity SBEs.

It was considered that in cases of practical use of this proposed method, the recorded data used in Exp-1 and in Exp-2 would be mixed together; therefore, in subsequent experiments, using two classifiers, namely, MLP-ANN and RBF-ANN, the effectiveness of the proposed method was validated for recording datasets constructed by mixing together the recorded data used in this study. Table 6 lists the performance outcomes of the proposed method for the mixture of the data used in Exp-1 and Exp-2 using two classifiers, i.e., MLP-ANN and RBF-ANN. These results suggest that, irrespective of the type of ANN, the proposed method can detect SBEs with higher accuracy than the conventional method, even for mixed recorded data.

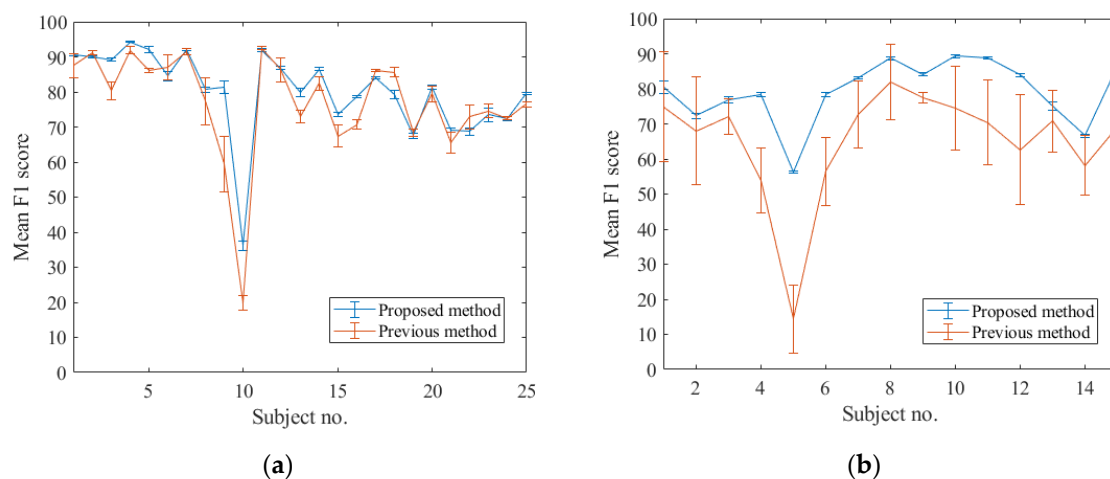


Figure 5. F1 score of the subjects obtained via the proposed or our previous method. (a) Exp-1 and (b) Exp-2 cases.

Table 6. Performance results of the proposed method based on the different classifiers applied on a dataset constructed by mixing data from Exp-1 and Exp-2.

Types of Classifiers Used in the Proposed Method	Accuracy [%]	Sensitivity [%]	Specificity [%]	PPV [%]	NPV [%]	F1 Score [%]
MLP-ANN	86.14 ± 6.96	80.97 ± 10.58	91.96 ± 7.29	83.84 ± 16.83	86.18 ± 12.74	80.58 ± 10.72
RBF-ANN	86.08 ± 7.07	81.28 ± 10.66	91.64 ± 7.24	83.34 ± 16.81	86.29 ± 12.80	80.53 ± 10.72

4. Discussion and Conclusions

In this study, we proposed a new method that used the auditory property-based features of NAPCC and ANN discriminators to automatically detect low-intensity SBEs from recorded data that were acquired using a noncontact microphone. The effectiveness of the proposed method was investigated by detecting SBEs from recorded data of 25 individuals which comprised silence and SBEs (Exp-1) and from recorded data of 15 individuals which included non-SBEs thought to be noise, generated in an actual environment (Exp-2). Leave-one-out cross-validation was used to evaluate the performance in each experiment. The results suggested that SBEs could be detected with an average accuracy of 85.83% in Exp-1 and of 85.99% in Exp-2. A comparison of performance with the MLP-ANN-based SBE detection method proposed in our previous work [12] showed that the proposed method was approximately 3% better in the case of Exp-1 and approximately 10% better in the case of Exp-2. In particular, the standard deviation in both Exp-1 and Exp-2 became smaller with the proposed method when compared with the conventional method. Hence, it is thought that the influence of each individual subject could be reduced. Furthermore, the large improvement in specificity and PPV suggests that the new method is useful for the effective and automatic detection of silent or apneic sections contained in sleep sounds.

To date, gammatone frequency cepstral coefficients (GFCC) [48] and BMM-based cepstrum coefficients [49] have been developed, but both require gammatone filter bank outputs or mean normalization. To achieve a more accurate human auditory perception, a DCGA filter was developed to extend the domain of the gammatone auditory filter. This filter bank accommodates the nonlinear behavior observed in human psychophysics and can be useful for perceptual signal processing [50]. The DCGA filter bank, which is the front-end of NAPCC, may be more noise-robust than the gammatone filter bank. Furthermore, the results obtained in this work showed that the noise robustness of NAPCC was improved by using the sigmoid normalization instead of the mean normalization used in GFCC and BMM-based cepstrum coefficients. Therefore, NAPCC which uses DCGA

and sigmoid normalization should improve the performance of GFCC and BMM-based cepstrum coefficients.

In particular, to detect low-intensity SBEs in the sleep sound recordings, the method developed in our previous study in 2018 outperformed the recent techniques published up to 2018 [12]. Since 2018, more recent technologies have been developed to classify snoring episodes and non-snoring episodes from sleep sounds.

Lim et al. proposed a recurrent neural network (RNN)-based classification method that was capable of classifying snoring episodes and non-snoring episodes with the use of features obtained from the sleep data recorded with a smartphone. The RNN-based classifiers achieved an accuracy of 98.9% using relatively small datasets [9]. However, snoring segments used in this work were created based on the peak point of snoring signals obtained via the peak-detection algorithm. This means that relatively high-intensity snoring was selected for the analysis.

Jiang et al. [51] proposed an automatic snore detection method using sound maps and a series of neural networks. The results demonstrated that the method is appropriate for identifying snores with an accuracy in the range of 91.8–95.1%. However, in this study, potential snoring episodes were segmented using the improved sub-band spectral entropy method which is based on sub-band energy calculation.

Shen et al. [8] proposed the use of the MFCC feature extraction method and the LSTM model for the binary classification of snoring data. The experimental results showed that the developed method yielded the highest accuracy rate of 87%. However, for the analysis, very weak snoring sounds were not labeled in the data presented on PSG used in this study.

Furthermore, a sleep sound classification method based on AIM has been proposed for sleep sounds extracted by an energy-based approach, and it has been confirmed that sleep sounds can be classified with high accuracy [13]. However, there is a need to use multiple acoustic features obtained from SAI which is converted from NAP using strobed temporal integration (STI).

This study has the following advantages. The proposed method can be conducted with low-computational costs because it eliminates the computationally expensive STI processing used in AIM and can be built using stages up to NAP. Additionally, it has been confirmed that the use of the proposed method allows for the detection of low-intensity SBEs with higher performance compared with our previous method [12], and the computational speed was also significantly improved. Given that the performance of the proposed method was superior to that of our previous method even in the case of Exp-2, it is suggested that the new feature (NAPCC) proposed in this study is an acoustic feature that is robust against noise.

However, our study has some limitations: (i) a relatively small size of the dataset, which cannot satisfy sound variety, was used; (ii) for SBEs of short duration, the performance of the proposed method was degraded because the output of the NAPCC corresponding to the SBEs became small in the NAPCC spectrogram, which is the ordered series of the NAPCC of each frame for the recorded data; (iii) the proposed method uses the DCGA filter bank approach which has the highest calculation cost in the NAPCC calculation procedure.

The proposed method is expected to contribute as a pretreatment step to OSAS screening based on snoring and respiratory sounds. It is thought to be useful for the effective and automatic identification of respiratory sound information, particularly apneic sections and silence, from sleep sounds acquired without contact.

Author Contributions: Conceptualization, T.E., O.J., N.T., and I.K.; methodology, T.E. and K.H.; software, K.H.; validation, T.E. and K.H.; formal analysis, K.H.; investigation, T.E., K.H., and N.T.; data curation, K.H., I.K., N.T., and O.J.; writing—original draft preparation, K.H.; writing—review and editing, T.E.; project administration, T.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted after obtaining approval from the ethics review boards of the Division of Science and Technology, Graduate School of Technology, Industrial and Social Sciences (No. 20012), Tokushima University, and Anan Kyoei Hospital (No. 39).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Mikiya Kanagawa for his help with the data analysis of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McLoughlin, I.; Zhang, H.; Xie, Z.; Song, Y.; Xiao, W. Robust sound event classification using deep neural networks. *IEEE ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 540–552. [[CrossRef](#)]
2. Parish, J.M.; Somers, V.K. Obstructive sleep apnea and cardiovascular disease. *Mayo Clin. Proc.* **2004**, *79*, 1036–1046. [[CrossRef](#)] [[PubMed](#)]
3. Senaratna, C.V.; Perret, J.L.; Lodge, C.J.; Lowe, A.J.; Campbell, B.E.; Matheson, M.C.; Hamilton, G.S.; Dharmage, S.C. Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Med. Rev.* **2017**, *34*, 70–81. [[CrossRef](#)] [[PubMed](#)]
4. Agnew, H.W., Jr.; Webb, W.B.; Williams, R.L. The first night effect: An EEG study of sleep. *Psychophysiology* **1966**, *2*, 263–266. [[CrossRef](#)]
5. Curcio, G.; Ferrara, M.; Piergianni, A.; Fratello, F.; De Gennaro, L. Paradoxes of the first-night effect: A quantitative analysis of antero-posterior EEG topography. *Clin. Neurophysiol.* **2004**, *115*, 1178–1188. [[CrossRef](#)]
6. Le Bon, O.; Staner, L.; Hoffmann, G.; Dramaix, M.; San Sebastian, I.; Murphy, J.R.; Kentos, M.; Pelc, I.; Linkowski, P. The first-night effect may last more than one night. *J. Psychiatr. Res.* **2001**, *35*, 165–172. [[CrossRef](#)]
7. Beattie, Z.T.; Hayes, T.L.; Guillemainault, C.; Hagen, C.C. Accurate scoring of the apnea-hypopnea index using a simple non-contact breathing sensor. *J. Sleep Res.* **2013**, *22*, 356–362. [[CrossRef](#)]
8. Shen, F.; Cheng, S.; Li, Z.; Yue, K.; Li, W.; Dai, L. Detection of snore from OSAHS patients based on deep learning. *J. Healthc. Eng.* **2020**, *2020*, 1–10. [[CrossRef](#)]
9. Lim, S.J.; Jang, S.J.; Lim, J.Y.; Ko, J.H. Classification of snoring sound based on a recurrent neural network. *Expert Syst. Appl.* **2019**, *123*, 237–245. [[CrossRef](#)]
10. Emoto, T.; Abeyratne, U.; Chen, Y.; Kawata, I.; Akutagawa, M.; Kinouchi, Y. Artificial neural networks for breathing and snoring episode detection in sleep sounds. *Physiol. Meas.* **2012**, *33*, 1675. [[CrossRef](#)]
11. Karunajeewa, A.S.; Abeyratne, U.R.; Hukins, C. Silence-breathing-snore classification from snore-related sounds. *Physiol. Meas.* **2008**, *29*, 227. [[CrossRef](#)] [[PubMed](#)]
12. Emoto, T.; Abeyratne, U.R.; Kawano, K.; Okada, T.; Jinnouchi, O.; Kawata, I. Detection of sleep breathing sound based on artificial neural network analysis. *Biomed. Signal Process. Control* **2018**, *41*, 81–89. [[CrossRef](#)]
13. Nonaka, R.; Emoto, T.; Abeyratne, U.R.; Jinnouchi, O.; Kawata, I.; Ohnishi, H.; Akutagawa, M.; Konaka, S.; Kinouchi, Y. Automatic snore sound extraction from sleep sound recordings via auditory image modeling. *Biomed. Signal Process. Control* **2016**, *27*, 7–14. [[CrossRef](#)]
14. Emoto, T.; Abeyratne, U.R.; Akutagawa, M.; Konaka, S.; Kinouchi, Y. High frequency region of the snore spectra carry important information on the disease of sleep apnoea. *J. Med. Eng. Technol.* **2011**, *35*, 425–431. [[CrossRef](#)] [[PubMed](#)]
15. Markandeya, M.N.; Abeyratne, U.R.; Hukins, C. Characterisation of upper airway obstructions using wide-band snoring sounds. *Biomed. Signal Process. Control* **2018**, *46*, 201–211. [[CrossRef](#)]
16. Ng, A.K.; San Koh, T.; Baey, E.; Lee, T.H.; Abeyratne, U.R.; Puvanendran, K. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Med.* **2008**, *9*, 894–898. [[CrossRef](#)] [[PubMed](#)]
17. Benavides, A.M.; Murillo, J.L.B.; Pozo, R.F.; Cuadros, F.E.; Toledano, D.T.; Alcázar-Ramírez, J.D.; Gómez, L.A.H. Formant frequencies and bandwidths in relation to clinical variables in an obstructive sleep apnea population. *J. Voice* **2016**, *30*, 21–29. [[CrossRef](#)] [[PubMed](#)]
18. Huang, Z.; Aarab, G.; Ravesloot, M.J.; Zhou, N.; Bosschieter, P.F.; van Selms, M.K.; Hilgevoord, A.A. Prediction of the obstruction sites in the upper airway in sleep-disordered breathing based on snoring sound parameters: A systematic review. *Sleep Med.* **2021**, *88*, 116–133. [[CrossRef](#)]
19. De Silva, S.; Abeyratne, U.R.; Hukins, C. A method to screen obstructive sleep apnea using multi-variable non-intrusive measurements. *Physiol. Meas.* **2011**, *32*, 445. [[CrossRef](#)]
20. Abeyratne, U.R.; De Silva, S.; Hukins, C.; Duce, B. Obstructive sleep apnea screening by integrating snore feature classes. *Physiol. Meas.* **2013**, *34*, 99. [[CrossRef](#)]
21. Marcal, T.A.; dos Santos, J.M.; Rosa, A.; Cardoso, J.M. OSAS assessment with entropy analysis of high resolution snoring audio signals. *Biomed. Signal Process. Control* **2020**, *61*, 101965. [[CrossRef](#)]
22. Dafna, E.; Tarasiuk, A.; Zigel, Y. Sleep-wake evaluation from whole-night non-contact audio recordings of breathing sounds. *PLoS ONE* **2015**, *10*, e0117382. [[CrossRef](#)] [[PubMed](#)]

23. Rathnayake, S.I.; Wood, I.A.; Abeyratne, U.; Hukins, C. Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1973–1981. [[CrossRef](#)] [[PubMed](#)]
24. Akhter, S.; Abeyratne, U.R.; Swarnkar, V. Characterization of REM/NREM sleep using breath sounds in OSA. *Biomed. Signal Process. Control* **2016**, *25*, 130–142. [[CrossRef](#)]
25. Xue, B.; Deng, B.; Hong, H.; Wang, Z.; Zhu, X.; Feng, D.D. Non-contact sleep stage detection using canonical correlation analysis of respiratory sound. *IEEE J. Biomed. Health. Inf.* **2019**, *24*, 614–625. [[CrossRef](#)]
26. Patterson, R.D.; Allerhand, M.H.; Giguère, C. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J. Acoust. Soc. Am.* **1995**, *98*, 1890–1894. [[CrossRef](#)]
27. Iitomi, Y.; Matsuoka, T. Experiments on perceiving the missing fundamental by using two harmonic components tone. *J. Acoust. Soc. Am.* **2007**, *121*, 3092. [[CrossRef](#)]
28. Patterson, R.; Robinson, K.; Holdsworth, J.; McKeown, D.; Zhang, C.; Allerhand, M. Complex Sounds and Auditory Images. In *Auditory Physiology and Perception*; Pergamon Press: Oxford, UK, 1992; pp. 429–446.
29. Emoto, T.; Abeyratne, U.R.; Shono, T.; Nonaka, R.; Jinnouchi, O.; Kawata, I.; Kinouchi, Y. Auditory image model for the characterisation of obstructive sleep apnoea. *Screening* **2016**, *6*, 8.
30. Irino, T.; Patterson, R.D.; Kawahara, H. Speech segregation using an event-synchronous auditory image and STRAIGHT. In *Speech Separation by Humans and Machines*; Springer: Boston, MA, USA, 2005; pp. 155–165.
31. Irino, T.; Patterson, R.D.; Kawahara, H. Speech segregation using an auditory vocoder with event-synchronous enhancements. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 2212–2221. [[CrossRef](#)]
32. Irino, T.; Patterson, R.D.; Kawahara, H. Speech Segregation Using Event Synchronous Auditory Vocoder. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (Cat. No.03CH37404), Hong Kong, China, 6–10 April 2003; Volume 5, p. V-525.
33. Walters, T.C. Auditory-Based Processing of Communication Sounds. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2011.
34. Lafren, J.B.; Talavaoe, T.M.; Thirukkonda, P.M.; Svirsky, M.A. Physiologically Based Analysis of Cochlear Implant Representations. In Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society, Houston, TX, USA, 23–26 October 2002; Volume 3, pp. 2078–2079.
35. Sola-Soler, J.; Jane, R.; Fiz, J.A.; Morera, J. Formant Frequencies of Normal Breath Sounds of Snorers may Indicate the Risk of Obstructive Sleep Apnea Syndrome. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 3500–3503.
36. Yadollahi, A.; Moussavi, Z. Formant Analysis of Breath and Snore Sounds. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 2563–2566.
37. Fleiss, J.L.; Levin, B.; Paik, M.C. The measurement of interrater agreement. *Stat. Meth Rates Proportions* **1981**, *2*, 22–23.
38. Sim, J.; Wright, C.C. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys. Ther.* **2005**, *85*, 257–268. [[CrossRef](#)] [[PubMed](#)]
39. Irino, T.; Patterson, R.D.; Kawahara, H. Speech Segregation based on Fundamental Event Information Using an Auditory Vocoder. In Proceedings of the Eighth European Conference on Speech Communication and Technology, Geneva, Switzerland, 1–4 September 2003.
40. Glasberg, B.R.; Moore, B.C.J. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *J. Acoust. Soc. Am.* **2000**, *108*, 2318–2328. [[CrossRef](#)] [[PubMed](#)]
41. Erzin, E.; Yemez, Y.; Tekalp, A.M. Multimodal speaker identification using an adaptive classifier cascade based on modality reliability. *IEEE Trans. Multimed.* **2005**, *7*, 840–852. [[CrossRef](#)]
42. Chanwoo, K.; Stern, R.M. Power-normalized cepstral coefficients (PNCC) for robust speech recognition. *IEEE ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 1315–1329.
43. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2000**, *97*, 105524. [[CrossRef](#)]
44. Delavarian, M.; Towhidkhah, F.; Dibajnia, P.; Gharibzadeh, S. Designing a decision support system for distinguishing ADHD from similar children behavioral disorders. *J. Med. Syst.* **2010**, *36*, 1335–1343. [[CrossRef](#)]
45. Yu, H.; Xie, T.; Paszczynski, S.; Wilamowski, B.M. Advantages of radial basis function networks for dynamic system design. *IEEE Trans. Ind. Electron.* **2011**, *58*, 5438–5450. [[CrossRef](#)]
46. Fath, A.H.; Madanifar, F.; Abbasi, M. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. *Petroleum* **2020**, *6*, 80–91. [[CrossRef](#)]
47. Akobeng, A.K. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr.* **2007**, *96*, 644–647. [[CrossRef](#)]
48. Zhao, X.; Wang, D. Analyzing Noise Robustness of MFCC and GFCC Features in Speaker Identification. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 7204–7208.
49. Munich, M.E.; Lin, Q. Auditory Image Model Features for Automatic Speech Recognition. In Proceedings of the Ninth European Conference on Speech Communication and Technology, Lisboa, Portugal, 4–8 September 2005.

-
50. Irino, T.; Patterson, R.D. A dynamic compressive gammachirp auditory filterbank. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 2222–2232. [[CrossRef](#)]
 51. Jiang, Y.; Peng, J.; Zhang, X. Automatic snoring sounds detection from sleep sounds based on deep learning. *Phys. Eng. Sci. Med.* **2020**, *43*, 679–689. [[CrossRef](#)] [[PubMed](#)]