AUTOMATIC CLASSIFICATION OF ORAL/NASAL SNORING SOUNDS BASED ON THE ACOUSTIC PROPERTIES

Tsuyoshi Mikami[⋆] *Yohichiro Kojima*[⋆] *Masahito Yamamoto*[†] *Masashi Furukawa*[†]

[⋆]Tomakomai College of Technology
Nishikioka 443, Tomakomai, 059-1275 Japan
[†]Graduate School of Information Science, Hokkaido University
Kita 14 Nishi 9, Sapporo, 060-0814 Japan

ABSTRACT

Snoring was once regarded as an indication of good sleep. But recently it has been known to be one of the symptoms which indicate sleep disordered breathing such as sleep apnea syndrome. Moreover, heavy snoring caused by oral breathing sometimes leads benign snorers to be apneics. Thus, it is important to detect oral snoring for medical treatment in the earlier stage, but we cannot know our own snoring. This paper describes a method to detect oral snoring by extracting the acoustic properties of snoring sounds and using the k-Nearest Neighbor classifier. As a result, over 92% of snoring sounds are successfully classified under the various cross validation evaluations.

Index Terms— Biomedical Signal Processing, Pattern Recognition, Snoring Sounds

1. INTRODUCTION

In this paper, we propose a novel method to classify oral/nasal snoring using the acoustic properties of snoring sounds: fundamental frequency and the maximum of the amplitude spectrum in a specific band. The purpose of this classification is to develop a home medical device which detects an abnormal oral-related snoring automatically at bedside. Snoring was once regarded as an indication of good sleep, but recently it has been known to be one of the symptoms which indicate sleep disordered breathing such as sleep apnea syndrome [1]. Especially, heavy snoring caused by oral breathing sometimes leads benign snorers to be apneics. Thus, it is important to detect oral snoring during sleep in the earlier stage, but we cannot know whether our own snoring is abnormal or not.

Many medical researchers have analyzed snoring sounds so far in an attempt to clarify the difference between the acoustic properties of snoring sounds in patients with and without obstructive sleep apnea (surveyed in [2]), but they have not focused on the acoustic properties of breathing route (mouth/nose) during snoring and no concrete method for the automatic classification has been established. In this paper, we analyzed the acoustic properties of snoring sounds and propose an automatic classification method so as to develop a home device that can detects oral snoring.

2. CLASSIFICATION PROCEDURE

2.1. Data Acquisition

A portable linear PCM (Pulse Code Modulation) sound recorder, Olympus LS-10, is used to record snoring sounds. Sampling frequency and quantization rate are set to 14.7 kHz and 16 bit respectively. Snoring sounds are recorded from seven subjects and the recording time is about 15 seconds per person and per breath.

All subjects are asked to simulate snoring by breathing deeply enough to vibrate the soft palate in their throat. While producing oral snores, the subjects' nostrils are completely closed with their fingers, and on the other hand they are asked to let their mouth completely closed while producing nasal snores. Since two out of seven subjects cannot produce snoring by breathing nasally, as also reported by Liistro, et al[3], we abandoned obtaining the nasal simulated snores from these persons.

Such snoring, called *simulated snoring* in common, is not the one generated from a person during sleep, but it has traditionally been adopted in some medical studies [3][4]. Herzog and coworkers analyzed the difference between acoustic properties of simulated snoring and those of natural (nocturnal) snoring, and demonstrated that they are quite similar to each other, especially in the case of periodic snoring[4]. According to these studies, we decided to deal with simulated snoring sounds in this study.

2.2. Subsequence Extraction

First of all, a snoring sound with each inhalation (called *an episode*) is cut out manually one by one from the recorded sleep sounds (see fig.1). Since the acoustic properties of snoring sounds are nonstationarily changing as time passes even in an episode[5], we extracted short-time subsequences from all episodes by sliding the window across the episode. This

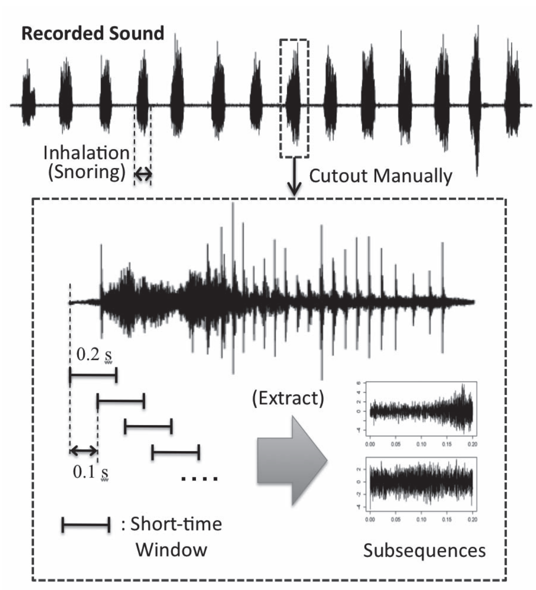

Fig. 1. An extraction method of episodes and subsequences of snoring sounds from a recorded sound.

technique is commonly used in speech recognition. The windows prepared for extracting subsequences are 0.2 seconds in length and shifted 0.1 seconds. The ith extracted subsequence is expressed as $x_i(t)$. As a result, we can obtain 710 oral and 511 nasal subsequences from all subjects, and they are our classification targets. Since two out of seven subjects cannot produce nasal snoring as mentioned above, the number of nasal subsequences is somewhat lower than that of oral ones.

2.3. Feature Extraction

2.3.1. Fundamental Frequency

Figures 2 and 3 show some examples of subsequences $x_i(t)$ and their FFT amplitude spectra expressed as $|X_i(f)|$. According to these figures, it seems easy to find out qualitatively some differences between the acoustic properties of oral snores and those of nasal ones.

First, we focused on the fundamental frequency, which is known as the eigenfrequency of the soft palate vibration[3], because the period of nasal snoring sounds are lower than that of the oral ones. Many pitch detection algorithms have been proposed so far, but we used in this paper *Harmonic Product Spectrum* (HPS) method, because the harmonic peaks identified in lower frequency domain in these figures are suitable for the HPS algorithm. The HPS is defined as follows:

$$H_i(f) = \prod_{m=1}^{r} |X_i(mf)| \tag{1}$$

where r is the number of harmonic peaks in the frequency domain. Thus, $H_i(f)$ has a single prominent peak at the fundamental frequency by multiplying the down-sampled amplitude spectra. So we can solve the fundamental frequency f_b by using the following criterion:

$$H_i(f_\mathrm{b}) = \max\ H_i(f) \tag{2}$$

But there is a variety of the number of harmonic peaks in the snore spectra. It is important to adjust the parameter r to a suitable value.

2.3.2. The Maximum of the Amplitude Spectrum in the Specific Band

According to figures 2 and 3, there are some intensity peaks around 900 Hz in the amplitude spectra of oral snores, whereas no such peaks exist in the spectra of nasal ones. Therefore, we considered such intensity peaks as a useful property to discriminate oral snores from nasal ones, and defined *the maximum of the amplitude spectrum in the specific band* as follows:

$$|X_i(f_m)| = \max_{f_1 \le f \le f_2} |X_i(f)| \tag{3}$$

Namely, the maximum of the amplitude spectrum is obtained at f_m Hz, which is also greater than or equal to f_1 Hz and less than or equal to f_2 Hz. In this case, it is necessary to adjust f_1 and f_2 to suitable values so as to realize the best performance.

2.4. Classification

A combination of two acoustic properties $(|X_i(f_m)|, f_b)$ estimated from the ith subsequence is defined as 2-dimensional feature vector expressed as $\mathbf{x}_i = (|X_i(f_m)|, f_b)$. In this paper, we adopt *k-Nearest Neighbor* (kNN) classification method, which assigns the class label which is the most frequent among the k reference data closest to the input whose class is unknown.

3. PERFORMANCE EVALUATION

3.1. 10-fold Cross Validation Test

The classification performance is evaluated using *10-fold cross validation* (10-fold CV) test. In this test, all feature vectors are divided into 10 groups, $G_1, G_2, \cdots, G_{10}$, at random. The data belonged to G_1 are used as the test data and the remainder as the reference data. Then, the kNN method estimates the classification result of the test data using the reference data. The classification rate of the input data, r_1, is calculated by comparing their correct class labels. The same procedure is also done for the groups $G_2, G_3, \cdots, G_{10}$ respectively, and the corresponding classification rates $r_2, r_3, \cdots, r_{10}$ are also calculated. Finally, the classification rate of all data is obtained by

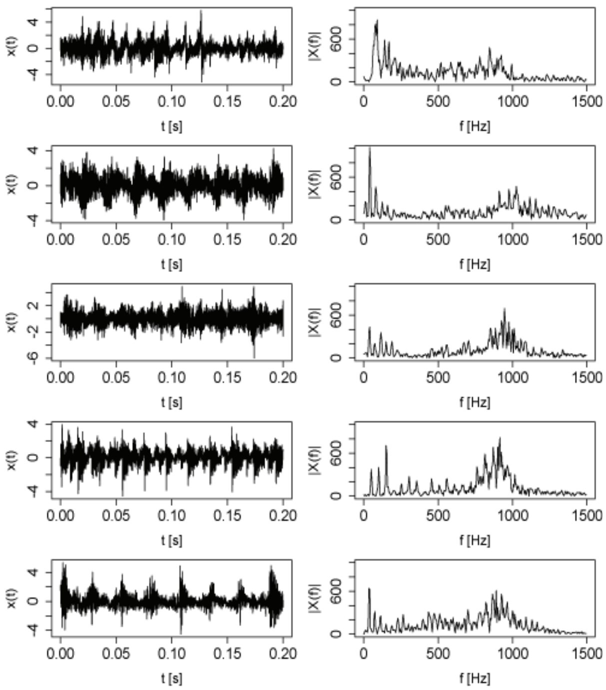

Fig. 2. Subsequences of oral snoring and their amplitude spectra.

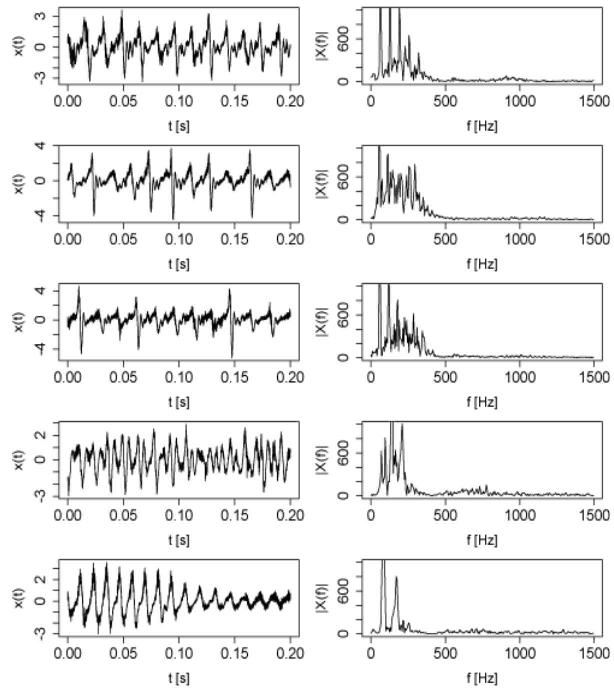

Fig. 3. Subsequences of nasal snoring and their amplitude spectra.

$R = \sum_{j=1}^{10} |G_j| r_j / N$, where N is the number of all data. So far, the 10-fold CV test has been widely used in pattern recognition studies, but, in order to demonstrate the usefulness of our method more objectively, we consider two more different ways of dividing the data into groups.

3.2. Leave-One-Out Test

Next we tried to use as many reference data as possible, so we assigned only one datum to the group G_j. Namely, the number of groups is equal to the number of all data. The other procedures are the same as 10-fold CV. This evaluation test is called *Leave-One-Out* (LOO) test in common and has been adopted in many studies.

3.3. Leave-*Episode*-Out Test

On further consideration, it is possible that one subsequence may be quite similar to the ones extracted from the *same* episode. Even if they do not overlap each other, subsequences extracted from the same episode may be generated from the same vibration dynamics provoked by the same inhalation. Thus, in step 1, we defined the number of groups as the number of episodes and assigned the data extracted from the same episode into the same group. We call this evaluation method *Leave-Episode-Out test* in this study. According to this, there are not subsequences extracted from the same episode in a reference data set.

3.4. Leave-*Subject*-Out Test

An individual difference may be what we must consider the most in this study. It may not be deniable that the difference between the subjects is larger than that between their breathing routes. But we cannot evaluate such a difference using 10-fold CV or LOO test. Accordingly, we assigned the data obtained from the same subject into the same group, and the number of groups is the same as the number of subjects. The other procedures are the same as 10-fold CV. We call this evaluation method *Leave-Subject-Out* test in this study.

4. RESULTS

In this study, three parameters r, f_1, f_2 for feature extraction are determined to 3, 500 Hz,1500 Hz respectively. These values are obtained by maximizing the classification rate through experiments under the LSO test. Figure 4 shows scatter plots in the feature space where oral and nasal subsequences are represented with triangles and circles respectively. Oral snores are more widely scattered than nasal ones, but they are well separated from each other except a few outliers.

Figure 5 shows the classification results with different number of neighbors, k, under the four evaluation tests. According to this, classification rates under the 10-fold CV, LOO, and LEO tests are almost the same. But in the case of LSO test, the rate is about 3% lower than those calcu-

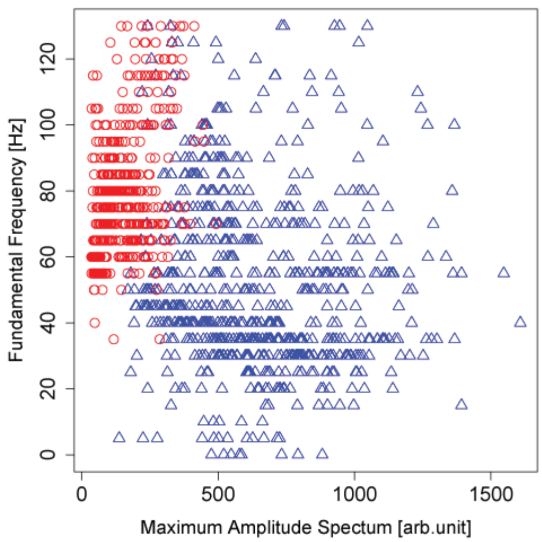

Fig. 4. Scatter plots of nasal (circles) and oral (triangles) snoring sounds on the 2-dimensional feature space

lated under the other evaluations. This indicates that there is a little individual difference in the acoustic properties of snoring sounds. But we can obtain a good performance; the classification rate is over 92%.

5. DISCUSSION

There are some intensity peaks from 500 Hz up to 1500 Hz in the frequency domain of oral snores, but we cannot find any in that of nasal ones. Such intensity peaks probably shows a *formant-like* resonance property in subject's throat modeled with a linear filter. Emoto, et al., [6] have demonstrated that the formant frequency of snoring sounds indicates the difference between benign snorers and apneics. Koutsourelakis, et al., [7] have mentioned that many apneics spend more time breathing orally than benign snorers during sleep. Based on these bibliographical review, it is possible that such intensity peaks indicate the formant frequency that also indicates the useful information about oral snoring and its relation to sleep apnea. This should be further analyzed using real snoring sounds obtained from apnea patients.

6. CONCLUSION AND FUTURE WORKS

Oral and nasal snoring can be successfully classified with good accuracy using two acoustic properties we focused on. In the future, more data should be collected for objective evaluation and the relation between the formant-like intensity peaks and sleep apnea should be clarified theoretically and experimentally.

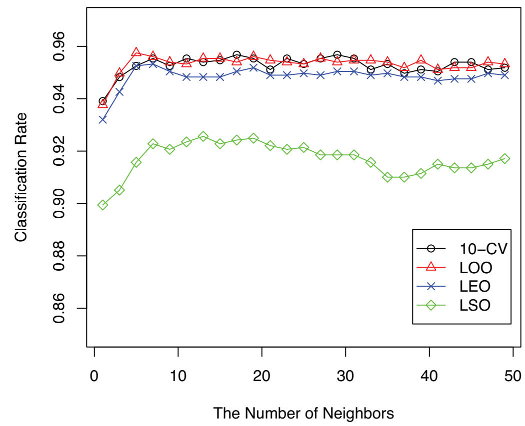

Fig. 5. Scatter plots of nasal (circles) and oral (triangles) snoring sounds on the 2-dimensional feature space

This study was supported in part by the Graint-in-Aid for Regional R&D Proposal-Based Program from Northern Advancement Center for Science and Technology of Hokkaido Japan.

7. REFERENCES

[1] M. Ichioka, "Respiratory organs and sleep disorders," *Folia Pharmacol. Jpn.*, vol. 129, pp. 432–435, 2007.

[2] D. Pevernagie, R. M. Aarts, and M. D. Meyer, "The acoustics of snoring," *Sleep Med. Rev.*, vol. 14, pp. 131–144, 2010.

[3] G. Liistro, D. Stanescu, and C. Veriter, "Pattern of simulated snoring is different through mouth and nose," *J. Appl. Physiol.*, vol. 70, no. 6, pp. 2746–2741, 1991.

[4] M. Herzog, E. Schieb, T. Bremert, B. Herzog, W. Hosemann, H. Kaftan, and T. Kuhnel, "Frequency analysis of snoring sounds during simulated and nocturnal snoring," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 265, no. 12, pp. 1553–1562, 2008.

[5] F. Dalmasso and R. Prota, "Snoring: analysis, measurement, clinical implications and applications," *Eur. Respir. J.*, vol. 9, pp. 146–159, 1996.

[6] T. Emoto, U. Abeyratne, T. Kusumoto, M. Akutagawa, E. Kondo, I. Kawata, T. Azuma, S. Konaka, and Y. Kinouchi, "Discriminating apneic snorers and benign snorers based on snoring formant extracted via a noise-robust linear prediction technoque," *J. Jpn. Soc. Med. & Bio. Eng.*, vol. 48, no. 1, pp. 115–121, 2010.

[7] I. Koutsourelakis, E. Vagiakis, C. Roussos, and S. Zakynthinos, "Obstructive sleep apnoea and oral breathing in patients free of nasal obstruction," *Eur. Respir. J.*, vol. 28, pp. 1222–1228, 2006.