Classification of Oral/Nasal Simulated Snores Based on the Acoustic Properties

Tsuyoshi MIKAMI∗∗∗, Yohichiro KOJIMA∗∗, Kazuya YONEZAWA∗∗∗, Masahito YAMAMOTO† and Masashi FURUKAWA†

∗ Tomakomai College of Technology
Nishikioka 443, Tomakomai, Hokkaido 059–1275, Japan
E-mail: mikami@jo.tomakomai-ct.ac.jp, kojiasan@gt.tomakomai-ct.ac.jp
∗∗ Department of Clinical Research, Hakodate National Hospital
Kawaharacho 18-16, Hakodate, Hokkaido 041–8512, Japan
E-mail: kyonezaw@hnh.hosp.go.jp
∗∗∗ Department of Clinical Research, Hakodate National Hospital
Kawaharacho 18-16, Hakodate, Hokkaido 041–8512, Japan
E-mail: kyonezaw@hnh.hosp.go.jp
† Graduate School of Information Science and Technology, Hokkaido University
Kita 14 Nishi 9, Sapporo, Hokkaido 060–0814, Japan
E-mail: masahito@complex.ist.hokudai.ac.jp, mack@complex.ist.hokudai.ac.jp

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. Especially, loud snoring caused by oral breathing during sleep is often found in many apnea/hypopnea patients. 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. According to the FFT amplitude spectra, nasal snoring sounds consist of only lower frequency components less than 500Hz, whereas oral snoring sounds consist of unique intensity peaks at 1kHz and lower frequency components less than 500Hz as well. According to the bibliographical point of view, such lower frequency components indicate the palatal snoring, and the intensity peak at around 1kHz indicates the tongue base snoring. Therefore, it is obvious that nasal snoring sounds can be regarded as simple palatal snoring whereas oral snoring sounds are a mixture of palatal and tongue base snoring sounds. So, we focused on the fundamental frequency and maximum of the amplitude spectrum in a specific band. In this paper, the Harmonic Product Spectrum (HPS) method is used for estimating the fundamental frequency and the k-Nearest Neighbor method is adopted for classifying oral/nasal snoring sounds. As a result, over 89% of snoring sounds are successfully classified under the four kinds of cross validation evaluations.

Key words: 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 irregular 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). Under normal circumstances, breathing during sleep is primarily nasal rather than oral(2), but numerous investigations have shown that loud habitual snoring is due to nasal obstruction(3)(4), and nasal obstruction alters airflow dynamics and leads to oral breathing during sleep(5). Since oral breathing tends to make the upper airway more collapsible(6), loud snoring caused by oral breathing is found in many sleep apnea/hypopnea patients. Thus, it
is important to detect oral snoring during sleep in the earlier stage in consideration of the medical treatment, but we cannot know whether our own snoring is irregular 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 (OSA)(7)(8)(9)(10)(11)(12)(13). Perez-Padilla, et al(7) firstly demonstrated that apnea patients had residual energy at 1kHz whereas the nonapneic snorers did not. So, they said that the ratio of power above 800Hz to power below 800Hz could be used to separate snorers from OSA patients. Fiz, et al(8) found the presence of a fundamental frequency and several harmonics in snoring sounds of many simple snorers and a low frequency peak with the second energy scattered on a narrower band and without clearly identified harmonics in those of OSA patients. On the other hand, Herzog, et al(9) showed that patients with primary snoring revealed peak intensities between 100 and 300 Hz, whereas OSA patients showed peak intensities above 1kHz. The PSG and BMI correlated with peak intensity of the power spectrum. Hara, et al(10) have reported that the sound spectrum of the simple snorers shows a single peak at a lower frequency, whereas the snores of OSA patients show multiple power peaks of various amplitudes.

Some researchers have focused on the formant analysis generally adopted in speech recognition technique. Sola-Soler, et al(11) have showed significant differences in formant frequencies variability between simple snorers and OSA patients even when non-postapneic snores were considered. Ng, et al(12) have also found quantitative differences in formant frequencies between apneic and benign snorers by the use of a Linear Predictive Coding (LPC) technique. Apneic snorers exhibit higher formant frequencies than benign snorers, especially F1, which can be related to the pathology of OSA. Emoto, et al(13) focused on the standard deviation of the estimated formant frequencies (F1) in snoring episodes over 6 hours, which is relatively higher in OSA patients than in simple (benign) snorers.

Accordingly, conventional studies have analyzed snoring sounds of simple snorers and those of apnea patients using FFT or LPC techniques, but their results are quite various. Thus, many researchers have still tried to clarify quantitative differences between acoustic properties of benign snores and those of apneic ones. Our approach, which is somewhat different, is to classify snoring sounds into nasal or oral. This classification can be applied to the automatic detection of oral snoring sounds which indicate the possibility of OSA occurrence for normal persons. Although open mouth and the corresponding loud snoring are highly related to OSA, few researchers have focused on the acoustic properties of oral and nasal snoring sounds and their classification. Dalmasso, et al(15), reported that the shape of cross sectional area in the upper airway are very similar during oral snoring and during oronasal snoring. But the shape of cross sectional area and its relative values can change during nasal snoring, but they always remain remarkably different from the other two (mouth and oronasal routes). Liistro, et al(16), reported that the frequency of airflow and supraglottic pressure oscillations were less during mouth than during nasal simulated snoring. These studies are informative for our purpose, but no concrete technical methods have been established to classify oral and nasal snoring sounds. In this paper, we analyze the acoustic properties of simulated snoring sounds with oral and nasal breathing and propose a concrete method for their classification.

2. Acquisition of Classification Targets

2.1. Apparatus, Subjects, and Simulated Snoring

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 44.1 kHz and 16 bit respectively. Snoring sounds are recorded from 15 subjects shown in table 1 and the recording time is about 30 seconds per person and per breath. Before recording, we explained the objective of this study and obtained the informed consent from all subjects.

The 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.

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\(^\text{(16)(17)(29)(30)}\). Similarities between simulated and natural (nocturnal) snoring are described in section 2.4 from the bibliographical point of view.

2.2. Definition of Snoring and Extraction of Snoring Episodes

First of all, a snoring sound produced with each inhalation (called a snoring episode) should be cut out one by one from the recorded sounds (Fig. 1), but in many cases the recorded sound contains non-snoring sounds (such as exhalation noise) as well. In order to distinguish snoring episodes from exhalation noises, the definition of snoring sounds should be considered.

According to some medical researchers\(^\text{(18)(19)}\), snoring is defined as the sound of pharyngeal vibration triggered by airflow turbulence across a narrowed upper airway. From a technical point of view, Abeyratne et al.\(^\text{(14)}\) proposed a paradigm to solve the issue of defining a snore, and figured out that sounds perceived as snores by humans are characterized by repetitively released packets of energy, which are responsible for creating the vibratory sound particular to snores and which define the pitch of snoring.

However, such vibration dynamics is complex and the acoustic properties are quite various\(^\text{(20)}\), because they depend highly on the anatomical site of snoring, such as soft palate, tongue, epiglottis, tonsils, and the mixture of them.\(^\text{(21)}\). Therefore, it is not easy to cut out the snoring episodes automatically\(^\text{(13)}\). In fact, many automatic extraction methods of snoring episodes have been proposed so far\(^\text{(22)(23)(24)(25)}\). For instance, Cavusoglu, et al.\(^\text{(22)}\) have tried to classify snores and non-snores according to their sub-band energy distributions in the frequency domain. Karunajeewa, et al.\(^\text{(23)}\) have focused on more various features - number of zero crossings, energy of the signal, normalized autocorrelation coefficient at 1 ms delay and the first predictor coefficient of LPC analysis - so as to classify snoring, breath, and silence. Accordingly, no decisive methods have been proposed yet and we did not apply any
Table 1: Detail of all subjects and the obtained snoring episodes and subsequences in this study

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th># of episodes</th>
<th># of subseq.</th>
<th>Apnea/Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>32</td>
<td>n:9, o:9</td>
<td>n:69, o:66</td>
<td>Benign</td>
</tr>
<tr>
<td>B</td>
<td>55</td>
<td>n:8, o:9</td>
<td>n:110, o:109</td>
<td>Benign</td>
</tr>
<tr>
<td>C</td>
<td>63</td>
<td>n:7, o:8</td>
<td>n:47, o:80</td>
<td>Apnea</td>
</tr>
<tr>
<td>D</td>
<td>52</td>
<td>n:6, o:8</td>
<td>n:21, o:43</td>
<td>Apnea</td>
</tr>
<tr>
<td>E</td>
<td>21</td>
<td>n:8, o:8</td>
<td>n:156, o:128</td>
<td>Apnea</td>
</tr>
<tr>
<td>F</td>
<td>35</td>
<td>n:0, o:9</td>
<td>n:0, o:101</td>
<td>Benign</td>
</tr>
<tr>
<td>G</td>
<td>46</td>
<td>n:8, o:8</td>
<td>n:94, o:131</td>
<td>Apnea</td>
</tr>
<tr>
<td>H</td>
<td>40</td>
<td>n:9, o:9</td>
<td>n:55, o:161</td>
<td>Benign</td>
</tr>
<tr>
<td>I</td>
<td>67</td>
<td>n:9, o:9</td>
<td>n:122, o:81</td>
<td>Benign</td>
</tr>
<tr>
<td>J</td>
<td>63</td>
<td>n:9, o:9</td>
<td>n:90, o:127</td>
<td>Benign</td>
</tr>
<tr>
<td>K</td>
<td>62</td>
<td>n:11, o:8</td>
<td>n:57, o:39</td>
<td>Apnea</td>
</tr>
<tr>
<td>L</td>
<td>56</td>
<td>n:9, o:8</td>
<td>n:99, o:98</td>
<td>Benign</td>
</tr>
<tr>
<td>M</td>
<td>43</td>
<td>n:8, o:7</td>
<td>n:47, o:120</td>
<td>Benign</td>
</tr>
<tr>
<td>N</td>
<td>38</td>
<td>n:8, o:9</td>
<td>n:48, o:55</td>
<td>Benign</td>
</tr>
<tr>
<td>O</td>
<td>22</td>
<td>n:9, o:7</td>
<td>n:52, o:45</td>
<td>Benign</td>
</tr>
</tbody>
</table>

automatic methods that have been proposed so far and decided to cut out the sound at each inhalation/exhalation (called candidates in this paper) manually from the recorded sounds.

On the other hand, we ordinary people can empirically distinguish snoring sounds from simple breathing noises(13). So, we asked three persons to judge whether the candidates are snoring episodes or not. Only the candidates three persons judged to be snores are adopted as snoring episodes in this paper. This procedure is the same as the one used in ref.(13).

2.3. Subsequence Extraction

Since the acoustic properties of snoring sounds are nonstationarily changing as time passes even in a snoring episode(15)(26), we extracted short-time subsequences from all episodes by sliding the window across the episode. This 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 $i$th extracted subsequence is expressed as $s_i(t)$, and in order to remove the effect of the microphone position $s_i(t)$ is normalized to $E[s_i(t)] = 0$ and $E[s_i(t)^2] = 1$ where $E[·]$ is an expectation operator. As a result, we can obtain 1384 oral and 1107 nasal subsequences from all subjects, and they are our classification targets.

2.4. Similarities between Natural and Simulated Snoring

Figs. 2 and 3 show the waveforms and the amplitude spectra of the subsequences extracted from snoring episodes produced by breathing orally and by breathing nasally. In the oral simulated snores, we can find an intensity peak at over 1kHz. Such acoustic property is also found in the natural snores the sound source of which is the tongue base(21), and the tongue base snoring is known to occur with an open oral airway(27). The same result is also reported by the other researcher(28). These conventional reports about the acoustic properties of natural oral/oronasal snores are consistent with those of simulated oral snores we obtained. In addition, Perez-Padilla et al indicated that the intensity peaks over 1kHz are found in OSA patients(7). Since many OSA patients tend to breathe orally during sleep as is described in section 1, it is possible that such intensity peaks are related to open mouth during snoring.

In the case of nasal simulated snores, we can find clearly periodic waveforms and the sound spectra which consist of lower frequency components below 500Hz. Such acoustic property is also found in the natural snores the sound source of which is the soft palate(21)(27). It has been reported that the palatal snoring occurs with the oral airway closed(27), and the inspiratory nasal snores have a fundamental peak with associated harmonic peaks(7). These conventional reports about the acoustic properties of natural nasal snores are consistent with those of simulated nasal snores we obtained. Based on these conventional studies, it is possible that the simulated snoring sounds are similar (not necessarily equal) to the natural ones produced by snorers during sleep. Moreover, simulated snoring has also been adopted in some
Fig. 2 Subsequences of oral snoring and their amplitude spectra.

Fig. 3 Subsequences of nasal snoring and their amplitude spectra.
medical studies\(^{(16)(17)(29)(30)(20)}\), and thus we decided to deal with simulated snoring sounds in this paper.

3. Classification Method

3.1. Feature Extraction

3.1.1. Fundamental Frequency

Figures 2 and 3 show some examples of subsequences \(s_i(t)\) and their FFT amplitude spectra expressed as \(|S_i(f)|\). We annotated on the upper right in the spectrum panels who and what episodes the subsequences are obtained from. For instance, “Subj-A, n4” indicates that the subsequence is extracted from the 4th nasal episodes of subject A. 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.

Liistro, et al\(^{(16)}\) found by observing the pharynx with cineradiography that during nasal snoring the uvula presents vibrations of relatively high frequency, whereas the whole soft palate vibrates with a lower frequency during oral snoring. Since this acoustic property is also found in our observed data shown in Figures 2 and 3, we adopt the fundamental frequency of snoring sounds as the first acoustic property for the classification. Many pitch detection algorithms have been proposed so far, but we used in this paper Harmonic Product Spectrum (HPS) method\(^{(31)}\). This method is useful for the vibrational sounds that have a unique fundamental frequency and its harmonics. Since such properties are also found in lower frequency domain (less than about 500Hz) in Figures 2 and 3, the HPS method is suitable to estimate the fundamental frequency of the snoring sounds.

The HPS of the amplitude spectrum of the \(i\)th subsequence is defined as:

\[
H_i(f) = \prod_{m=1}^{r} |S_{i,m}(mf)|
\]

where \(r\) is the number of harmonic peaks, \(|S_{i,m}(mf)|\) is the amplitude spectrum, and \(m\) is a scaling parameter. Figure 4 shows an example of HPS calculation when \(r\) is set to four. In this case, the HPS is calculated by \(H_i(f) = |S_{i,1}(f)| \times |S_{i,2}(f)| \times |S_{i,3}(f)| \times |S_{i,4}(f)|\). Namely, if the fundamental frequency is 80Hz, the HPS has the most prominent peak at 80Hz because the harmonic peaks are also found at 160Hz, 240Hz, and 180Hz in the amplitude spectrum. In this paper, the amplitude spectra of snoring sounds are low-pass filtered in advance with a cut-off frequency 500 Hz, because the vibration components lie in the frequency domain less than
about 500-600 Hz\(^{(27)(21)}\). As a result, \(H_i(f)\) has a single prominent peak at the fundamental frequency, because only the peak at the fundamental frequency is enhanced by multiplying the down-sampled amplitude spectra (Fig.4). Accordingly, we can solve the fundamental frequency \(f_b\) by using the following criterion:

\[
H_i(f_b) = \max H_i(f)
\]

But there is a variety of the number of harmonic peaks in the snore spectra. It is important to adjust the parameter \(r\) to the suitable value. But even if \(r\) is not determined to the correct number of harmonics, \(H_i(f_b)\) is sufficiently enhanced by multiplying the down-sampled amplitude spectra.

3.1.2. The Maximum of the Amplitude Spectrum in a Specific Band   According to figs. 2 and 3, there are some intensity peaks in a specific band over 1kHz, whereas no such peaks exist in nasal snoring sounds. Agrawal, et al\(^{(21)}\) have reported that such peaks over 1kHz indicate the tongue base snoring, caused by the turbulence which occurs when the airflow passes through the narrowed upper airway, by observing the pharynx with nasendoscopy, and in the case of soft palate snoring the spectral components are found in less than 500Hz. They also found the mixture of snoring sounds produced at the tongue base and those at the soft palate, which are similar to our results of oral snores. In addition, according to the fact that open mouth tends to make the upper airway more collapsible\(^{(1)}\), it is natural that the tongue base snores occur with oral breathing. Therefore, we considered such intensity peaks as a useful property to discriminate oral snores from nasal ones, and calculated the maximum of the amplitude spectrum in a specific band as follows:

\[
M = \max_{f_1 \leq f \leq f_2} |S_i(f)|
\]

Namely, the maximum of the amplitude spectrum is obtained in a specific band which is 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 the suitable values so as to realize the best performance.

3.1.3. A Criterion for the Parameter Adjustment   A combination of two acoustic properties \((M, f_b)\) is defined as the 2-dimensional feature vector expressed with \(x_i = (x_{i1}, x_{i2}) = (M, f_b)\) where \(M\) and \(f_b\) are estimated from the \(i\)th subsequence. But it is necessary to adjust some parameters, \(r\) in Eq.(1) and \(f_1, f_2\) in Eq.(3). In this paper, the ratio of within-class variance and between-class variance (called the variance ratio\(^{(32)(33)}\)) is used for an evaluation criterion for parameter adjustment. For a two-class problem, the variance ratio is defined as

\[
\gamma = \frac{(m_1 + m_2)^2}{\sigma_1^2 + \sigma_2^2}
\]

where \(m_j\) is the mean of and \(\sigma_j\) is the standard deviation of the feature values, and \(j = 1, 2\) indicates the class label (oral/nasal) respectively. If the data belonged to the same class are not widely scattered (i.e. \(\sigma_i\) is lower) and those belonged to the different class exist as far as possible from each other (i.e. \(|m_1 - m_2|\) is higher ), the variance ratio becomes higher. The higher variance ratio means it possible to discriminate oral from nasal snores easier in the feature space. Thus, these parameters are determined to the values with which the variance ratio is the maximum.

3.2. Classification   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. Though the kNN method is conceptually quite simple and the parameter is only one (\(k\); the number of neighbors), it is easy to obtain a nonlinear classification boundary.

Many classification methods have been proposed so far, such as Neural Networks, Decision Tree, Support Vector Machines, and so on\(^{(34)(33)}\). But for adopting them it is necessary to adjust many hyper parameters in advance and sometimes it causes the classification boundary
too complex and therefore the classification rate decreases in test examples\(^{(32)}\). In this paper, we focused mainly on the concrete method to extract the acoustic properties of breathing route during snoring for the purpose of oral snore detection. Accordingly, as for the classifier, we decided to adopt the simplest method, kNN classifier, which is available without adjusting many hyper parameters. Naturally, it cannot be denied that the other methods are more superior than the kNN method, and, therefore, the performance comparison is necessary for selecting the most suitable classifier. But this will be analyzed in the future.

4. Performance Evaluation

4.1. \(m\)-fold Cross Validation Test

The classification performance is evaluated using the \(m\)-fold cross validation (\(m\)-fold CV) test defined as the following procedure:

1. Divide all data \(x_i (i = 1, 2, \ldots, N)\) into \(m\) groups \(G_1, G_2, \ldots, G_m\).
2. Set \(j \leftarrow 1\)
3. Adopt \(G_j\) as a set of test data and the other data assigned to the remainder sets \(G_{j\neq j}\) as reference data for \(k\)-NN classification.
4. Calculate the classification rate \(r_j\) of the data belonged to \(G_j\) by the use of kNN method.
5. Set \(j \leftarrow j + 1\) and go to step 3 while \(j < m\).
6. Calculate the classification rate of all data with \(R = \frac{\sum_{j=1}^{m} |G_j| r_j/N}{|G_j|}\) where \(|G_j|\) is the number of data allocated to \(G_j\).

In this paper, \(m\) is set to 10 because 10-fold CV test has been widely used in pattern recognition studies\(^{(32)}\). But, in order to evaluate the usefulness of our method more objectively, we have to consider three more different ways of dividing all the data into groups.

4.2. 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 \(m\)-fold CV. This evaluation test is called Leave-One-Out (LOO) test in common and has also been adopted in many studies.

4.3. 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, the 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 (LEO) test in this study.

4.4. Leave-Subject-Out Test

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

5. Results and Discussion

5.1. Optimal Value of \(r\) and its Relation to the Acoustic Properties

First, the variance ratio of the fundamental frequency is calculated when the number of multiplication (\(r\)) for the HPS method is adjusted to 1-6 respectively. Figure 5 shows that the
highest ratio is obtained with $r = 3$, but when $r$ is set to the value greater than or less than 3, the ratio becomes lower. This is caused by two types of waveforms generally found in snoring sounds, simple-waveform and complex-waveform\(^{(20)}\). In the frequency domain, complex-waveform snores are characterized by multiple, equally-spaced peaks of power (comb-like spectrum). Otherwise, simple-waveform snores have only 1-3 peaks and in many cases the first or the second peak is the most prominent, so the waveform in the time domain looks quasi-sinusoidal.

An example of the incorrect estimation of the fundamental frequency is shown in figure 6, where the maximum peak does not lie at the fundamental frequency with $r = 1, 2$ in the left panels and with $r = 4$ in the right panels.

In the case of left panels, the reason is that the peak at the third harmonic frequency is the highest in the amplitude spectrum. Many complex-waveform snores do not have a single dominant peak, but two or more peaks which have almost the same power (just like comb). Thus, it is not rare that the two-times multiplication ($r = 2$) does not sufficiently enhance the peak at the fundamental frequency for the complex-waveform snores. In the case of right panels, on the other hand, since simple-waveform snores have only three peaks at most, four or more times multiplication is too much and lessens the peak of the fundamental frequency by multiplying the non-peak value at the fourth harmonic frequency. Thus, a slight protuberance occurred at a half of the fundamental frequency (visually recognizable in the panel of $r = 3$ at around 30-40Hz) becomes unfavorably competitive and the highest peak no longer exist at the fundamental frequency when $r = 4$.

5.2. Optimal Value of $f_1$ and $\Delta f$ and its Relation to the Acoustic Properties

Next, the most suitable value of the other two parameters ($f_1, \Delta f$) are estimated by maximizing the variance ratio, but $\Delta f$ is used as $f_2 = f_1 + \Delta f$ in Eq.(3). The left panel in figure 7 is a contour map of the variance ratio when $f_1$ is set to from 400 Hz up to 1kHz and $\Delta f$ from 10 Hz up to 1.2kHz. The maximum is obtained when $f_1$ and $\Delta f$ are 700 Hz and 720 Hz respectively. This result indicates that the intensity peak found in oral snores ranges from 700 Hz to 1420 Hz. Compared with Figures 2, it is quite natural to understand that this range can lead to the best value to extract the innate properties of oral snores.

From a review of the literature, we discuss what kind of phenomenon such intensity peak reflects. According to Agrawal and coworkers’ report\(^{(21)}\), the site of snoring is the soft tissues in the upper airway such as the soft palate, tongue base, epiglottis, and tonsils. But in many cases the main site is the soft palate and/or the tongue base\(^{(27)(35)}\). Agrawal\(^{(21)}\) also demonstrated that in the case of tongue base snores an intensity peak is found in a specific band over 1 kHz while the palatal snores consist of lower frequency components (less than 500 Hz). The similar results are also reported by other researchers\(^{(27)(28)}\).

In comparison with our results, oral snoring sounds do not have only lower frequency
components below 500 Hz, but include higher frequency components at around 1kHz as well. From this and from a review of the literature, oral snoring sounds are regarded as a mixture of the palatal snoring (lower frequency) and the tongue base snoring (higher frequency).

5.3. Optimal Value of $k$ and Classification Results

In this section, the number of neighbors ($k$) for the kNN classifier is adjusted under the four cross validation evaluation. Figure 8 shows the classification rate when $k$ is set to from 1 to 50 with 2 steps. The classification rate is largely converged to 0.92 with $k > 7$ under the 10-CV, LOO, and LEO evaluations. In the case of LSO test, the classification rate reaches 0.88 when $k$ is set to 7, but gradually increases up to 0.89 when $k=19$. 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 0.89 at least if the parameters are suitably adjusted. Figure 9 shows scatter plots in the feature space where oral and nasal subsequences are represented with blue and red points respectively. Oral snores are more widely scattered than nasal ones, but they are well separated from each other except a few outliers. In general, if the number of neighbors ($k$) is set to lower value, the classification boundary is so complex that it cannot correctly discriminate some test data around the classification boundary. From figure 9, it is obvious that the boundary becomes smoother if $k$ is set to 19 compared with $k = 7$. As a result, the better performance is obtained with $k = 19$ at least under every evaluation.

Table 2 shows the detail result of the classification with $k = 19$, where the rates are about 92% in total under the three evaluation tests except the LSO. Under the LSO test the rate is about 3% lower than that under the other tests. In general, if the number of data is large enough to achieve the objective high performance, the classification rates are almost the same.
under every evaluation test\(^{(32)}\). Thus, the number of our data is not so small that we cannot evaluate our method objectively, but, in consideration of the individual difference, it is not so large that we can achieve (completely) objective results. So, it is necessary to obtain more data from more subjects in the future.

5.4. Classification Results of Oronasal Simulated Snoring

Finally, we show and discuss the classification result if the \(k\)NN classifier receives oronasal snoring sounds as the input data. Oronasal snoring sounds are obtained from the same subjects in table 1, who are asked to simulate snoring by breathing both orally and nasally with open mouth and without closing the nostrils. Since oronasal snoring seems to be found more generally than (complete) oral or nasal snoring, it is necessary to discuss what results are obtained if such snoring sounds are given to our proposed method.

Fig.10 shows some examples of waveforms and the amplitude spectra of oronasal snores. The second panel from the top is similar to the nasal snores in fig.3, whereas the other panels are similar to the oral snores in fig.2. In addition, fig.11 shows scatter plots of oronasal snores on the 2-dimensional feature space. Compared with fig.9, the distribution largely overlaps both oral and nasal distribution. Thus, we can find that two acoustic properties of oronasal snores include both those of oral snores and those of nasal ones, and there are no peculiar properties of oronasal snores.
Fig. 9 Scatter plots of nasal (red) and oral (blue) snoring sounds and the classification boundary estimated by kNN classifier with $k = 7$ (left) and with $k = 19$ (right).

Table 2 Classification rate with the optimal parameters ($r = 3$, $f_1 = 700$, $f_2 = 1420$, $k = 19$) under the four different evaluation methods.

<table>
<thead>
<tr>
<th>Eval.</th>
<th>Nasal (%)</th>
<th>Oral (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-fold</td>
<td>1014/1107 (91.60%)</td>
<td>1289/1384 (93.14%)</td>
<td>2303/2491 (92.45%)</td>
</tr>
<tr>
<td>LOO</td>
<td>1015/1107 (91.69%)</td>
<td>1283/1384 (92.70%)</td>
<td>2298/2491 (92.25%)</td>
</tr>
<tr>
<td>LEO</td>
<td>1012/1107 (91.42%)</td>
<td>1281/1384 (92.56%)</td>
<td>2293/2491 (92.05%)</td>
</tr>
<tr>
<td>LSO</td>
<td>964/1107 (87.08%)</td>
<td>1268/1384 (91.62%)</td>
<td>2232/2491 (89.60%)</td>
</tr>
</tbody>
</table>

Fig. 10 Waveforms and amplitude spectra of oronasal simulated snores.

In contrast with oral and nasal snores, it is difficult for subjects to precisely control the airflow while breathing, namely, difficult to equally divide the airflow into oral and nasal cavity while oronasal breathing, because of nasal congestion and/or breathing habit which depend on the subject. Accordingly, in consideration of individual difference, we analyzed
the classification result of oronasal snores in every subject under the LSO test (shown in Table 3). Except for subject J, the majority of oronasal snores are classified into oral class, while a small number of oronasal ones into nasal class. The rate of oral class is about 81.2% in average.

Therefore, the acoustic properties of oronasal snores include mainly those of oral ones. Dalmasso, et al, described that the shape of cross sectional area in the upper airway are very similar during oral snoring and during free snoring (oronasal route)(15). This report supports our results. As a result, our method can detect open mouth during snoring whether the nostrils are closed or not, because oronasal snores are also produced by open mouth. According to the fact that open mouth is related to sleep apnea as is described in section 1, our method is also applicable to oronasal snores.

6. Conclusion

From the acoustic properties of snoring sounds, nasal snores consist of lower frequency components less than 500 Hz, whereas oral snores have both the lower frequency components and the intensity peak at around 1kHz. And it is found that the fundamental frequency of oral snores tends to be lower than that of nasal ones. In this paper, we focused on the difference of these acoustic properties to classify oral/nasal snoring sounds. We adopted the HPS method to find the fundamental frequency and calculated the maximum of the amplitude spectra in a specific band, which is greater than 700 Hz and less than 1420 Hz. As a result, oral and nasal
snores can be successfully classified with good accuracy, 89% at least, using two acoustic properties we focused on and the kNN classification method. Moreover, we clarified what kind of phenomena the acoustic properties we obtained reflect from the medical and physiological literature and demonstrated that the majority of oronasal snoring sounds are classified into oral class by the use of our method. This means that our method can detect open mouth during snoring, which is known to be related to OSA, whether the nostrils are closed or not.

There are still some problems to develop a home medical device which can detect oral snoring with only a microphone. In the future, we will deal with the following problems. (1) More data should be collected for objective evaluation and should be compared with natural snoring sounds. (2) It is necessary to compare kNN with the other classification methods to find the best method for classifying oral/nasal snoring. (3) All snore episodes were cut out manually from the recorded data in this paper, but this procedure must be done automatically.

Acknowledgement

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

References


