Snoring sound classification using multiclass classifier under actual environments

Keisuke Nishijima* and Ken'ichi Furuya*
*Oita University, Oita, Japan
E-mail: { k-nisijima, furuya-kenichi }@Oita-u.ac.jp

***Abstract*— The problem with conventional snoring sound identification methods is that their performance declines when the snoring sound is identified in the actual environment. Therefore, it is necessary to cope with the stationary and nonstationary environmental sounds that cause the decrease. In this research, we tried to cope with stationary environmental sounds by spectrum subtraction method for noise suppression. Non-stationary environmental sounds were regarded as one class for each type of environmental sound. We tried to identify the snoring sounds by multikernel learning, which is a multiclass extension of a support vector machine and by multilayer perceptron, which is a kind of neural network.**

I. Introduction

This research is aimed at the construction of a system to easily detect sleep states using a common mobile information terminal, thus eliminating the need for dedicated hardware. It collects sleep sounds through a mobile information terminal, analyzes the sleep state, accumulates data at the server, and provides the user with the analysis results. An overview of the system is shown Fig. 1. Currently, we are working on the identification of snoring sounds and environmental sounds.

Conventional research on snoring identification involves learning using a support vector machine (SVM) [1] based on the acoustic features of snoring [2, 3]. The average sound pressure level (SPL) and mel-frequency cepstrum coefficients (MFCCs) are learned and identified by the SVM as acoustic features. In these studies, the sleeping environment was assumed to be silent, which led to the deterioration of snoring sound classification performance owing to the presence of environmental sounds in an actual living environments.

In this paper, we consider the difficulty of characterizing nonsnoring sound classes when nonstationary environmental sounds with various characteristics are considered. Therefore, we used the method of environmental adaption that uses the environmental sounds of the sleep environment and the spectral subtraction method for noise suppression for the stationary environmental sound. For various nonstationary environmental sounds, we tried to improve the identification accuracy using multiclass classification by an SVM using multikernel learning (MKL), and multilayer perceptron (MLP), which is one of the deep learning methods.

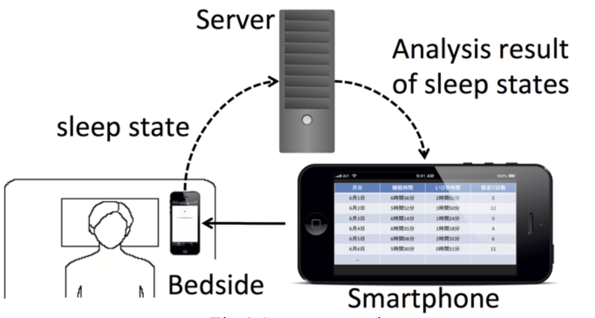

Fig.1 System overview

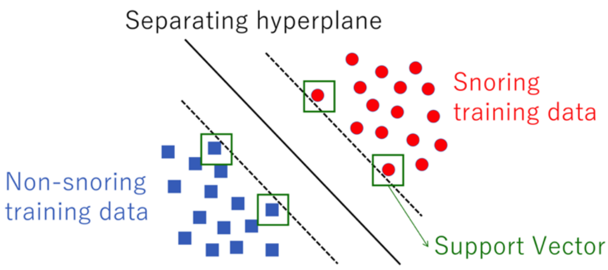

Fig.2 Overview of SVM

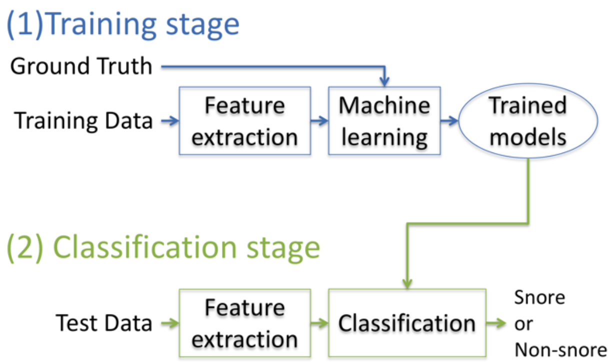

Fig.3 Snoring sound classification with SVM

II. Conventional Research

Waida [1] investigated the peak frequency, envelope, MFCCs, and formant frequency as acoustic features for the snoring sound detection algorithm. Kashina [2] added sound pressure levels to these acoustic features, and used three classifiers: Fisher's linear discriminant analysis, linear discriminant analysis, and SVM using a linear kernel. Then, the combination of acoustic features and classifiers were evaluated.

As a result, it was confirmed that the MFCCs are effective as an acoustic feature, and the SVM is effective as a classifier.

An SVM classifies the data by determining the best hyperplane that separates all the data points of a snoring sound class from those of a nonsnoring sound class (Fig. 2). Snoring sounds are detected by using a model learned by the SVM. The snoring sound classification process is shown in Fig. 3. Snoring sound classification consists of a training stage and a classification stage. In the training stage, after the noise suppression of the training data noise, the acoustic features are extracted. Machine learning is performed using these acoustic features and ground truths, and a learning model is created. In the classification stage, noise suppression is performed on the test data, and then feature extraction is performed. The acoustic features and trained model are used to classify the test data into the snoring and nonsnoring sound classes.

III. Approach

Conventional research assumes a noiseless environment but when the method is applied to the actual living environments, the classification accuracy is decreased. The decrease is attributed to stationary environmental noise such as air conditioning noise, and nonstationary environmental noise such as the creaking of the wall. For stationary environmental sounds, we tried to improve the classification accuracy by introducing noise suppression before the feature extraction and the method of environmental adaptation using the environmental sounds of the sleep environment. For nonstationary environmental sounds, considering these sounds with various features as just one class of sounds could be the cause of accuracy degradation. Therefore, we tried to improve this by classifying nonstationary environmental sounds into different classes. For multiclass classification, we consider MKL-SVM and MLP.

IV. Proposed Method

We tried to improve classification accuracy by noise suppression and adaptation to the sleep environment for stationary environmental sounds, and by multiclass classification for nonstationary environmental sounds. Multiclass classification utilizes two methods: MKL-SVM which extends the conventional method, and MLP.

A. Noise Suppression

The spectral subtraction (SS) method is used for noise suppression. In the frequency domain, noise $\hat{N}(\omega)$ is estimated from sleep sound $X(\omega)$ in which there is no snoring at the beginning of sleep. By subtracting the estimated power spectrum of noise, a sleep sound $\hat{S}(\omega)$, in which the noise is suppressed, is obtained. Let $\angle X(\omega)$ be the phase spectrum, α the subtraction parameter, and β the flooring parameter.

$$|\hat{S}(\omega)|^2 = max\left\{ |X(\omega)|^2 - \alpha|\hat{N}(\omega)|^2 , \beta|X(\omega)|^2 \right\} \tag{1}$$

$$\hat{S}(\omega) = |\hat{S}(\omega)| \cdot \exp(j\angle X(\omega)) \tag{2}$$

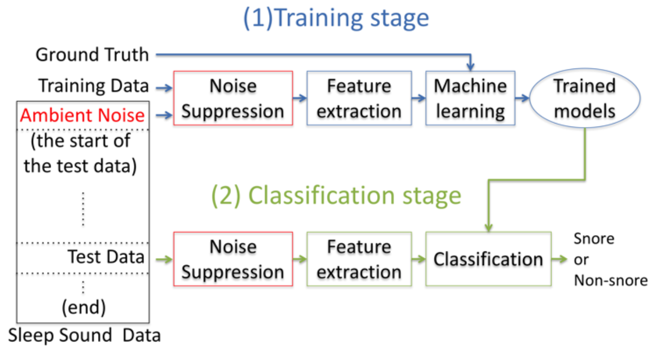

Fig.4 Snoring sound classification with SVM using ambient noise adaptation

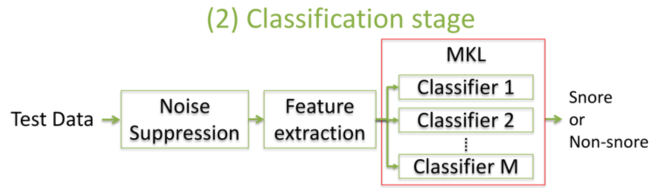

Fig. 5 Snoring sound classification with MKL

The parameter α controls the amount of noise subtracted from the noisy signal. The flooring parameter β is a positive value close to zero.

B. Adaptation to the Sleep Environment

Usually, it takes some time to go to sleep after going to bed. During that time, no snoring sound is generated, and only environmental sounds of the sleeping environment are included. By using this environmental sound data as training data for learning by SVM, an SVM model adapted to stationary environmental sound is obtained (Fig. 4). This makes it possible to classify snoring sounds adapted to the sleeping environment. However, it is necessary to consider how much should be added to the training data.

C. Snoring Sound Classification by Multiclass

In the conventional method, two-class classification was used to distinguish between snoring and environmental sounds. If the sleep sound consists only of snoring sound and stationary environmental sounds, it can be classified as the sleep sound of the actual sleeping environment. However, the problem is that classification performance declines in the two-class classification when nonstationary environmental sounds, such as the creaking of the wall, are added. In this method, environmental sounds other than snoring are treated as multiple classes instead of one class.

D. Snoring Sound Classification by MKL

The use of multiple SVMs for two-class classification by the conventional method enables multiclass classification by multikernel learning [4]. The one-versus-one method creates a two-class classifier for all class combinations. This is the same process as the training stage in Fig. 4. In the classification stage, the extracted acoustic features as shown in Fig. 5 are input into all created classifiers. The classes are decided by majority voting based on the classification results of all classifiers;

finally, the snoring or nonsnoring classes are generated. Let M be the number of classifiers to be created, and k the number of classes to be classified.

$$M = \frac{k(k-1)}{2} \tag{3}$$

E. *Snoring Sound Classification by MLP*

In MKL classification, as the number of classes increases, the number of classifiers also increases, and learning and classification take time. Therefore, we consider using MLP, a kind of neural network, that can classify multiple classes with one classifier. In MLP, learning is performed in multiple classes for each feature of the environmental sounds, and the output of the MLP is classified into two classes based on snoring sounds and the environmental sounds.

Fig. 6 shows the process flow for classifying snoring and environmental sounds.The sleep sound is divided into frames, and after the noise suppression, the acoustic features are extracted. Each feature is normalized and passed on the input unit. One feature corresponds to one unit of the input layer. In the MLP, the input layer receives data, and the output of each layer becomes the input of the next layer. The units of each layer combine the weights of the input features and assign the importance of each features. The network classifies the input based on this and outputs it in the output layer. The number of units of the output layer is determined by the number of classes. Based on the output result of MLP, the final classification results for the snoring sounds or the environmental sounds can be obtained.

V. EXPERIMENT

The objective of the experiment is to evaluate the classification accuracy of the proposed method for snoring and environmental sounds using noise suppression, adaptation to environmental sounds, and multi-class classification (MLP, MKL-SVM).

A. *Recording of Sleep sounds and Creation of data*

The sleep sounds were obtained by placing the mobile information terminal at the bedside in the home of the subject. An iPod touch and the iOS application were used for recording.

There were 10 subjects and three days of recorded sleep sounds. Since one subject's sleep sound did not include snoring, data for nine subjects were used to create the experimental data.

Seven classes of snoring sounds, nonsnoring sounds (stationary), running train sounds, alarm sounds, creaking sounds, crowing sounds, and running car sounds were extracted from the recorded sleep sounds. The experimental conditions are listed in Table 1. The ground truth used for learning was created manually. As acoustic features, the average SPL in a frame was used as one dimension, and MFCCs as 12 dimensions. A total of 13-dimensional input vectors were used.

B. *Parameters of SS Method*

In order to apply the SS method, an experiment was conducted to study the appropriate parameters. The flooring

Table 1 Experimental conditions

Sampling frequency	44.1 KHz
Quantization bit rate	16 bit
Frame length	25 ms
Shift length	10 ms
FFT size	2048

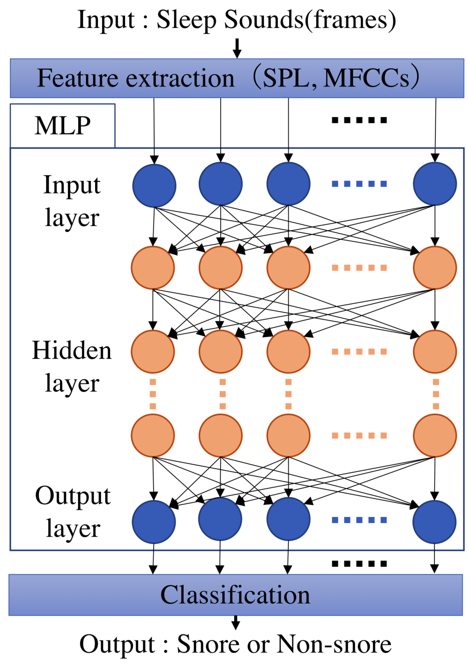

Fig. 6 Snoring sound classification with MLP

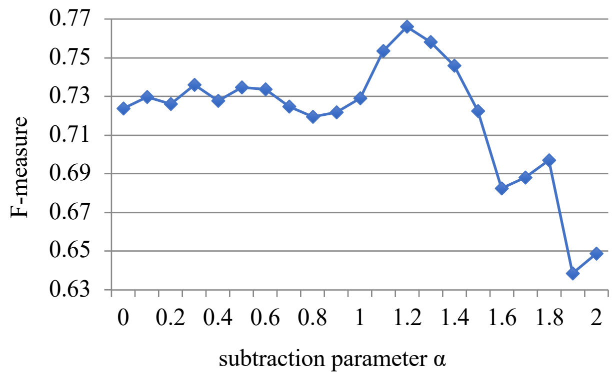

Fig. 7 Classification performance of each subtraction parameter α

parameter β was set to 0 by preliminary experiments. The subtraction parameter α was changed from 0.0 to 0.1 in steps of 2.0. The data for one subject was used for training, and the remaining data for eight subjects were used for the test. The data for each subject includes 448 of snoring class and 1036 of non-snoring (stationary) class. An F-measure was used to evaluate the classification accuracy. The experimental results are shown in Fig.7. The best result was obtained when the subtraction parameter α = 1.2. The F-measure improved from 0.72 with no noise suppression (α = 0) to 0.77 when α = 1.2.

C. Parameter of Adaptation to the Sleep Environment

The relationship between the number of environmental sound data used in learning for adaptation to the sleep environment and classification accuracy was verified experimentally. One group of data was selected for every nine subjects, and it was verified by nine-fold cross-validation. That is, eight groups were used as training data, learning was performed, the one remaining group was evaluated as test data, and the average of all combinations was evaluated. The number of snoring and nonsnoring sound classes in one group was 180, respectively, and the sleep environment sound was added to the nonsnoring class at the time of training. However, the total number of nonsnoring classes during training did not change, and the ratio of the number of nonsnoring classes in training data and the sleep environment sound was changed. An F-measure was used to evaluate the classification accuracy. The experimental results are shown in Fig. 8. The F-measure increased from 0.85 to 0.88 by adding 92% of environmental sound.

D. Hyperparameters of MLP

We implemented MLP using Deep Learning for Java (DL4J) [6] and determined the hyperparameters using a grid search. For classification, 144 of snoring sounds, non-snoring sounds (stationary), running train sounds, alarm sounds, creaking sounds, crowing sounds, and running car sounds were used, respectively. Each was randomly divided into two groups, with half as training data and the other half as test data. Preliminary results show that the optimization algorithm used the stochastic gradient descent method; the activation function used the ReLU (ramp function); the batch size is 1700 so that learning converges as monotonically as possible; and the epoch number is 2000 . The hidden layer is of five types [1, 2, 3, 4, 5], and the number of units in the hidden layer is of 14 types [7, 13, 26, 39, 52, 65, 130, 260, 390, 520, 650, 650, 1300, 2600, 3900]. Two types of learning rates [0.1, 0.01] were used, and the optimum value was determined in these ranges. The evaluation index used accuracy as a percentage of correctly detected data in the test data. However, since classes other than snoring sounds are considered as nonsnoring classes, only classes that were detected as snoring classes were misidentified as nonsnoring classes. The best classification accuracy when changing the hidden layer from 1 to 5 is shown in Fig. 9. Table 2 shows the best number of hidden units for each number of hidden layers. The combination with the best average accuracy for each class is 1 hidden layer, 1300 hidden units, and learning rate of 0.01. At that time, the classification accuracy is 95.8% for snoring sounds, 98.8% for nonsnoring sounds (stationary), and 86.9% for average of nonstationary. As the number of hidden layers increases, the classification accuracy of the snoring class decreased; conversely, the classification accuracy of nonsnoring sounds (stationary) tends to improve. However, the difference in the classification accuracy of the best parameter of each hidden layer is close but not significant.

In neural network learning, a large amount of learning data is required. In this experiment, the number of training data used may be insufficient, which could have affected the accuracy.

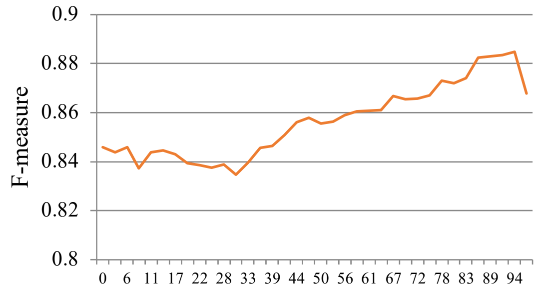

Fig. 8 Classification performance of each ratio of ambient noise

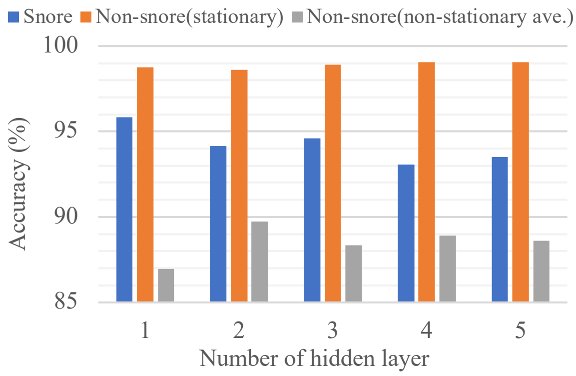

Fig. 9 Accuracy of each hidden layers

Table 2 Hidden layers and number of best hidden units

hidden layers	1	2	3	4	5
units per layer	1300	130	130	52	65

There is also the possibility that there are better combinations of network configuration and hyperparameters.

E. Classification Accuracy of MKL-SVM and MLP

The two learning methods of MKL-SVM, and MLP were compared. For classification, snoring sounds, nonsnoring sounds (stationary), running train sounds, alarm sounds, creaking sounds, crowing sounds, and running car sounds were used 144 data, respectively. Each was randomly divided into two groups, with half as training data and the other half as test data.

MKL-SVM was implemented using LIBSVM [5]. MKL-SVM made 21 one-versus-one classifiers for seven classifications. Each classifier used the conventional method with noise suppression and adaptation to the sleep environment. MLP was implemented the same way as in subsection D. From the result of D, a three-layer perceptron consisting of the input layer, hidden layer (one layer), and output layer was used, hyper parameters are 1300 hidden units, learning rate is 0.01, and batch learning was performed with a size of 1700 and an epoch number of 2000.

The evaluation index used the accuracy as a percentage of correctly-detected data in the test data. Fig. 10 shows the classification accuracy of each class of MKL-SVM and MLP. The classification accuracy of the snoring sounds was 95.8% of MLP and 96.8% of MKL-SVM. Nonsnoring sounds (stationary) was 96.3% of MLP and 94.4% of MKL-SVM. The classification accuracy of the average of the five nonstationary sounds was 81.1% of MLP and 79.7% of MKL-SVM. The classification accuracy of snoring sounds was 1% better for MKL-SVM, and the classification accuracy for nonsnoring sounds (stationary) and the average of five nonstationary sounds was better for MLP.

F. *Classification Accuracy of Snoring Sound*

For comparison, two-class classifiers with SVM were added to MKL-SVM and MLP. The classification accuracy of snoring sounds was compared by three methods. SVM used the conventional method with noise suppression and adaptation to the sleep environment. MKL-SVM and MLP used the same method as in subsection E. For classification, snoring sounds, nonsnoring sounds (stationary), running train sounds, alarm sounds, creaking sounds, crowing sounds, and running car sounds were used 144 data, respectively. Each was randomly divided into two groups, with half as training data and the other half as test data.

Table 3 shows the classification accuracy between snoring and nonsnoring sounds. For the nonsnoring sounds, the conventional SVM had poor classification accuracy of the nonstationary environmental sound. Therefore, the nonsnoring sounds (stationary) and the average of the five classes of nonstationary environmental sounds ware evaluated separately. The classification accuracy of snoring sounds was improved by 13.2% for MKL-SVM and 12.2% for MLP compared with the conventional SVM (two-class classification). Nonsnoring sounds (stationary) improved MKL-SVM by 1.7% and MLP by 5.3%. The average of five nonstationary sounds was 13.9% for MKL-SVM and 16.1% for MLP. MKL-SVM is the method with the best classification accuracy for snoring sounds, while MLP has the best classification accuracy for nonsnoring sounds. MLP was the best on average for all classes.

Table 3 Classification accuracy of snoring and environmental sounds

Method	snoring	nonsnoring (stationary)	aveage of nonstationary
SVM	83.6%	93.5%	70.8%
MKL-SVM	96.8%	95.2%	84.7%
MLP	95.8%	98.8%	86.9%

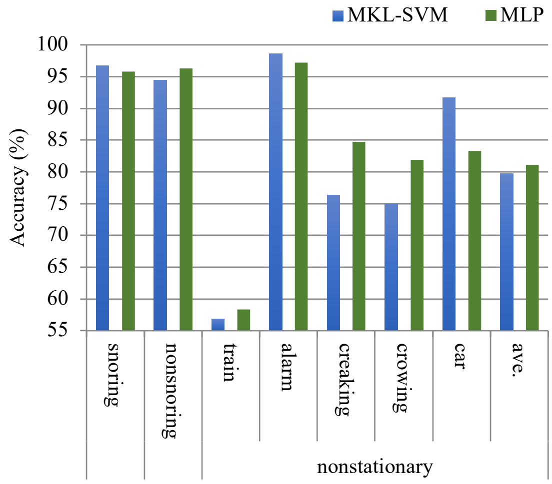

Fig. 10 Classification performance of MKL-SVM and MLP

VI. Conclusions

The problem with snoring sound classification in a actual environment is that classification accuracy is reduced by the presence of stationary and nonstationary environmental sounds. To solve this, we adopted the spectral subtraction method for noise suppression and adaptation to the sleeping environment to cope with the stationary environmental sounds. In addition, the classification accuracy was improved by classifying nonstationary environmental sounds into multiple classes. Compared with the conventional SVM, the classification accuracy for nonstationary environmental sounds was greatly improved by multiclass classification. MKL-SVM has the best in the classification accuracy for snoring sounds while MLP was the best in the other classes. MLP was the best on average for all classes. The optimization of hyperparameters in MLP and the increase in the number of training data will be further studied.

Acknowledgment

This work was supported by JSPS KAKENHI grant numbers 19K12044, 18K11377 and 16K00245.

References

[1] Kashina, Y., Waida, M., Nakashima, H., Miyoshi, K., “Classifier implementation and evaluation for snoring detection system, ” Proceedings of 2014 Spring Meeting of Acoustical Society of Japan, pp. 801–804, 2014.

[2] Waida, M., Kashina, Y., Nakashima, H., Miyoshi, K., “Sounds feature extraction for breath state detection algorithm, ” Proceedings of 2014 Spring Meeting of Acoustical Society of Japan, pp. 805–808, 2014.

[3] Bishop, C.M., “Machine Learning and Pattern Recognition, ” Springer, Berlin, 2006.

[4] Chih-Chung Chang , Chih-Jen Lin, “LIBSVM: A Library for Support Vector Machines,” 2013. http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

[5] LIBSVM, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[6] DL4J, https://deeplearning4j.org/index.html

[7] K. Nishijima, Shingo Uenohara and Ken'ichi Furuya, “Evaluating Classification Methods in Snore Activity Detection,” Proc. of the 12th International Conference on Complex, Intelligent, and Software Intensive Systems, pp.921-926, 2018.

[8] Nishijima, K., Uenohara, S., Furuya, K., “Estimation of noise suppression parameters for maximizing snoring activity detection performance,” Proceedings of International Conference on Consumer Electronics-Taiwan, ICCE-TW 2017, pp. 21–22. IEEE, 2017.

[9] Nishijima, K., Uenohara, S., Furuya, K., “A study on the optimum number of training data in snore activity detection using SVM,” Proceedings of 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS2016), pp. 582–584, 2016.