



Title	Analysis of Upper Airway MRIs and Snoring Sounds for Automatic Classification of Obstructive Sleep Apnea Syndrome
Author(s)	三上, 剛
Citation	北海道大学. 博士(情報科学) 乙第6931号
Issue Date	2014-09-25
DOI	10.14943/doctoral.r6931
Doc URL	<a href="http://hdl.handle.net/2115/57261">http://hdl.handle.net/2115/57261</a>
Type	theses (doctoral)
File Information	Tsuyoshi_Mikami.pdf



[Instructions for use](#)

A Thesis  
学位論文

# Analysis of Upper Airway MRIs and Snoring Sounds for Automatic Classification of Obstructive Sleep Apnea Syndrome

閉塞型睡眠時無呼吸症候群の自動識別を目的とした  
上気道 MR 画像といびき音の解析



submitted in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy  
in  
Information Science and Technology

博士  
(情報科学)

Tsuyoshi Mikami

三上 剛

Division of Computer Science and Information Technology,  
Graduate School of Information Science and Technology,  
Hokkaido University

北海道大学大学院 情報科学研究科  
情報理工学専攻

September 2014

# Contents

<b>Abstract</b>	<b>2</b>
<b>Abstract (in Japanese)</b>	<b>5</b>
<b>Acknowledgements (in Japanese)</b>	<b>8</b>
<b>Publications</b>	<b>10</b>
<b>1 Sleep Apnea Syndrome and Biomedical Signals</b>	<b>14</b>
1.1 Introduction . . . . .	14
1.2 Social Background . . . . .	14
1.3 Sleep Apnea Syndrome (SAS) . . . . .	16
1.3.1 Definition of SAS and AHI . . . . .	16
1.3.2 Obstructive Sleep Apnea Syndrome (OSAS) . . . . .	17
1.3.3 Central Sleep Apnea Syndrome (CSAS) . . . . .	19
1.3.4 Mixed Sleep Apnea Syndrome (MSAS) . . . . .	19
1.4 Screening and Diagnosis of SAS . . . . .	20
1.4.1 Simple Screening Test . . . . .	20
1.4.2 Polysomnography (PSG) . . . . .	21
1.4.3 Diagnosis . . . . .	23
1.5 Medical Images of Upper Airway Soft Tissues . . . . .	24
1.5.1 Image-based Diagnosis . . . . .	24
1.5.2 Cephalogram . . . . .	24

1.5.3	Upper Airway MR Images . . . . .	25
1.5.4	Related Studies . . . . .	26
1.6	Snoring Sounds . . . . .	29
1.6.1	Sound-based Diagnosis . . . . .	29
1.6.2	Relation of Oral Snoring to OSAS . . . . .	30
1.6.3	Simulated Snoring Sounds . . . . .	30
1.6.4	Similarities between Natural and Simulated Snoring . . . . .	31
1.6.5	Related Studies . . . . .	33
1.7	Summary . . . . .	36
	References . . . . .	37
<b>2</b>	<b>Morphological Analysis of the Upper Airway Soft Tissues in MR Images for Classification of Severe OSAS</b>	<b>44</b>
2.1	Introduction . . . . .	44
2.2	Subjects and Instrument . . . . .	46
2.3	Feature Extraction Method . . . . .	47
2.3.1	Tongue Region Segmentation . . . . .	47
2.3.2	Morphological Feature of the Tongue Region . . . . .	49
2.3.3	Sectional Area of the Narrowest Upper Airway . . . . .	50
2.4	Evaluation Method . . . . .	50
2.4.1	Correlation Coefficient with Apnea/Hypopnea Index . . . . .	50
2.4.2	Classification of the Severe OSAS using Multi-Regression Analysis . . . . .	52
2.4.3	Comparison with the Other Features in Conventional Studies	54
2.5	Results . . . . .	57
2.6	Discussion . . . . .	59
2.6.1	Correlation with the Severity . . . . .	59
2.6.2	Classification Ability of the Severe OSAS . . . . .	61
2.7	Conclusion . . . . .	63

References . . . . .	64
<b>3 Acoustic Analysis and Classification of Oral and Nasal Snoring Sounds</b>	<b>69</b>
3.1 Introduction . . . . .	69
3.2 Acquisition of Classification Targets . . . . .	71
3.2.1 Apparatus, Subjects, and Simulated Snoring . . . . .	71
3.2.2 Definition of Snoring and Extraction of Snoring Episodes . .	72
3.2.3 Subsequence Extraction . . . . .	76
3.3 Classification Method . . . . .	76
3.3.1 Feature Extraction . . . . .	76
3.3.2 Classification . . . . .	80
3.4 Performance Evaluation . . . . .	81
3.4.1 $m$ -fold Cross Validation Test . . . . .	81
3.4.2 Leave-One-Out Test . . . . .	82
3.4.3 Leave- <i>Episode</i> -Out Test . . . . .	82
3.4.4 Leave- <i>Subject</i> -Out Test . . . . .	82
3.5 Results and Discussion . . . . .	83
3.5.1 Optimal Value of $r$ and its Relation to the Acoustic Properties	83
3.5.2 Optimal Value of $f_1$ and $\Delta f$ and its Relation to the Acoustic Properties . . . . .	86
3.5.3 Optimal Value of $k$ and Classification Results . . . . .	86
3.5.4 Classification Results of Oronasal Simulated Snoring . . . .	89
3.6 Conclusion . . . . .	92
References . . . . .	93
<b>4 Spectral Classification of Oral and Nasal Snoring Sounds using a Support Vector Machine</b>	<b>99</b>
4.1 Introduction . . . . .	99
4.2 Data Acquisition and Preprocessing . . . . .	101

4.2.1	Data Acquisition . . . . .	101
4.2.2	Similarity between simulated and actual snoring sounds . . .	101
4.2.3	Episode Selection and Subsequence Extraction . . . . .	102
4.3	Feature Extraction Method . . . . .	106
4.4	Classification Method . . . . .	107
4.4.1	Support Vector Machine . . . . .	107
4.4.2	Kernel Functions . . . . .	108
4.4.3	A Criterion for Finding the Best Parameters . . . . .	108
4.4.4	Comparison with the Other Classifiers . . . . .	110
4.5	Performance Evaluation Methods . . . . .	112
4.5.1	10-fold Cross Validation Test . . . . .	112
4.5.2	Leave-Subject-Out Test . . . . .	112
4.6	Results and Discussion . . . . .	113
4.6.1	The Optimal Parameter of $\Delta f$ and the Best Choice of Kernels	113
4.6.2	Individual Differences and Reliability . . . . .	114
4.6.3	Comparison with Our Conventional Method . . . . .	115
4.6.4	Miss-classified Spectral Features . . . . .	118
4.6.5	Computational Time . . . . .	119
4.6.6	Comparison with Other Classifiers . . . . .	120
4.7	Conclusion . . . . .	124
	References . . . . .	125
<b>5</b>	<b>Conclusion</b>	<b>131</b>
5.1	Contributions of this Thesis . . . . .	131
5.2	Future Works . . . . .	132

# List of Figures

1.1	The Frequency of Traffic Accidents in Sleep Apnea Patients for Five Years ( $n = 3.7$ million, $p < 0.01$ )[1] . . . . .	15
1.2	Anatomic structure of the upper airway during OSAS (cited from Wikimedia Commons, <a href="http://commons.wikimedia.org/wiki/File:Obstruction_ventilation_apnee_sommeil.svg">http://commons.wikimedia.org/wiki/File:Obstruction_ventilation_apnee_sommeil.svg</a> [3]) . . . . .	16
1.3	Pediatric polysomnography (cited from Wikimedia Commons, <a href="http://en.wikipedia.org/wiki/Polysomnography">http://en.wikipedia.org/wiki/Polysomnography</a> mediaviewer/ File:Pediatric polysomnogram.jpg) . . . . .	21
1.4	Screenshot of a PSG of a person in REM sleep (cited from Wikimedia Common, <a href="http://commons.wikimedia.org/wiki/File:Sleep Stage REM.png">http://commons.wikimedia.org/wiki/File:Sleep Stage REM.png</a> [9]) . . . . .	22
1.5	An example of upper airway MR image in a male OSAS patient . . .	25
1.6	An example of craniofacial and soft-tissue parameters on cephalometric radiographs[19] . . . . .	26
1.7	The vibration of tissues during snoring with oral breath (left) and with nasal breath (right) . . . . .	29
1.8	Waveforms and amplitude spectra of oral simulated snoring sounds	32
1.9	Waveforms and amplitude spectra of nasal simulated snoring sounds	33
2.1	Sagittal and transverse planes of MR Images (cited from Wikimedia Commons, <a href="http://upload.wikimedia.org/wikipedia/commons/e/e1/Human_anatomy_planes.svg">http://upload.wikimedia.org/wikipedia/commons/e/e1/Human_anatomy_planes.svg</a> ) . . . . .	47
2.2	Anatomic chart of the upper airway and the tongue region . . . . .	48

2.3	Selection of the cross-sectional MR image through the narrowest airway . . . . .	49
2.4	Upper airway MRIs of the subjects from benign to severe OSA . . . . .	51
2.5	Directions whose correlation coefficients are greater than 0.5 . . . . .	52
2.6	Procedure to determine the neck circumference . . . . .	55
2.7	Procedure to determine the transverse planes for tongue volume estimation . . . . .	56
2.8	Correlation coefficients of $L_i$ ( $i = 1, 2, \dots, 48$ ) with AHI (Left) and relationship between $F_5$ and AHI (Right) . . . . .	59
2.9	Directions whose correlation is higher than 0.5 (region A and B) and the corresponding regions in MR images of moderate and severe OSAS patients. . . . .	62
2.10	Modified Mallampati Grade; Class I (soft palate, uvula, faces, pillars visible), Class II (soft palate, uvula, faces visible), Class III (soft palate, base of uvula visible), Class IV (only hard palate visible), (cited from Wikimedia Commons, <a href="http://commons.wikimedia.org/wiki/File:Mallampati.svg">http:// commons.wikimedia.org/wiki/ File:Mallampati.svg</a> ) . . . . .	62
3.1	A portable linear PCM recorder for recording snoring sounds (Olympus LS-11) . . . . .	73
3.2	An extraction method of episodes and subsequences of snoring sounds from a recorded sound. . . . .	74
3.3	An example of the harmonic product spectrum calculated with $r = 4$ . . . . .	75
3.4	Subsequences of oral snoring and their amplitude spectra. . . . .	78
3.5	Subsequences of nasal snoring and their amplitude spectra. . . . .	79
3.6	The variance ratio of the fundamental frequency calculated with the number of multiplication, $r$ , being set to 1 to 6 . . . . .	83
3.7	Two examples of incorrect estimation of the fundamental frequency when $r$ is determined to less than 3 (right) and greater than 3 (left) . . . . .	84



3.8	Contour plots of classification rate obtained from various $f_1$ and $\Delta f$ under the LSO test . . . . .	85
3.9	Classification rates with the number of neighbors, $k$ , set to 1 to 50 with 2 steps under the four cross-validation tests . . . . .	87
3.10	Scatter plots of nasal (red) and oral (blue) snoring sounds and the classification boundary estimated by kNN classifier with $k = 1, 5, 7, 13, 19,$ and $23$ . . . . .	88
3.11	Waveforms and amplitude spectra of oronasal simulated snores . . .	90
3.12	Scatter plots of oronasal simulated snores on the 2-dimensional feature space . . . . .	91
4.1	Simple chart of the snoring sites and the airflow while oral and nasal breath . . . . .	100
4.2	The waveforms and amplitude spectra of nasal snoring sounds . . .	103
4.3	The waveforms and amplitude spectra of oral snoring sounds . . . .	104
4.4	Discrete amplitude spectra and elements of the feature vector. . . .	105
4.5	Classification rate against various $\Delta f$ parameter estimated with SVM	115
4.6	The average of feature values on each dimension when $\Delta f$ is determined to 500. . . . .	116
4.7	The waveforms and amplitude spectra of non-vibrational sounds in oral (upper two) and nasal (lower two) snoring . . . . .	117
4.8	The waveforms and amplitude spectra of miss-classified oral (upper two) and nasal (lower two) snoring sounds . . . . .	119
4.9	Recognition rate when using $k$ -Nearest Neighbor method with various $k$ numbers . . . . .	121
4.10	Recognition rate when using MLP with various numbers of neurons in hidden layer . . . . .	121
4.11	Recognition rate when using LVQ with various numbers of codebook vectors . . . . .	122

4.12 Recognition rate of all classifiers under the LSO test . . . . . 123

# List of Tables

1.1	Correlation between AHI and Morphologic Properties in Conventional Studies . . . . .	28
2.1	Photographing conditions of the sagittal and transverse planes in the upper airway MRIs . . . . .	46
2.2	Contingency table made by comparing the classification results and the actual severity . . . . .	54
2.3	Overview of the oropharyngeal features having been focused on in conventional studies . . . . .	57
2.4	The mean and standard deviation of feature values in patients with $AHI < 30$ (from benign to moderate) and with $AHI \geq 30$ (serious) . . . . .	60
2.5	Sensitivity, specificity, and efficiency of multi-regression equations with the cut-off AHI determined to 30 . . . . .	60
2.6	Average, standard deviation, and correlation coefficient with AHI of the proposed and the conventional features both estimated from the upper airway MR images . . . . .	61
2.7	Correlation between AHI and Morphologic Properties in Conventional Studies . . . . .	63
3.1	Detail of all subjects and the obtained snoring episodes and subsequences in this study . . . . .	72
3.2	Classification rate with the optimal parameters ( $r = 3, f_1 = 700, f_2 = 1420, k = 19$ ) under the four different evaluation methods . . . . .	89

3.3	Classification results of oronasal simulated snores . . . . .	92
4.1	Kernel functions for the SVM used in this study . . . . .	109
4.2	The concrete values of kernel specific parameters for obtaining the best classification accuracy under the LSO test when $\Delta f$ is determined to 500. . . . .	110
4.3	Four classifiers adopted to discriminate oral and nasal snoring sounds for comparison . . . . .	111
4.4	Classification accuracies of SVMs with five different kernels respectively under the 10-fold CV and LSO test when $\Delta f$ is determined to 500. . . . .	114
4.5	Classification accuracies of SVMs with five different kernels respectively under the 10-fold CV and LSO test when $\Delta f$ is determined to 500. . . . .	114
4.6	Classification accuracy estimated with the conventional method in our previous work[15] . . . . .	118
4.7	Specific parameters of classification methods . . . . .	120

*Dedicated to my family, and  
to all the people who have supported the author*

# Abstract

Many serious accidents caused by drowsiness while driving or working with heavy machinery have been reported in recent years. In many cases, such abnormal drowsiness is caused by Sleep Apnea Syndrome (SAS), which, therefore, should be recognized to be not only an individual disease but also a social problem. The early detection and suitable treatment are urgently necessary, but SAS patients very often have subjective symptoms of it. Accordingly, the number of the latent SAS patients in Japan is estimated to be over 2 millions.

A polysomnography (PSG) is presently one of the most reliable and widely used measuring methods to assess sleep apnea at hospital. It is a screening method which directly measures the condition of sleep apnea and consists of about 10 kinds of contact sensors such as EEG, EOG, airflow via nose and mouth, ECG, SpO<sub>2</sub>, etc. The measured data of PSG is used for a medical technologist to estimate an Apnea/Hypopnea Index (AHI, the average number of apnea/hypopnea events per hour during sleep), which indicates the severity of sleep apnea. The AHI is an objective index for SAS diagnosis, but the PSG requires the patients to sleep in the hospital overnight. In addition, the severity is sometimes estimated to be lower than in reality because of the daily variance and the *first night effect*. Since the PSG is also time and energy consuming for measuring and analyzing data, a different approach to SAS classification has been expected recently.

By the way, if Obstructive SAS (OSAS), which is known to be over 80% of SAS, is only focused on, we can find two specific properties of OSAS in the upper airway soft tissues. One is the enlargement of the upper airway soft tissues which

---

directly causes the physical closure of the upper airway. The other is a vibration (snoring) which occurs in the narrowed upper airway which the inspiratory airflow rapidly passes through. Especially, loud snoring caused by open mouth during sleep is a typical symptom of OSAS. Since one of the two or both features are found in OSAS patients, it is necessary to analyze both those features for OSAS classification. According to this, this study focuses on the morphological features of the soft tissues in the upper airway MRIs and the acoustic properties of snoring sounds, and proposes some technical methods for automatic OSAS classification. This thesis is organized as follows.

In chapter 1, background and objectives of this study are described. The significance of classifying OSAS by using the upper airway MR images and the snoring sounds is described, together with the brief explanation, the present situation, and some drawbacks of standard screening methods for SAS. In addition, the related studies are introduced.

In chapter 2, the upper airway MR images of OSAS patients are analyzed. The tongue morphology in the sagittal plane is quantified and the correlation coefficient with AHI is calculated. As a result, the direction from the center of gravity in the tongue region to the back of the tongue and to the hyoid bone are mostly correlated with the severity of OSAS. This result is related to the medical criteria which have been adopted for visual examination of oropharynx and X-ray images. Next, the cross sectional area of the narrowest airway in the transverse plane is calculated and the linear equation for predicting AHI is solved with multi-regression analysis using the tongue morphology and the cross sectional area. The severe OSAS is classified by thresholding the prediction value. Consequently, it is clarified that the tongue morphology and the cross sectional area are useful for the purpose of severe OSAS classification.

In chapter 3, oral snoring which is generally found in OSAS patients is focused on and the difference between the acoustic properties of oral snoring sounds and those of nasal ones are clarified. In order to record snoring sounds with oral

---

and nasal breath, this study used a method called *simulated snoring*, which has been adopted in some physiological and medical studies. The simulated snores are produced by an awake person who is asked on purpose to simulate snoring. They are not completely equal to actual nocturnal snoring from acoustical point of view, but much similarity has been reported in conventional studies. According to the FFT amplitude spectra, the fundamental frequency of nasal snores tends to be higher than that of oral ones, and do not have many harmonic components as well. In oral snoring sounds, a *formant*-like intensity peak is found at around 1kHz, whereas in nasal ones no such peak is found. Based on these acoustic properties, the fundamental frequency and the maximum of the spectrum peak in a specific band are calculated and used as feature values for k-Nearest Neighbor classification. As a result, about 89% snoring sounds are correctly classified, but some snoring sounds are not periodic and do not have a fundamental frequency. It is clarified that these sounds are not correctly classified.

In chapter 4, the spectral shape up to 2kHz is adopted as the multi-dimensional feature vectors of the snoring sounds and a Support Vector Machine (SVM) is used for the classification. In order to find an appropriate choice of kernels, typical kernel functions such as linear, polynomial, sigmoid, Gaussian, Laplacian are considered. In addition, a chi-square kernel and Kullback-Leibler (KL) kernel, which are often used for spectrum similarity, are also adopted. As a result, a classification accuracy when using the KL kernel is the highest and about 5% is improved in comparison with the method in chapter 3. Moreover, the result is fairly good from the point of view of individual difference, reliability, and computational time.

Finally, chapter 5 describes the contribution of this thesis and the future works by summarizing all the results and discussions in chapter 2, 3, and 4.



## Abstract (in Japanese)

近年、自動車の運転や重機作業の従事者が作業中に傾眠状態に陥り重大な事故につながった事例が数多く報告されており、その多くは睡眠時無呼吸症候群 (Sleep Apnea Syndrome; 以下 SAS) が原因の睡眠不足にあることが指摘されている。そのため、SAS は個人疾患ではなく、社会問題として認識されるべきであり、早期発見と適切な治療が急務であるが、SAS は自覚することが難しいことから、潜在的な患者は国内で 200 万人に及ぶと推定されている。

現在、SAS の検査方法として最も信頼性が高く、広く病院で用いられているゴールド・スタンダードな手法は PSG (Polysomnography, 睡眠ポリグラフ) である。PSG は、脳波、眼球運動 (EOG)、オトガイ筋電、鼻と口の気流、心電、血中酸素飽和度 (SpO<sub>2</sub>) など 10 種類程度の接触型センサーを装着し、睡眠中に生じる無呼吸の病態を直接計測する検査方法である。PSG によって獲得された計測データを用いて、SAS の重症度を示す AHI (Apnea/Hypopnea Index, 1 時間あたりの無呼吸または低呼吸の回数) が計算され、治療のための客観的な指標として用いられる。しかし、PSG は検査入院を要することから患者の負担も大きく、日差変動もあることから実際より軽症と診断される場合もある。また、PSG は大掛かりな検査であり、取得したデータの分析にも手間がかかるため、PSG とは異なるアプローチから出来るだけ簡便に SAS を識別できる検査手法に近年注目が集まっている。

一方、SAS の約 8 割以上を占める閉塞型 SAS (Obstructive SAS; 以下、OSAS) に限定した場合、上気道軟部組織に関して OSAS 特有の特徴が見られる。1 つは物理的に気道の閉塞が生じる上気道軟部組織の肥大であり、これは MR 画像により確認することができる。もう 1 つは、狭小化した気道に吸気が急速に流入することによって生じる上気道軟部組織の振動 (いびき) であり、特に開口を伴う大音

---

量のいびきは OSAS 固有の特徴として知られている。OSAS はこれらの特徴が単独または複合的に発現するため多面的な解析が必要と考える。そのため、本研究では上気道 MR 画像における軟部組織の形態的特徴といびきの音響的特徴に着目し、OSAS を識別するための情報科学的手法を提案する。論文は以下の構成になっている。

第1章は序論であり、本研究の背景と目的を述べている。SAS の概要および SAS の検査方法の現状と問題点について取り上げ、上気道 MR 画像といびき音に着目して OSAS を識別する必要性について述べている。また、関連研究について概説し、本論文の位置づけを明らかにしている。

第2章では、OSAS 患者の上気道 MR 画像について着目した。矢状面図における舌領域の形態を定量化し重症度 AHI との相関係数を求めた。その結果、舌の形態的な特徴としては、舌領域の中心から舌背方向と舌骨方向の長さが AHI と最も相関が高いことが明らかになった。これは、従来から診断で用いられている定性的な口咽頭視診の指標と、X 線画像を用いた場合（セファログラム）の評価指標にそれぞれ対応することが分かった。さらに、横断面図における気道の最狭部断面積を求め、舌部の特徴量にこの値を加えて重回帰分析を行い重症度 AHI の予測式を求めた。予測値のしきい値処理による重度 OSAS の識別能力について検証し、上気道 MR 画像の矢状面図における舌領域の形態的特徴と横断面図における最狭気道断面積を用いることの有用性について明らかにした。

第3章では、OSAS 患者に多く見られる口呼吸に伴ういびきに着目し、呼吸様式（口呼吸、鼻呼吸）に伴う音響特性の相違について検証した。口呼吸のみの場合と鼻呼吸のみの場合のいびき音を録音するため、医学生理学研究で従来から採用されている模擬いびき（Simulated Snoring, いびきのかきまね）の方法を用いた。尚、実際のいびきと模擬いびきの等価性については従来研究でも検証されており、完全に等価とは必ずしも言えないが多くの類似性が指摘されている。FFT による振幅スペクトルを用いて検証したところ、鼻呼吸いびきは口呼吸いびきに比べ基本周波数が高く、倍音成分が少ない比較的単純な波形であることが確認された。一方で、口呼吸いびきにはフォルマント様の密度ピークが 1 kHz 付近にみられたが、鼻呼吸いびきの場合には見られなかった。以上の音響特性に着目し、基本周波数と

---

1kHz 前後の帯域におけるスペクトルの最大値を特徴量として求め、k 最近傍法を用いて口呼吸いびきと鼻呼吸いびきの識別を行った。その結果、90 % 近くの識別率を得たが、いびき音の中には基本周波数のピークが存在しない非周期波形も多く、そのようないびき音は正しく識別されないことも判明した。

第4章では、第3章で明らかになった非周期波形のいびき音の識別も考慮し、振幅スペクトル全体の概形を多次元の特徴ベクトルとして定義した。また、識別器としてサポートベクタマシン (SVM) を用いることにした。SVM のカーネル関数として、パターン識別で標準的に用いられるもの (線形, 多項, Sigmoid, Gauss, Laplace) に加え、スペクトルの類似度として近年その有効性が指摘されているカイ二乗カーネルと Kullback-Leibler (KL) カーネルも導入した。その結果、KL カーネルを用いた場合が最も識別率が高く、第3章で提案した手法と比較して識別率が5%近く向上した。また、SVM を用いた結果に関して、データの個人差、結果の信頼性、計算時間について考察し、いずれの点においても良好な結果を得ることが出来た。

第5章は結論である。本研究で得られた結果についてまとめ、今後の課題について述べている。

## Acknowledgements

この論文は2014年までの6年間に行われた研究成果によって構成されています。この間、3編のフルペーパーと1編のショートペーパーが学術雑誌および論文誌へ掲載され、国際会議および国内学会の講演会やシンポジウムにおいて延べ18回の発表を行ってきました。これらの実績を残すに当たり大変多くの方々のご支援ご鞭撻を頂きました。ここに深く感謝致しますと共に、今後も更なる研鑽を積み博士の名に恥じない研究成果を挙げる所存です。

山本雅人先生（北海道大学 大学院情報科学研究科情報理工学専攻 自律系工学研究室 教授）には、本論文の主査として、本学位論文ならびにその基盤となった学術ジャーナルに掲載された3編のフルペーパーについてきめ細かいご指導を頂きました。

栗原正仁先生（北海道大学 大学院情報科学研究科情報理工学専攻 知能ソフトウェア研究室 教授）、鈴木恵二先生（同 調和系工学研究室 教授）、小野哲雄先生（同 ヒューマンコンピュータインタラクション研究室 教授）、飯塚博幸先生（同 自律系工学研究室 准教授）には、副査として本論文を纏めるに当たって多くのご助言を頂きました。

古川正志先生（北海道大学 名誉教授、北海道情報大学 教授）には、本論文を構成する3編のフルペーパーに関してご指導を頂きました。また、筆者が旭川高専の学生だった頃から、学問や研究の基本的な事柄に関して大変多くのことを学びました。

米澤一也先生（国立病院機構函館病院 臨床研究部長）には、SAS 専門医のお立場から本研究に関して医学的見地からご助言を頂きました。また、SAS 患者様のいびき音の録音、上気道MRI データのご提供、同病院における研究倫理委員会に

---

おける承認手続き，医学系の各種フォーラムへの参加，MRI データの分析結果に関するディスカッションなど，数多くのご支援とご指導を頂きました。

吉村 斎先生（苫小牧工業高等専門学校 情報工学科 教授）には，古川先生を始めとする諸先生方をご紹介くださりました。また，筆者が苫小牧高専に赴任した当初から長年にわたり大変多くのご支援を頂きました。

今哲二先生（今内科消化器科クリニック 院長）には，SAS の専門医である米澤先生をご紹介して頂きました。また，志賀捷浩先生（志賀医院 院長）と筆者の弟夫婦である三上健氏，三上満江氏には，今先生から米澤先生までに繋がるご縁を頂きました。

小島洋一郎先生（苫小牧工業高等専門学校 理系総合学科 教授）には，これまで何度も筆者と共同研究の機会を頂きました。先生から多くのご助言を頂いただけでなく，高専教員としての研究テーマの着眼点など大変多くのことを学び，本研究テーマを着想する原点となりました。

苫小牧工業高等専門学校 情報工学科の諸先生方には，10 年以上もの長きに渡りなかなか実績を積むことの出来なかった筆者に対し温かいご支援を頂きました。

最後に，これまで筆者を支えてくれた妻と子供達，両親，義母と義姉，ならびに各講演会やシンポジウムにおいて有益なご助言を頂いた諸先生方や学生さん達に感謝申し上げます。

# Publications

## Journal Papers

1. Tsuyoshi Mikami, Kazuya Yonezawa, Yohichiro Kojima, Masahito Yamamoto, Masashi Furukawa, "Usefulness of Image Analysis on Upper Airway MRI for the Classification of Severe Obstructive Sleep Apnea Syndrome (OSAS)", *Medical Imaging and Information Sciences*, Vol.31, No.1, pp.13-18, 2014 (in Japanese)
2. Tsuyoshi Mikami, Yohichiro Kojima, Kazuya Yonezawa, Masahito Yamamoto, Masashi Furukawa, "Spectral Classification of Oral and Nasal Snoring Sounds with Support Vector Machines", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.17, No.4, pp.611–621, Fuji Tech. Pub., 2013
3. Tsuyoshi Mikami, Yohichiro Kojima, Kazuya Yonezawa, Masahito Yamamoto, Masashi Furukawa, "Classification of Oral/Nasal Simulated Snores based on the Acoustic Properties", *Journal of Biomechanical Science and Engineering*, Vol.7, No.4, pp.433–448, JSME, 2012
4. Tsuyoshi Mikami, Yohichiro Kojima, "Automatic Classification of Breathing Route during Sleep using Snoring Sounds", *IEEJ Transactions on Electronics, Information, and Systems*, Vol.131, No.8, pp.1516–1517, 2011

---

## International Conference Papers (Reviewed)

1. Tsuyoshi Mikami, Yohichiro Kojima, Kazuya Yonezawa, Masahito Yamamoto, "Image-based Classification of Severe Obstructive Sleep Apnea using the Mophology of Upper Airway Soft Tissues", *Proceedings of 2014 IEEE International Symposium on Biomedical Imaging (ISBI 2014)*, Beijing, China, April 29 - May 3, 2014
2. Tsuyoshi Mikami, Yohichiro Kojima, Kazuya Yonezawa, Masahito Yamamoto, Masashi Furukawa, "Mophology Analysis of the Tongue Region in the Upper Airway MRI for Sleep Apnea Diagnosis", *Proceedings of 2013 SICE Annual Conference*, Nagoya, September 14-17, 2013
3. Tsuyoshi Mikami, Yohichiro Kojima, Kazuya Yonezawa, Masahito Yamamoto, Masashi Furukawa, "Correlation of Morphological Properties with the Severity of Obstructive Sleep Apnea", *Proceedings of 35th Annual Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2013)*, Osaka, July 7-9, p.3234, 2013
4. Tsuyoshi Mikami, Yohichiro Kojima, Masahito Yamamoto, Masashi Furukawa, "Automatic Classification of Oral/Nasal Snoring Sounds based on the Acoustic Properties", *Proceedings of the 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*, Kyoto, March 25-30, pp.609-612, 2012
5. Tsuyoshi Mikami, Yohichiro Kojima, Masahito Yamamoto, Masashi Furukawa, "An SVM-based Classification of Oral and Nasal Snoring Sounds with Kullback-Leibler Kernel", *Proceedings of SICE Annual Conference*, Akita, August 20-23, pp.1795-1797, 2012
6. Tsuyoshi Mikami, Yohichiro Kojima, Masahito Yamamoto, Masashi Furukawa, "Neural Classification of Snoring Sounds for the Detection of Oral Breathing

---

during Snoring”, *Proceedings of the 5th IASTED International Conference on Biomedical Engineering*, Innsbruck, February 16-18, pp.723-059, 2011

7. Tsuyoshi Mikami, Yohichiro Kojima, Masahito Yamamoto, Masashi Furukawa, ”Recognition of Breathing Route during Snoring for Simple Monitoring of Sleep Apnea”, *Proceedings of SICE Annual Conference*, Taipei, August 18-21, pp.3433-3434, 2010
8. Tsuyoshi Mikami, ”Detecting Nonlinearity in Prediction Residuals of Snoring Sounds”, *Proceedings of ICROS-SICE International Joint Conference*, Fukuoka, August 18-21, pp.5256-5259, 2009
9. Tsuyoshi Mikami, ”Detecting Nonlinear Properties of Snoring Sounds for Sleep Apnea Diagnosis”, *Proceedings of the 2nd IEEE International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, Shanghai, May 16-18, pp.1173-1176, 2008

### **Presentation in Domestic Conferences or Symposiums**

1. Kazuya Yonezawa, Tsuyoshi Mikami, Tongue Morphology Analysis of the Upper Airway MRI for Sleep Apnea Diagnosis, The 67th Annual Meeting of the Japanese Society of National Medical Services, Kanazawa, 2013
2. Tsuyoshi Mikami, Kazuya Yonezawa, Acoustic Properties of Oral and Nasal Snoring Sounds, The 67th Annual Meeting of the Japanese Society of National Medical Services, Kanazawa, 2013
3. Kazuya Yonezawa, Tsuyoshi Mikami, Condition of the Pharyngeal Soft Tissues during Nasal CPAP, The 67th Annual Meeting of the Japanese Society of National Medical Services, Kanazawa, 2013
4. Yusaku Shimono, Tsuyoshi Mikami, Tongue Region Segmentation in the Upper Airway MR Images, The 2013 Hokkaido Division Conference of the Japan Society of Precision Engineering, Kitami, August, 2013



- 
5. Tsuyoshi Mikami, Kazuya Yonezawa, Yoichiro Kojima, Masahito Yamamoto, Masashi Furukawa, Correlation of Tongue Morphology in the Upper Airway MRI with the Severity of Obstructive Sleep Apnea, The 12th Microsymposium on Complex Systems, Sapporo, March, 2013
  6. Tsuyoshi Mikami, Studies on the Conditions of Sleep Apnea Syndrome: Collaboration of Medicine and Engineering, The 5th Hakodate Sleep Disordered Breathing Forum, November, 2012
  7. Tsuyoshi Mikami, Kazuya Yonezawa, Yoichiro Kojima, Masahito Yamamoto, Masashi Furukawa, Acoustic Properties of Snoring Sounds in Sleep Apnea Patients, The 87th Conference of Japanese Society of Medical Instruments, Sapporo, June, 2012
  8. Tsuyoshi Mikami, Kazuya Yonezawa, Yoichiro Kojima, Masahito Yamamoto, Masashi Furukawa, An SVM-based Classification of Oral and Nasal Snoring Sounds with Kullback-Leibler Kernel, The 11th Microsymposium on Complex Systems, Sapporo, March, 2012
  9. Tsuyoshi Mikami, Yoichiro Kojima, Masahito Yamamoto, Masashi Furukawa, Is there chaos in snoring sounds?, The JSME Robotics and Mechatronics Symposium (ROBOMECH), Asahikawa, June, 2011

## **Awards**

1. The Best Poster Award, The 67th Annual Meeting of the Japanese Society of National Medical Services, 2013

# Chapter 1

## Sleep Apnea Syndrome and Biomedical Signals

### 1.1 Introduction

This chapter gives the background and objective for this thesis, including some basic knowledge of Obstructive Sleep Apnea Syndrome (OSAS). Firstly, the definition of Sleep Apnea Syndrome (SAS) and Apnea/Hypopnea Index (AHI), the most commonly used severity scale, are explained in the former sections. Also, the current state of the sleep apnea screening is introduced and some problems are clarified from a medical and scientific point of view.

The later sections describe two biomedical signals related to OSAS, upper airway MR images and snoring sounds, and state the reason why these biomedical data are necessary for SAS classification. In addition, conventional studies related to image- and sound-based classification of SAS are surveyed in the later sections.

### 1.2 Social Background

*Sleep Apnea Syndrome* (SAS) has been focused on recently, as some of the serious traffic accidents are caused by the daytime heavy sleepiness of the drivers. Such sleepiness is one of the most serious problems in SAS, because SAS patients do

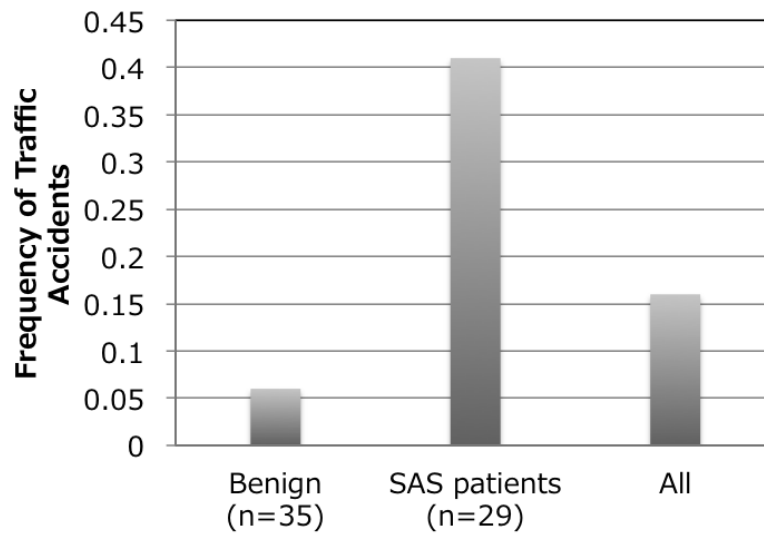


Figure 1.1: The Frequency of Traffic Accidents in Sleep Apnea Patients for Five Years ( $n = 3.7$  million,  $p < 0.01$ )[1]

not have a good sleep due to an unconscious awakening caused by apneas during sleep. In Western countries, this was known well long ago. According to an article published in *Wake up America* in 1990, the Three Miles Island nuclear accident and the Space Shuttle Challenger accident are mentioned as the ones caused by sleep disordered breathing including SAS.

Moreover, it has been reported that the frequency of traffic accidents caused by moderate or severe sleep apnea is seven times higher than that caused by normal drivers (figure 1.1)[1]. However, most people tend not to be conscious of their apneas or even their abnormal sleepiness. Toga [2] examined the drivers being aware of heavy sleepiness and whether they are taken sleep apnea or not. According to Findley's study[1], there should be many drivers being aware of heavy sleepiness, but actually the number of the drivers is quite low. This indicates that only the people who know SAS well are conscious of their abnormal sleepiness and many people do not have much information about SAS. It is, therefore, quite important to make the illness better known, and is desirable for apnea patients to perceive their own symptoms as soon as possible.

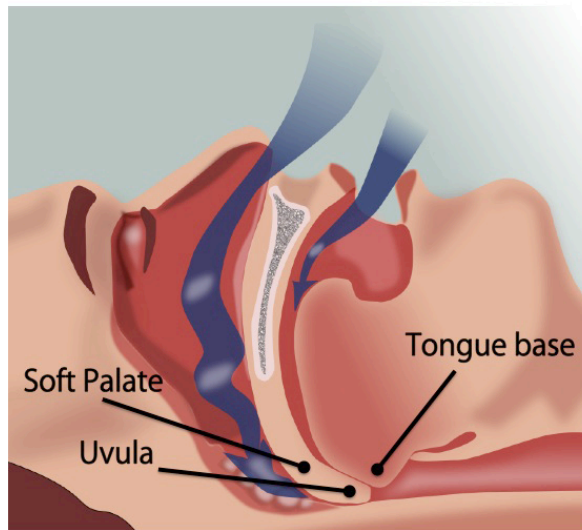


Figure 1.2: Anatomic structure of the upper airway during OSAS (cited from Wikimedia Commons, [http://commons.wikimedia.org/wiki/File:Obstruction\\_ventilation\\_apnee\\_sommeil.svg](http://commons.wikimedia.org/wiki/File:Obstruction_ventilation_apnee_sommeil.svg) [3])

Loud snoring, a typical symptom of SAS, often becomes an annoying problem for the family of SAS patients at home. Non of their families can even sleep in the next room, to say nothing of the same bedroom. Also, SAS patients often doze off in the office or the public facilities and snore loudly as well. Now, SAS is not necessarily a personal problem but a social one, which annoys the other people and often causes a serious accident. Therefore, it is quite important that you do know your own sleep condition or perceive the occurrence of your sleep apneas.

## 1.3 Sleep Apnea Syndrome (SAS)

### 1.3.1 Definition of SAS and AHI

Guilleminault, et al, [4] firstly advocated the concept of SAS in 1970s. An *apnea* is defined as the cessation of airflow through the nose and mouth for more than 10 seconds, and the SAS is defined as over 30 apnea events in 7 hours during sleep, which are also found in non-REM sleep. In the case of severe patients, more than

1 minute cessation are often found. During apnea, the oxygen in the patient's body gradually decreases and the carbon dioxide increases on the other hand. So, respiratory effort gradually increases.

After that, a *hypopnea*, more than 50% reduction in airflow, has also been considered important. Today, *more than 5 apnea or 10 apnea/hypopnea events per hour* are the criteria of SAS that have widely been adopted worldwide.

Thus, the number of apnea/hypopnea events per hour is used as an index of sleep apnea severity called *Apnea-Hypopnea Index* (AHI). The American Academy of Sleep Medicine (AASM) recommends grading sleep apnea as follows:

- Benign (AHI < 5)
- Mild ( $5 \leq \text{AHI} < 15$ )
- Moderate ( $15 \leq \text{AHI} < 30$ )
- Severe (AHI  $\geq 30$ )

The AHI can precisely be estimated from polysomnography (PSG) and roughly estimated from simple test. In the case of AHI>20 estimated from PSG or AHI> 40 from simple test, the nasal CPAP, the most representative ventilation therapy for OSAS, is covered by insurance in Japan.

From a physiological point of view, there are three types of sleep apnea; Obstructive Sleep Apnea Syndrome(OSAS), Central Sleep Apnea Syndrome (CSAS), and Mixed Sleep Apnea Syndrome (MSAS). But in many patients it is rare that only one of the three types is found in the PSG results. In OSAS patients, for instance, more obstructive apneas are recognized but a few central/mixed apneas are also found during one night.

### 1.3.2 Obstructive Sleep Apnea Syndrome (OSAS)

Over 80 percent of SAS patients in Japan are categorized to Obstructive sleep apnea syndrome (OSAS), the most common form of apnea[5]. It occurs when complete or

partial airway closure is emerged by the relaxation of muscles in our throat during sleep. There is a tendency for those who have a narrowed throat caused by their obesity or have a small jaw to be suffered from the disease. Snoring is a typical symptom of OSAS. Since many Japanese have a small jaw, OSAS easily occurs in non-obese people. The target of this study is limited to the OSAS.

The most common obstructive sleep apnea symptoms include:

- Daytime sleepiness or fatigue
- Dry mouth or sore throat upon awakening
- Headaches in the morning
- Trouble concentrating, forgetfulness, depression, or irritability
- Night sweats
- Restlessness during sleep
- Sexual dysfunction
- Snoring
- Sudden awakenings with a sensation of gasping or choking
- Difficulty getting up in the mornings

The repeated episodes of apneas lead to many times of unconscious awakening and prevent from getting good sleep. This causes an Excessive Daytime Sleepiness (EDS), which may interfere with your social life. It is known that the EDS leads to a traffic accident.

The medical treatment of OSAS is mainly a loss in weight if the patients are overweight, but it is not rare that the patients are not completely recovered. The apneas in slight patients are sometimes reduced when the patients lie on their side during sleep. But for the severe patients and the moderate patients with hypertension, cardiovascular disease, or diabetes, nCPAP (nasal Continuous Positive

Airway Pressure) is drastically effective, which requires patients to wear a nasal mask through which the airflow in their throat is preserved during sleep with the mechanical air pressure.

### 1.3.3 Central Sleep Apnea Syndrome (CSAS)

In contrast to OSAS, the upper airway is not obstructed in CSAS patients but the apnea is caused by a functional abnormality of the respiratory center. Namely, the brain does not order your muscles to breathe due to the abnormality. CSAS may occur in general with OSAS, or alone. It is known that cardiac insufficiency is complicated with CSAS.

The main symptom of central sleep apnea is temporary blockages of breathing during sleep. Although snoring is a very strong symptom of obstructive sleep apnea, snoring is usually not found with central sleep apnea. Symptoms may also include:

- being very tired during the day
- waking up often during the night
- going to the bathroom often during the night
- having headaches in the early morning
- poor memory and difficulty concentrating
- mood problems

### 1.3.4 Mixed Sleep Apnea Syndrome (MSAS)

Mixed Sleep Apnea Syndrome (MSAS) is a condition characterized by symptoms of both CSAS and OSAS. It often begins as central apnea and develops into the obstructive type.

## 1.4 Screening and Diagnosis of SAS

### 1.4.1 Simple Screening Test

Firstly, a simple screening test, which can be used at home, is introduced in this section. Basically, SAS screening is to measure the following three states of the patients:

- breathing (airflow through the nose or mouth)
- respiratory effort (movement of the chest or abdomen)
- sleep condition (sleeping hours and quality)

Simple screening instruments on the market are generally composed of a nasal/oral thermistor, a chest sensor, an abdomen sensor, and a pulse oximeter. The chest and abdomen sensors measure the physical movement of the chest and abdomen while breathing, which indicates the respiratory effort. The thermistor placed under the nose measures the airflow while breathing. In this method, brain signal indicating the sleep condition is not measured.

The pulse oximeter measures the oxygen saturation level in the blood. By using the thermistors and chest/abdomen sensors, apnea events can be detected correctly, but hypopnea events cannot. Thus, the screening results are not so reliable in the case of mild or moderate SAS patients who have more hypopneas than apneas. Likewise the pulse oximetry is not so reliable for mild or moderate sleep apneas, because the blood oxygen saturation level is not so lessened. Since the EEG signals are not recorded by simple screening, the sleeping hours are based on the self-declaration of the patient.

By using the simple screening method, it is, therefore, easy to measure the patient's apneas at home, but mild/moderate sleep apneas are overlooked.





Figure 1.3: Pediatric polysomnography (cited from Wikimedia Commons, [http://en.wikipedia.org/wiki/Polysomnography\\_mediaviewer/File:Pediatric\\_polysomnogram.jpg](http://en.wikipedia.org/wiki/Polysomnography_mediaviewer/File:Pediatric_polysomnogram.jpg))

#### 1.4.2 Polysomnography (PSG)

The PSG is a gold standard for screening SAS and has been used in many hospitals dealing with SAS. The PSG consists of many different sensors as follows:

- EEG (Electroencephalogram; brain)
- EOG (Electrooculography; eye movements)
- EMG (Electromyogram; muscle activity or skeletal muscle activation)
- ECG (Electrocardiogram; heart rhythm)
- pulse oximetry (blood oxygen saturation)
- thermistor (nasal and oral airflow)
- sound probe (snoring sounds)
- chest and abdomen sensors (respiratory effort)

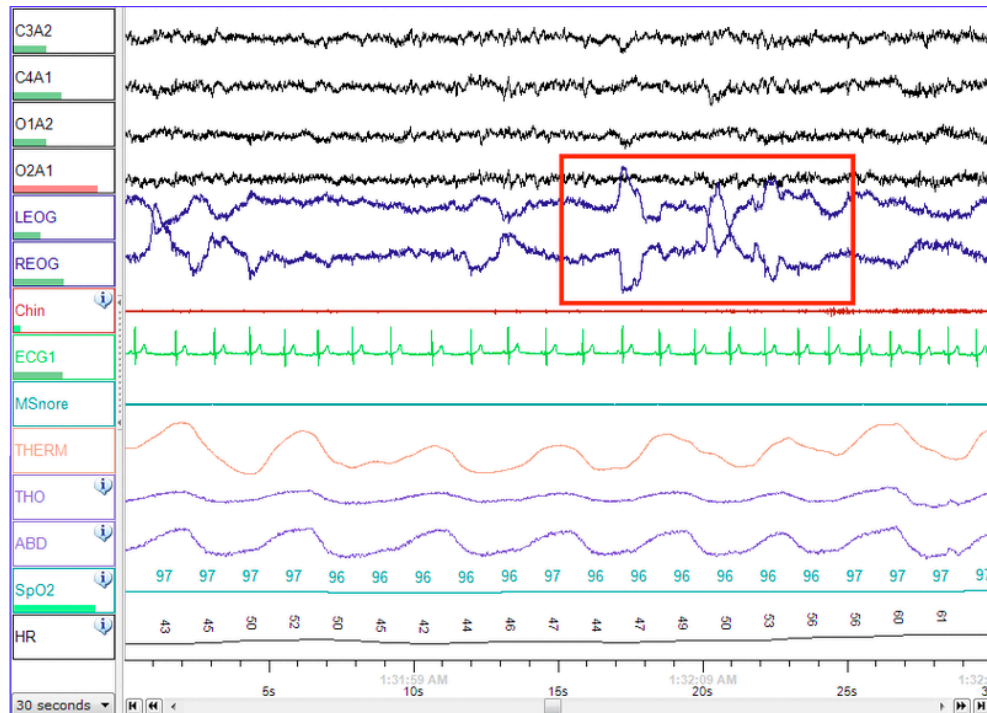


Figure 1.4: Screenshot of a PSG of a person in REM sleep (cited from Wikimedia Common, [http://commons.wikimedia.org/wiki/File:Sleep\\_Stage\\_REM.png](http://commons.wikimedia.org/wiki/File:Sleep_Stage_REM.png)[9])

- body position sensor (body position)

Some of the sensors are the same as the ones adopted in a simple screening, but sleep condition can be evaluated by the EEG, EOG, and EMG placed on the geniohyoid muscle. In addition, it is possible to evaluate the relation between the deepness of the sleep and the apnea/hypopnea events.

The PSG is a general method, but some drawbacks have also been reported as follows[6][7][8].

#### 1. Time Consuming Tasks

In order to estimate AHI, sleeping hours must be estimated through EEG. So, this method requires patients to sleep at hospital during screening and therefore the large volumes of data are obtained. Every instrument has a program that automatically estimates the deepness of the sleep and detects

apnea/hypopnea events, but the program is not reliable enough. So, it is necessary for medical technologists to manually confirm and modify the result of the program.

2. Many Contact Sensors

Many contact sensors must be attached on patients' body and so the patients often cannot sleep as deeply as usual. This does not lead to the deep muscle relaxation usually found at home and may cause the underestimation of the severity.

3. Daily Variance and First Night Effect

In sleep studies, a famous phenomenon has been reported by Arens, et al, [10]. This so-called *First Night Effect* is that the first night of laboratory sleep contains more awake periods and less Stage I-REM sleep. This effect rapidly adapts out by the second night of sleep, but at hospital several times PSG screening is rather unfavorable. This phenomenon may also cause the underestimation of the severity.

Based on these drawbacks, other methods that can estimate the severity more simply have recently been focused on.

### 1.4.3 Diagnosis

On account of the drawback called the First Night Effect, the medical specialist often diagnoses SAS by using not only the AHI but the other criteria such as the cephalometry measured from X-ray images, Modified Mallampati Grade estimated by visual examination of the oropharynx. Especially, the enlargement of the upper airway soft tissues are clearly identified on the MR images. The experienced medical specialists can qualitatively estimate the severity of the OSAS patients by the visual examination from the MR images of the upper airway. But the sufficient experience is necessary for the estimation.

On the other hand, snoring is a typical symptom of OSAS. Medical specialists have not used these sounds for the diagnosis of OSAS so far, but it has been known that the snoring sounds of OSAS patients have some peculiar properties. From the middle of 1990s, some researchers have analyzed the snoring sounds of OSAS patients to clarify the relation to AHI, but it was and is difficult to realize such estimation from the acoustic properties of snoring sounds.

## 1.5 Medical Images of Upper Airway Soft Tissues

### 1.5.1 Image-based Diagnosis

Visual examination using the X-ray images or MRIs has naturally been adopted so far, because in OSAS patients the enlargement of the upper airway soft tissues is in common identified. Figure 1.5 shows an example of X-ray image and MRI of the upper airway. But those medical images have their own advantages and disadvantages for the diagnosis of OSAS, which are described in the following sections.

### 1.5.2 Cephalogram

Cephalogram is a lateral X-ray image of the head. Prior to photographing, the head is immobilized by the ear-rods, and is positioned so that the distance from the head to the X-ray tube and to the film are determined to the constant. The radiation always passes through the same position of the head. The usefulness of cephalogram is to know the position of the hard tissues such as hyoid bone and the upper/lower jawbone.

In this method, measurement points are firstly determined on the cephalogram, and the distance and the angle among those points are measured. For the upper airway X-ray image of OSAS patients, the position of the upper/lower jawbone and the tongue are focused on so as to identify the narrowness of the airway. For instance, the distance from the hyoid bone to the mandibular plane (MP-



Figure 1.5: An example of upper airway MR image in a male OSAS patient

H)(See figure 1.6) indicates the elongated tongue morphology often found in OSAS patients.

According to conventional cephalometric studies[11][12][13][14][15][16][17][18], the lower jaw of SAS patients tends to be small and positioned backward. Especially, this feature is often found in Japanese people. But since the soft tissues are not clearly photographed in X-ray images, the determination of measurement points requires much experience of image-based diagnosis.

### 1.5.3 Upper Airway MR Images

On the other hand, soft tissues are clearly recognized in MR images in comparison with X-ray images. In addition, three-dimensional anatomical features can be evaluated. Thus, many researchers have focused on some 3-dimensional features of the pharyngeal/laryngeal organization to clarify the relation to the severity of OSAS.

However, those conventional studies have only analyzed quite simple features:

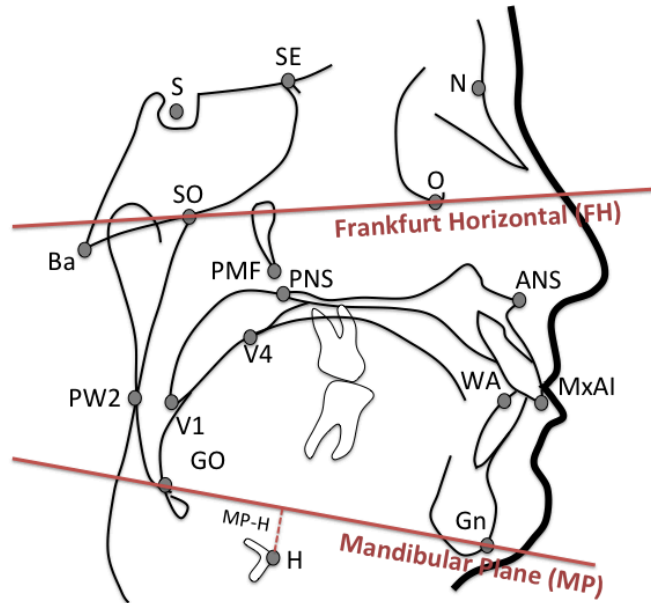


Figure 1.6: An example of craniofacial and soft-tissue parameters on cephalometric radiographs[19]

the volume of the tongue and the capacity of the upper airway[20][21], the length around the neck[22][23], fat volume[22] and etc[10]. It is quite easy to measure those features from the upper airway MR images, but the acquisition of more detailed features can be expected by the use of some image processing techniques.

#### 1.5.4 Related Studies

In this section, conventional studies of image-based analysis are introduced. The author gives notice here that the explanation of the conventional studies in this section are cited respectively from the original papers.

Kitamura, et al[11], studied the usefulness of cephalometry and pharyngeal findings in determining efficient primary diagnosis of OSAS. Persons with  $AHI \geq 15$  in PSG were considered to indicate OSAS patients. They analyzed the correlation between AHI and other parameters and conducted stepwise multiple regression analysis to predict AHI, and studied the screening performance of prediction equations using a receiver operating characteristic (ROC) curve. Of the 8 cephalometric

parameters examined, the length of the soft palate (PNS-P;  $p = 0.011$ ) and the distance from the mandibular plane to the hyoid bone (MP-H;  $p < 0.001$ ) correlated significantly with AHI. Two indices of the pharyngeal finding and body mass index (BMI) also significantly correlated with the AHI (MMP;  $p < 0.001$ , tonsil size ;  $p = 0.005$ , BMI ;  $p < 0.001$ ).

Yu, et al[13], evaluated the cephalometric features of OSAS patients and to elucidate the relationship between cephalometric variables and severity of the apnea-hypopnea index (AHI). According to their results, patients with OSAS in both subgroups showed several significant cephalometric features compared with simple snorers: (1) inferiorly positioned hyoid bone, (2) enlarged soft palate, and (3) reduced upper airway width at soft palate. Stepwise regression analysis showed that anterior displacement of the hyoid bone and retroposition of the mandible were the dominant overall determinants for AHI in patients with OSAS, and that narrowing of the bony oropharynx and inferior displacement of the hyoid bone were dominant determinants for AHI in nonobese patients.

Iida-Kondo[24], calculated the tongue volume/oral cavity volume ration (TV/OCV ratio) in the oral cavity using MRI for both OSAS patients and normal controls. The subjects comprised 20 male patients with OSAS ( $AHI \geq 5.0$ , with a diagnosis of OSAS) and 20 normal male adults ( $AHI < 5.0$ , with no history of OSAS) as the controls. They performed MRI to acquire T1- and T2-weighted images. They estimated tongue volumes on the basis of the cross-sectional area of each image, then using the tongue volume data, they calculated TV/OCV ratios. In the normal control group, mean ( $\pm$ SD) body mass index (BMI) was  $21.68 \pm 1.73$  and the mean TV/OCV ratio was  $86.98 \pm 3.16\%$ , whereas these values were  $25.0 \pm 15.94$  and  $90.56 \pm 2.14\%$ , respectively, in the OSAS patient group. The TV/OCV ratio of the OSAS patient group was significantly higher than that of the normal control group ( $p < 0.01$ ).

Yucel, et al[15], examined changes of the upper airway cross-sectional area in each phase of respiration in different degrees of severity of OSAS with dynamic CT

Table 1.1: Correlation between AHI and Morphologic Properties in Conventional Studies

Feature	Method	Corr. Coef.	<i>p</i> -value	Literature
PNS-P	Cephalogram	0.194	0.011	Kitamura, et al [2]
		0.240		Yu, et al [4]
MP-H	Cephalogram	0.449	< 0.001	Kitamura, et al [2]
		0.310	< 0.01	Sakakibara, et al [5]
		0.420	< 0.001	Yucel, et al [6]
TGL	Cephalogram	0.320	< 0.01	Yu, et al [3]
		0.213	< 0.05	Sakakibara, et al [5]
H-VL	Cephalogram	0.430	< 0.01	Yu, et al [4]
		0.425	< 0.001	Sakakibara, et al [5]
BMI	Other	0.310	< 0.001	Kitamura, et al [2]
		0.525		Sakakibara, et al [5]
MMP	Visual Exam.	0.316	< 0.001	Kitamura, et al [2]
Tonsil Size	Visual Exam.	0.212	0.005	Kitamura, et al [2]

and investigated whether these changes have any correlation with sleep apnea severity parameters, including PSG and cephalometry. Forty seven patients and twenty four controlled subjects (habitual snorers) were studied. Cross-sectional area of the airway at the level of oropharynx and hypopharynx were obtained in each phase of quiet tidal breathing and at the end of both the forced inspiration and expiration. Six standard cephalometric measurements were made on the lateral scout view. All parameters were compared between controls and mild/moderate and severe OSAS groups. As a result, patients with severe OSAS had significantly narrower cross-sectional area at the level of uvula in expiration, more inferiorly positioned hyoid bone, and thicker soft palate compared with patients with mild/moderate OSAS ( $p < 0.05$ ) and the control group ( $p < 0.05$ ). In addition, severe OSAS patients had bigger neck circumference than those in the control group ( $p < 0.05$ ).

Sakakibara, et al[14] evaluated the cephalometric features in Japanese OSAS patients. In their study, 48 cephalometric variables were measured in 37 healthy males and 114 male OSAS patients. As a result, the non-obese OSAS patients showed several cephalometric defects compared with their BMI-matched normal



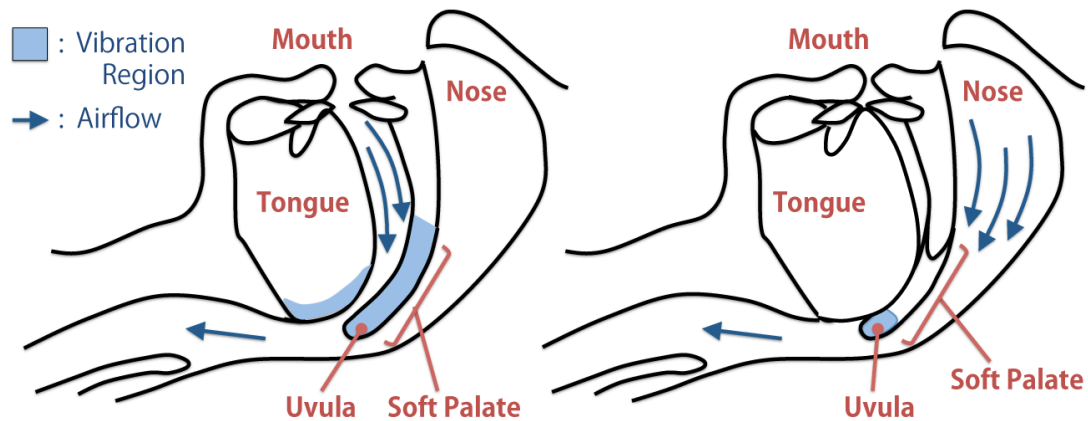


Figure 1.7: The vibration of tissues during snoring with oral breath (left) and with nasal breath (right)

controls: 1) decreased facial A-P distance at cranial base, maxilla and mandible levels and decreased bony pharynx width; 2) enlarged tongue and inferior shift of the tongue volume; 3) enlarged soft palate; 4) inferiorly positioned hyoid bone; and 5) decreased upper airway width at four different levels.

Moreover, the volume of the tongue and the airway capacity[20][21], fat volume[22], and the length around the neck[22][23] have been analyzed quantitatively. In conventional studies, the relation with AHI has also been discussed, but the more detailed morphological features of the soft tissues have not yet been analyzed so far.

## 1.6 Snoring Sounds

### 1.6.1 Sound-based Diagnosis

Snoring is one of the typical symptoms of OSAS, but they are not so much used for the diagnosis. But some of the characteristics of snoring sounds in OSAS patients have been empirically reported so far. The snoring of OSAS patients is known to be very loud and accompanied by gasps or choking. The acoustic properties have not been completely clarified quantitatively on account of its complicated mechanism.

Since it is, however, easy to record snoring sounds with only a microphone and a sound recorder, many researchers have explored the possibility for the use of medical diagnosis.

### 1.6.2 Relation of Oral Snoring to OSAS

The sound signal mostly related to sleep apnea is a snoring sound. 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 apnea syndrome [25].

However, benign persons often snore during sleep by the influence of muscle relaxation caused by the alcohol, fatigue, and/or advanced age. Such kind of snoring tends to be transient and not serious, but loud habitual snoring is a malign signal of OSAS.

Snoring sounds are mainly composed of both the vibrational sounds of the upper airway soft tissues (soft palate, uvula, tongue base, epiglottis, and tonsils) and the noise passing through the narrowed airway (figure 1.7). So, snoring sounds are a complex mixture of two or more vibration sounds and some noise. It depends upon the vibration and fluid dynamics influenced by breathing route during sleep, the degree of muscle relaxation, and a body position during sleep.

Oral snoring is a typical symptom of OSAS. Open mouth during sleep moves the lower jaw downward and the tongue base tends to obstruct the airway. Therefore, a person whose jaw is small, found in many Japanese people, tends to have OSAS even though he/she is not stout. This cause is different from the enlargement of the upper airway soft tissues. Thus, it is important to detect oral snoring to classify OSAS occurring from another reason which cannot be identified from the MR images.

### 1.6.3 Simulated Snoring Sounds

Simulated snoring sounds are generated from awake persons who are asked to breathe deeply enough to vibrate the soft palate, uvula and/or the other soft tissues

in their throat. Since this method is useful to control airflow through the nose and mouth precisely, some of the medical researchers have dealt with this method to analyze the relation between the respiratory route and the acoustic properties of snoring sounds. Liistro and colleagues observed[26] a decrease in the sagittal diameter of the oropharynx followed, during simulated snoring, by high-frequency oscillations of soft palate and pharyngeal walls. They found that the pattern of soft palate oscillations was different while snoring through the nose or mouth by the use of simulated snoring technique. Moreover, Dalmasso, et al[27], Lofaso, et al[28], Perez-Padilla, et al[33], and Beck, et al[29] have also adopted simulated snoring for medical or physiological studies.

In this study, the simulated snoring technique is also used to obtain oral and nasal snoring sounds. Then, the problem is whether the acoustic properties of simulated snoring sounds are equivalent to those of the actual ones. The similarity between them is discussed in the next subsection.

#### 1.6.4 Similarities between Natural and Simulated Snoring

Figures 1.8 and 1.9 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[30], and the tongue base snoring is known to occur with an open oral airway[31]. The same result is also reported by the other researcher[32]. 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 OSAS patients[33]. Since many OSAS 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

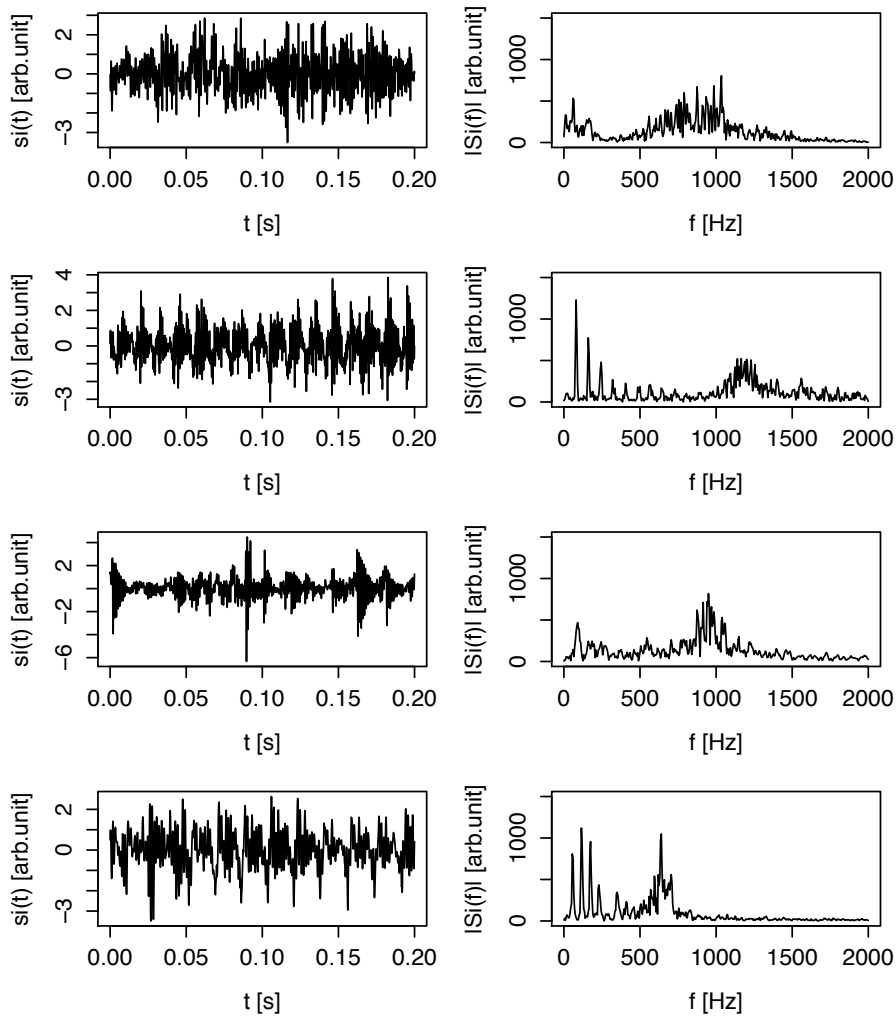


Figure 1.8: Waveforms and amplitude spectra of oral simulated snoring sounds and the sound spectra which consist of lower frequency components below 500Hz. Such acoustic property is also found in the natural snores the sound source of which is the soft palate[30][31]. It has been reported that the palatal snoring occurs with the oral airway closed[31], and the inspiratory nasal snores have a fundamental peak with associated harmonic peaks[33]. 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

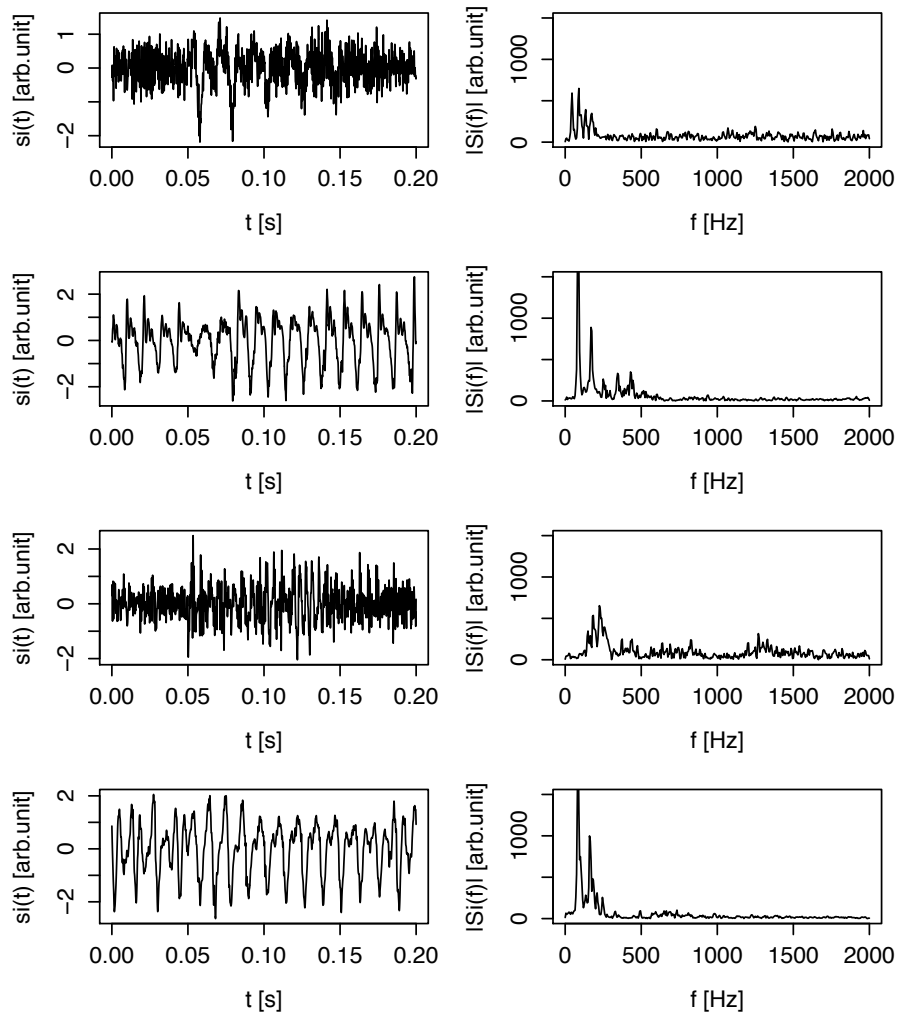


Figure 1.9: Waveforms and amplitude spectra of nasal simulated snoring sounds during sleep. Moreover, simulated snoring has also been adopted in some medical studies [26][34][35][28][29], and thus we decided to deal with simulated snoring sounds in this paper.

### 1.6.5 Related Studies

In this section, conventional studies of snoring sound analysis are introduced. The author gives notice here that the explanation of the conventional studies in this section are cited respectively from the original papers.

Perez-Padilla, et al, [33] 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 OSAS patients.

Fiz, et al[36] studied patients with simple snoring and OSAS, analyzing the acoustic properties of the snoring sounds. Spectral analysis of snoring sound showed the existence of two different patterns. The first pattern was characterized by the presence of a *fundamental frequency* and several harmonics. The second pattern was characterized by a low frequency peak with the sound energy scattered on a narrower band of frequencies, but without clearly identified harmonics. The seven simple snorers and two of the 10 patients with OSAS (AHI=13 and 14, respectively) showed the first pattern. The rest of the OSAS patients showed the second pattern. The peak frequency of snoring was significantly lower in OSAS patients, with all but one OSAS patient and only one simple snorer showing a peak frequency below 150 Hz. Significant differences in the sound power spectrum of snoring sound between subjects with simple snoring and obstructive sleep apnea patients.

Herzog, et al[37], examined simulated snoring under conditions awake, recorded the produced snoring sounds and compared those sounds with nocturnal snoring. Snoring sounds were analyzed by FFT and the intensity peaks 1-5 were evaluated. Rhythmic and non-rhythmic snoring events were distinguished depending on present obstructive apneas. Clinical and PSG data were correlated with the results of the frequency analysis of the snoring sounds. Simulated snoring sounds revealed a low frequency of 200 Hz in intensity peaks 1 and 2 with an increase up to 3kHz in peaks 3-5. Similar frequency patterns were detected in rhythmic nocturnal snoring. Non-rhythmic snoring events revealed frequency patterns between 2kHz and 3kHz in all intensity peaks. Simulated snoring resembles rhythmic nocturnal snoring with low-frequency intensity peaks, whereas non-rhythmic snoring revealed high frequencies. The examination during simulated snoring and frequency analysis of snoring sounds might contribute in locating the pathogenesis of snoring.

Hara, et al[38], determined whether the acoustic characteristics of snoring sounds differed between simple snorers and patients with obstructive sleep apnea syndrome (OSAS) by using a multidimensional voice program (MDVP) that analyzes various aspects of voice. Natural overnight snoring was recorded from each subject while they slept during PSG. For data analysis, four markers were used: peak frequency, soft phonation index (SPI), noise to harmonics ratio (NHR), and power ratio. As a result, the Mann-Whitney U test revealed significant differences between the SPI, NHR, and power ratio of simple snorers and patients with OSAS. Simple snorers had a high SPI value. OSAS-related snorers demonstrated a high NHR and low power ratio. MDVP can be used for snoring sound analysis as a noninvasive examination of sleep-related breathing disorders for differential diagnosis. However, a suitable option that is rapid and has an easy-to-use interface would be more advantageous for analyzing snoring sounds.

Sola-Soler, et al, [39] have showed significant differences in formant frequencies variability between simple snorers and OSAS patients even when non-postapneic snores were considered.

Ng, et al, [40] have also found quantitative differences in formant frequencies between apneic and benign snorers by the use of a Linear Predictive Coding (LPC) technique. Formant frequencies (F1, F2, and F3) were extracted from the LPC spectrum for analysis. The accuracy of this approach was assessed using receiver operating characteristic curves and notched box plots. As a result, quantitative differences in formant frequencies between apneic and benign snores are found in same- or both-gender snorers. Apneic snores exhibit higher formant frequencies than benign snores, especially F1, which can be related to the pathology of OSAS. This study yields a sensitivity of 88%, a specificity of 82%, and a threshold value of  $F1 = 470$  Hz that best differentiate apneic snorers from benign snorers (both gender combined).

Emoto, et al, [6] proposed a novel approach to the diagnosis of OSAS based on the formants of snoring sounds, extracted via a noise-robust linear prediction

technique. The proposed method and existing LPC-based method are compared via a measure, which indicates the standard deviation of first formant frequencies. The performance of the proposed method was evaluated on a database of clinical snoring sounds recorded overnight in the laboratory of a hospital sleep diagnostic center. Compared with existing LPC-based method, they showed that the proposed method can differentiate (sensitivity : 88.9%, specificity : 88.9%, AUC : 0.85) between benign snoring (Apnea Hypopnea Index,AHI =  $6.0 \pm 3.2$ ; 6188 episodes) and apneic snoring ( AHI =  $40.7 \pm 20.2$ ; 14066 episodes)

As far as the author knows, the other researchers worldwide such as Dalmaso, et al[27], Sola-Soler, et al[7], Abeyratne, et al[41], Osborne, et al[42], Osborne, et al[42], and so on. Conventional studies related to snoring sounds have been surveyed in detail by Pevernagie, et al [43] in 2009. According to the above-mentioned representative studies, the acoustic properties of snoring sounds are so various and complex that one cannot easily extract useful information about the sleep apnea and still does not reach to a general conclusion of the acoustic properties of OSAS-related snoring sounds.

## 1.7 Summary

SAS is not necessarily a personal problem but a social one, which annoys the other people by their loud snoring and sometimes causes a serious accident. The patients should receive a medical treatment in the earlier stage, but they cannot perceive their own symptoms during sleep.

In the present state of SAS screening, sleep conditions of the patients are recorded from EEG, ECG, SpO<sub>2</sub>, and the other contact sensors during sleep. They must sleep at hospital overnight, and many contact sensors prevent them to sleep as well as at home. This sometimes causes the underestimation of the severity. Thus, other approaches to simpler screening for SAS have been focused on recently.

In the upper airway MR images, the enlargement of the upper airway soft



tissues are clearly photographed and often used for visual examination of medical specialists. The enlargement is mainly caused by obesity and directly obstructs the upper airway. Indeed, it is known that the Body-Mass Index (BMI) is highly related to the severity of OSAS[12]. Snoring is a vibration of the upper airway soft tissues and generates a loud sound especially when oral breathing. Oral snoring is a typical symptom of OSAS and the detection is quite important. Open mouth during sleep moves the lower jaw downward. The tongue base also moves downward in accompaniment with the jaw and obstructs the airway. Therefore, many of the people whose jaw is small have OSAS even though they are not stout.

This study pays attention to these two biomedical signals which become different causes of OSAS respectively, and tries to extract the morphological and the acoustic properties of these signals in OSAS patients.

## References

- [1] L. J. Findley, M. E. Unverzagt, and P. M. Suratt. Automobile accidents involving patients with obstructive sleep apnea. *American Review of Respiratory Disorders*, Vol. 138, pp. 337–340, 1988.
- [2] H. Toga and M. Takahashi. Essentials of diagnosis and treatment in sleep-disordered breathing. *Medical Review*, 2006.
- [3] Wikipedia. Obstructive sleep apnea. [http://en.wikipedia.org/wiki/Obstructive\\_sleep\\_apnea](http://en.wikipedia.org/wiki/Obstructive_sleep_apnea).
- [4] C. Guilleminault, A. Tilkian, and W. C. Dement. The sleep apnea syndromes. *Annual Review of Medicine*, Vol. 27, pp. 465–484, 1976.
- [5] The homepage of WebMD. an american corporation which provides health information services. <http://www.webmd.com/>.
- [6] T. Emoto, U. R. Abeyratne, T. Kusumoto, M. Akutagawa, E. Kondo, I. Kawata, T. Azuma, S. Konaka, and Y. Konouchi. Discriminating apneic snorers and benign snorers based on snoring formant extracted via a noise-robust linear prediction technique. *Journal of Japanese Society for Medical and Biological Engineering*, Vol. 48, No. 1, pp. 115–121, 2010.
- [7] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. *Engineering in Medicine and Biology Society, 2003 Proc. of the 25th Annual International Conference of the IEEE*, Vol. 3, pp. 2527–2530, 2003.

- 
- [8] T. Mikami, K. Yonezawa, Y. Kojima, M. Yamamoto, and M. Furukawa. Usefulness of image analysis on upper airway mri for the classification of severe obstructive sleep apnea syndrome (OSAS). *Medical Imaging and Information Sciences (in press)*, 2014.
- [9] Wikipedia. Polysomnography. <http://en.wikipedia.org/wiki/Polysomnography>.
- [10] R. Arens, J. M. McDonough, A. M. Corbin, M. E. Hernandez, G. Maislin, R. I. Schwab, and A. I. Pack. Linear dimensions of the upper airway structure during development, assessment by magnetic resonance imaging. *American Journal of Respiratory and Critical Care Medicine*, Vol. 165, pp. 117–122, 2002.
- [11] T. Kitamura, A. Sakabe, N. Ueda, T. Shiomori, T. Udaka, T. Ohbuchi, and H. Suzuki. Usefulness of cephalometry and pharyngeal findings in the primary diagnosis of obstructive sleep apnea syndrome. *Nippon Jibiinkoka Gakkai Kaiho*, Vol. 111, No. 11, pp. 695–700, 2008.
- [12] The Research Group of Sleep-Disordered Breathing, editor. *Guideline for Diagnosis and Treatment of Adult SAS*. Medical Review, 2008.
- [13] X. Yu, K. Fujimoto, K. Urushibata, Y. Matsuzawa, and K. Kubo. Cephalometric analysis in obese and nonobese patients with obstructive sleep apnea syndrome. *Chest*, Vol. 124, pp. 212–218, 2003.
- [14] H. Sakakibara, M. Tong, K. Matsushita, M. Hirata, Y. Konishi, and S. Suet-sugu. Cephalometric abnormalities in non-obese and obese patients with obstructive sleep apnoea. *European Respiratory Journal*, Vol. 13, pp. 403–410, 1999.
- [15] A. Yucel, M. Unlu, A. Haktanir, M. Acar, and F. Fidan. Evaluation of the upper airway cross-sectional area changes in different degrees of severity of ob-

- 
- structive sleep apnea syndrome: Cephalometric and dynamic ct study. *American Journal of Neuroradiology*, Vol. 26, pp. 2624–2629, 2005.
- [16] T. Mochizuki. Examination of low palatal arch with long low-hanging soft palate in obstructive sleep apnea syndrome and cephalometry. *Nippon Jibi-inkoka Gakkai Kaiho*, Vol. 100, pp. 937–945, 1997.
- [17] K. A. Ferguson, T. Ono, A. A. Lowe, C. F. Ryan, and J. A. Fleetham. The relationship between obesity and craniofacial structure in obstructive sleep apnea. *Chest*, Vol. 108, pp. 375–385, 1995.
- [18] M. Partinen, C. Guilleminault, M-A. Quera-Salva, and A. Jamieson. Obstructive sleep apnea and cephalometric roentgenograms, the role of anatomic upper airway abnormalities in the definition of abnormal breathing during sleep. *Chest*, Vol. 93, pp. 1199–1205, 1988.
- [19] J. A. Dempsey, J. B. Skatrud, A. J. Jacques, S. J. Ewanowski, B. T. Woodson, P. R. Hanson, and B. Goodman. Anatomic determinants of sleep-disordered breathing across the spectrum of clinical and nonclinical male subjects. *Chest*, Vol. 122, No. 3, pp. 840–851, 2002.
- [20] R. Schwab, M. Psirstein, R. Pierson, A. Mackley, R. Hachadoorian, R. Arenas, G. Maislin, and A. Pack. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *American Journal of Respiratory and Critical Care of Medicine*, Vol. 168, pp. 522–530, 2003.
- [21] K. C. Welch, G. D. Foster, C. T. Ritter, T. A. Wadden, R. Arens, G. Maislin, and R. J. Schwab. A novel volumetric magnetic resonance imaging paradigm to study upper airway anatomy. *Sleep*, Vol. 25, No. 5, pp. 532–542, 2002.
- [22] I. L. Mortimore, I. Marshall, P. K. Wraith, R. J. Sellar, and N. Douglas. Neck and total body fat deposition in nonobese and obese patients with sleep apnea

- compared with that in control subjects. *American Journal of Respiratory and Critical Care Medicine*, Vol. 157, pp. 280–283, 1998.
- [23] D. Rodenstein, G. Doods, Y. Thomas, G. Liistro, D. C. Stanescu, C. Culee, and G. Aubert-Tulkens. Pharyngeal shape and dimensions in healthy subjects, snorers, and patients with obstructive sleep apnea. *Thorax*, Vol. 45, pp. 722–727, 1990.
- [24] C. Iida-Kondo, N. Yoshino, T. Kurabayashi, S. Mataka, M. Hasegawa, and N. Kurosaki. Comparison of tongue volume/oral cavity volume ratio between obstructive sleep apnea syndrome patients and normal adults using magnetic resonance imaging. *Journal of Medical and Dental Science*, Vol. 53, pp. 119–126, 2006.
- [25] M. Ichioka. Respiratory organs and sleep disorders. *Folia Pharmacologica Japonica (in Japanese)*, Vol. 129, pp. 432–435, 2007.
- [26] G. Liistro, D. Stanescu, and C. Veriter. Pattern of simulated snoring is different through mouth and nose. *Journal of Applied Physiology*, Vol. 70, No. 6, pp. 2746–2741, 1991.
- [27] F. Dalmaso and R. Prota. Snoring: analysis, measurement, clinical implications and applications. *European Respiratory Journal*, Vol. 9, pp. 146–159, 1996.
- [28] F. Lofaso, K. Leroux, M. A. Quera-Salva, G. Mroue, M. P. D’Ortho, D. Isabey, and B. Louis. Snoring detection during auto-nasal continuous positive airway pressure. *European Respiratory Journal*, Vol. 19, pp. 108–112, 2002.
- [29] R. Beck, M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *European Respiratory Journal*, Vol. 8, pp. 2120–2128, 1995.

- 
- [30] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. E. Camilleri. Sound frequency analysis and the site of snoring in natural and induced sleep. *Clinical Otolaryngology*, Vol. 27, pp. 162–166, 2002.
- [31] S. J Quinn, L. Huang, P. D. M. Ellis, and J. E. F. Williams. The differentiation of snoring mechanisms using sound analysis. *Clinical Otolaryngology*, Vol. 21, pp. 119–123, 1996.
- [32] A. K. Bieger-Farhan, N. K. Chadha, A. E. Camileri, P. Stone, and K. McGuinness. Portable method for the determination of snoring site by sound analysis. *Journal of Laryngology and Otology*, Vol. 118, No. 2, pp. 135–138, 2004.
- [33] J. R. Perez-Padilla, E. Slawinski, L. M. Difrancesco, R. Feige, J. E. Remmers, and W. A. Whitelaw. Characteristics of the snoring noise in patients with and without occlusive sleep apnea. *American Review of Respiratory Disorders*, Vol. 147, No. 3, pp. 635–644, 1993.
- [34] 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. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 12, pp. 1553–1562, 2008.
- [35] M. Herzog, T. Metz, A. Schmidt, T. Bremert, W. Hosemann B. Venohr, and H. Kaftan. The prognostic value of simulated snoring in awake patients with suspected sleep-disordered breathing: Introduction of a new technique of examination. *Sleep*, Vol. 29, No. 11, pp. 1456–1462, 2006.
- [36] J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mananas, and P. Caminal. Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. *European Respiratory Journal*, Vol. 9, No. 11, pp. 2365–2370, 1996.

- 
- [37] M. Herzog, A. Schmidt, T. Bremert, B. Herzog, W. Hosemann, and H. Kaftan. Analysed snoring sounds correlate to obstructive sleep disordered breathing. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 1, pp. 105–113, 2008.
- [38] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita. Acoustic analysis of snoring sounds by a multidimensional voice program. *Laryngoscope*, Vol. 116, No. 3, pp. 379–381, 2006.
- [39] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. *Proc. of Annual Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, Vol. 3, pp. 2527–2530, 2003.
- [40] A. K. Ng, T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, Vol. 9, No. 8, pp. 894–898, 2008.
- [41] U. R. Abeyratne, A. S. Wakwella, and C. Hukins. Pitch jump probability measures for the analysis of snoring sounds in apnea. *Physiological Measurement*, Vol. 26, No. 5, pp. 779–798, 2005.
- [42] J. E. Osborne, E. Z. Osman, P. D. Hill, B. V. Lee, and C. Sparkes. A new acoustic method of differentiating palatal from non-palatal snoring. *Clinical Otolaryngology*, Vol. 24, No. 2, pp. 130–133, 1999.
- [43] D. Pevernagie, R. M. Aarts, and M. D. Meyer. The acoustics of snoring. *Sleep Medicine Review*, Vol. 14, No. 2, pp. 131–144, 2010.

## Chapter 2

# Morphological Analysis of the Upper Airway Soft Tissues in MR Images for Classification of Severe OSAS

### 2.1 Introduction

PSG is a gold standard method for diagnosing SAS and is widely used in common. As is described in chapter 1, It consists of tens of contact sensors such as EEG, ECG, EMG, EOG, thermistor, pulse oximetry, and so on. The sensors should be equipped on the patients body so as to estimate AHI, which is used for objective index for the diagnosis and treatment. But PSG requires patients to stay in hospital overnight, so in many cases they cannot sleep as deeply as usual and the severity tends to be underestimated. Moreover, PSG is also influenced by the first night effect[1] also introduced in chapter 1. In order to avoid this effect, it is better to record the biosignals in the same hospital room for several consecutive nights, but actually the several times of recording are done only when the obviously irregular results are identified through the first night. Normally the PSG data recorded through the first night are used for medical diagnosis.

On the other hand, a simple monitoring instrument is adoptable for home monitoring. It is possible to begin a medical treatment without using the PSG



when the patient is clearly severe according to the home monitoring result. It can be considered that this method is not influenced as much as PSG by the first night effect because the patient is allowed to sleep at his/her home. But overnight recording is also necessary and this method tends to underestimate the severity.

By the way, according to the MR images, it can be found that the upper airway in OSA patients is narrower than those in benign persons. Especially, the enlargement of the upper airway soft tissues such as soft palate, uvula, tongue, and tonsils are the main factors of airway obstruction. Morphological features of them have been quantitatively analyzed so far by the use of cephalogram [2][3][4][5][6][7][8][9], and MR images [10][11][12][13][14][15].

Cephalogram is the measurement of the human head by X-ray imaging and is used to gauge the size and spacial relationships of the teeth, jaws, and cranium. Recently, it has been reported that the distance from the mandibular plane to the hyoid bone (MP-H) and the length of the soft palate (PNS-P) are highly correlated with AHI. In X-ray images, hard tissues such as bones or cartilage can clearly be recognized. However, since the soft tissues are not so clearly recognized in X-ray images, much experience in image diagnosis is necessary to visually determine the gaging points.

In the case of MR images, soft tissues are clearly recognized and three-dimensional evaluation is also practicable. In conventional studies, the volume of the tongue and the airway capacity[11][15], fat volume[13], and the length around the neck[13][14] have been analyzed quantitatively. Arens, et al [12] evaluated the relation of the length from the mandible bone to the clivus with AHI in a little more detail than conventional studies, but the detailed morphological features of the soft tissues have not sufficiently been analyzed so far.

This chapter analyzes quantitatively the tongue morphology and the cross sectional area of the narrowest upper airway in the sagittal and cross sectional planes of the upper airway MR images and discuss the relation between those features and the severity of OSA. Also, the severe OSA is detected by thresholding the severity

Table 2.1: Photographing conditions of the sagittal and transverse planes in the upper airway MRIs

	T1-weighted Sagittal Plane	T1-weighted Transverse Plane
TE	12 ms	12 ms
TR	450 ms	518 ms
Field of View	256×256 mm <sup>2</sup>	256×256 mm <sup>2</sup>
Slice Thickness	4.5 mm	5 mm
Slice Gap	0.5 mm	1 mm
Num. of Slices	20	25
Matrix	205×205	205×205

predicted with the linearly weighted image features obtained from multi-regression analysis. The performance is evaluated based on the sensitivity, specificity, and efficiency.

## 2.2 Subjects and Instrument

The subjects are 43 male OSAS patients (ages: 24-83, the median of the age: 52, AHI: 3.9-90.3, the median of AHI: 30.4), who have received the medical treatment at National Hospital Organization Hakodate Hospital. The upper airway MR images are obtained from all of them.

The instrument is the InteraMaster R8 made by PHILIPS Co.Ltd., and the magnetic field strength is 1.5T. The upper airway MR images analyzed here are composed of T1-weighted sagittal and transverse planes (see figure 2.1) under the conditions shown in table 2.1. While photographing, the subjects lie on their back in bed and their head is completely stabilized. The examples of the sagittal planes photographed from mild or severe patients are shown in figure 2.4.

The value of AHI is estimated from the result of PSG screening by medical technologists in NHO Hakodate Hospital. In this study, the patients whose AHI is greater than or equal to 30 is defined as *severe* SAS patients based on the AASM (American Academy for Sleep Medicine) Severity Standard [3][16]. The AHIs of

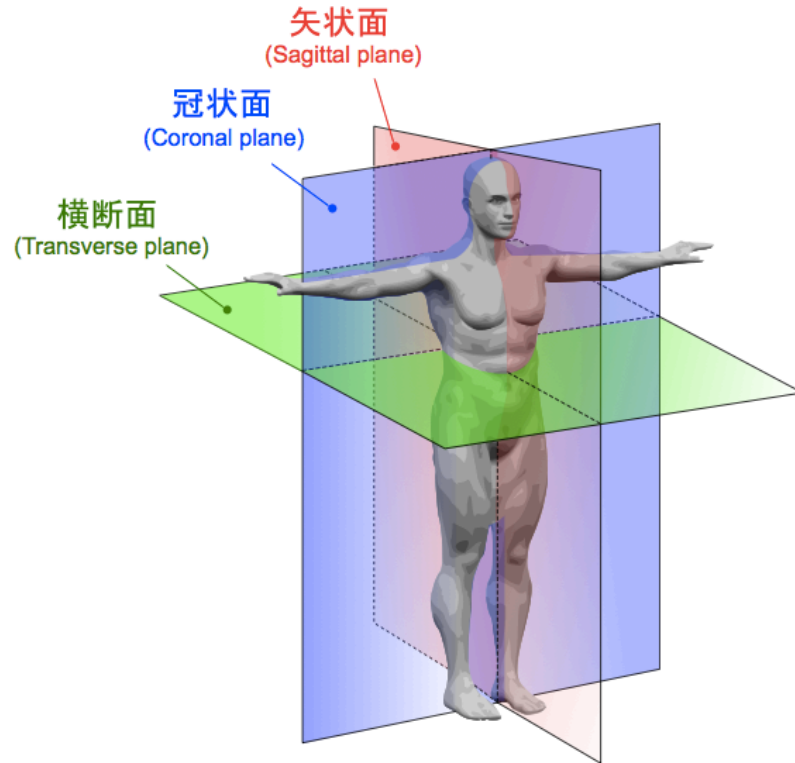


Figure 2.1: Sagittal and transverse planes of MR Images (cited from Wikimedia Commons, [http://upload.wikimedia.org/wikipedia/commons/e/e1/Human\\_anatomy\\_planes.svg](http://upload.wikimedia.org/wikipedia/commons/e/e1/Human_anatomy_planes.svg))

21 subjects are less than 30, while those of the other 22 subjects are greater than or equal to 30.

Before conducting the analysis, the author had obtained the permission to use the patients' medical data for the purpose of this study at the Ethical Committee at NHO Hakodate Hospital.

## 2.3 Feature Extraction Method

### 2.3.1 Tongue Region Segmentation

In order to evaluate the morphological features of the tongue region in the sagittal plane, it is necessary to perform the segmentation of the tongue region. In

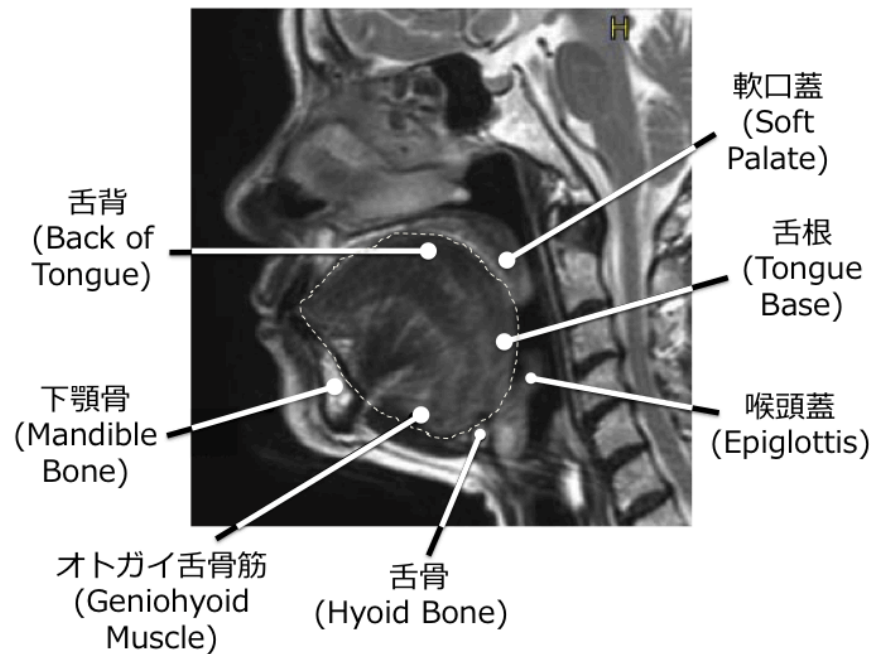


Figure 2.2: Anatomic chart of the upper airway and the tongue region

conventional studies, segmentation of many organs has been performed in various algorithms, such as brain[17][18], liver[19], and blood vessels[20][21]. But no methods of the tongue region segmentation have been proposed. Since conventional methods for the other organs are based on the own characteristics, they cannot necessarily be applied to the tongue segmentation. Thus, the author considered it difficult to develop an automation technique for the tongue region segmentation and for now decided to extract the tongue region manually one by one on the OsiriX, a medical image processing software.

The tongue region in this study is defined as the region enclosed by the dotted line in figure 2.2, which includes the tongue base and the geniohyoid muscle, and does not include the epiglottis, hyoid bone, and mandible bone. This region can easily be recognized.

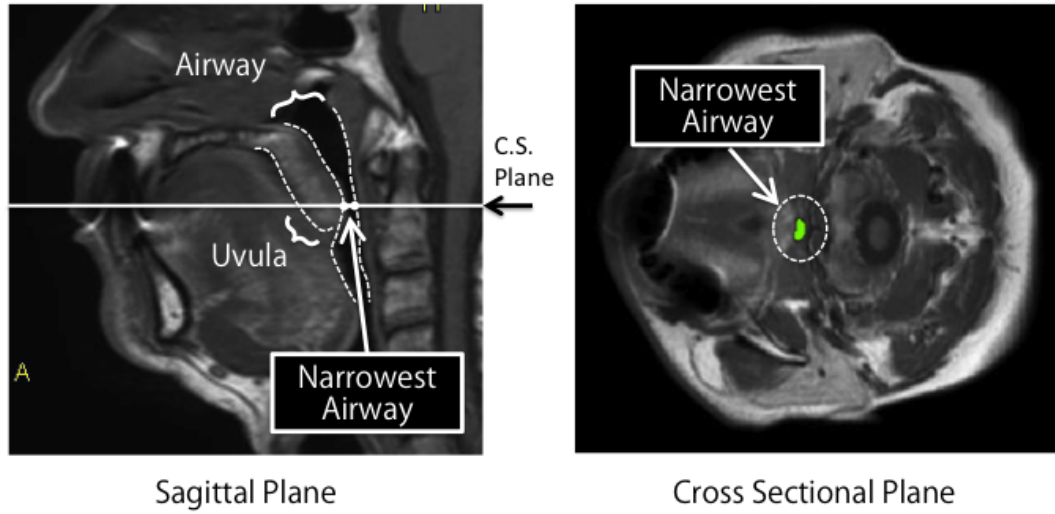


Figure 2.3: Selection of the cross-sectional MR image through the narrowest airway

### 2.3.2 Morphological Feature of the Tongue Region

The center of gravity of the whole extracted tongue region is firstly calculated. From this point, straight lines are radially extended at the fixed angular intervals of  $\theta = \pi/24$  radian, and the radius  $L_i$  ( $i = 1, 2, \dots, 48$ ) from the center to the circumference of the tongue region is estimated (figure 2.5). Since the patients lie on the bed while photographing, vertical direction in the image indicates the bed surface and is defined as zero radian. In addition,  $L_i(n)$  is defined as  $i$ th length of  $n$ th subject. Such feature extraction method has widely been used in many pattern recognition applications[22].

Precisely, the bending angle of the subject's head while photographing is slightly various in subject and this affects the definition of  $\theta$  direction. It is desirable that this direction should be revised by the bending angle of the head when the subject lies on bed. But since there are not any quantitative and objective methods to measure the bending angle from the MR images, the author thought that such revision is difficult under the present conditions. However, the photographed images are obtained from the patients whose head is completely stabilized guided by

the radiological technologists for medical diagnosis. Accordingly, it is considered that the patients' body positions when photographing are sufficiently common for objective evaluation.

### 2.3.3 Sectional Area of the Narrowest Upper Airway

Next, the author focused on the cross sectional plane of the MRI to evaluate the *narrowness* of the upper airway. It is expected to obtain useful features for airway obstruction caused by the enlargement of tonsils, which cannot be evaluated using the sagittal plane. The MR images are not the ones taken from the OSA patients during sleep, but it is known that the airway narrowing can also be found in awake OSA patients[23]. Many cross sectional planes are taken from the top of the patient's head, but we adopted the image whose sectional area is the narrowest compared with the other slices (figure 2.3).

The narrowest cross sectional area in the upper airway is evaluated by the number of image pixels. As same as the tongue region segmentation, the author painted the sectional area of the upper airway in green color manually using a mouse on the image processing software OsiriX. The green pixels are automatically counted by a computer program on the statistical software R, and in this study the number of counted pixels ( $S(n)$ ) is defined as the sectional area of the narrowest airway from the MR image of  $n$ -th subject. So,  $S(n)$  is not an actual area whose unit is  $\text{cm}^2$  but the number of pixels in the image whose unit is "pixels".

## 2.4 Evaluation Method

### 2.4.1 Correlation Coefficient with Apnea/Hypopnea Index

The quantified morphological features estimated from the upper airway MRIs are evaluated by calculating the correlation coefficient,  $R$ , (Pearson's correlation) with Apnea/Hypopnea Index (AHI). Let  $X(n)$  be one of the morphological features estimated from the  $n$ th subject, and  $\text{AHI}(n)$  is the value of AHI estimated from

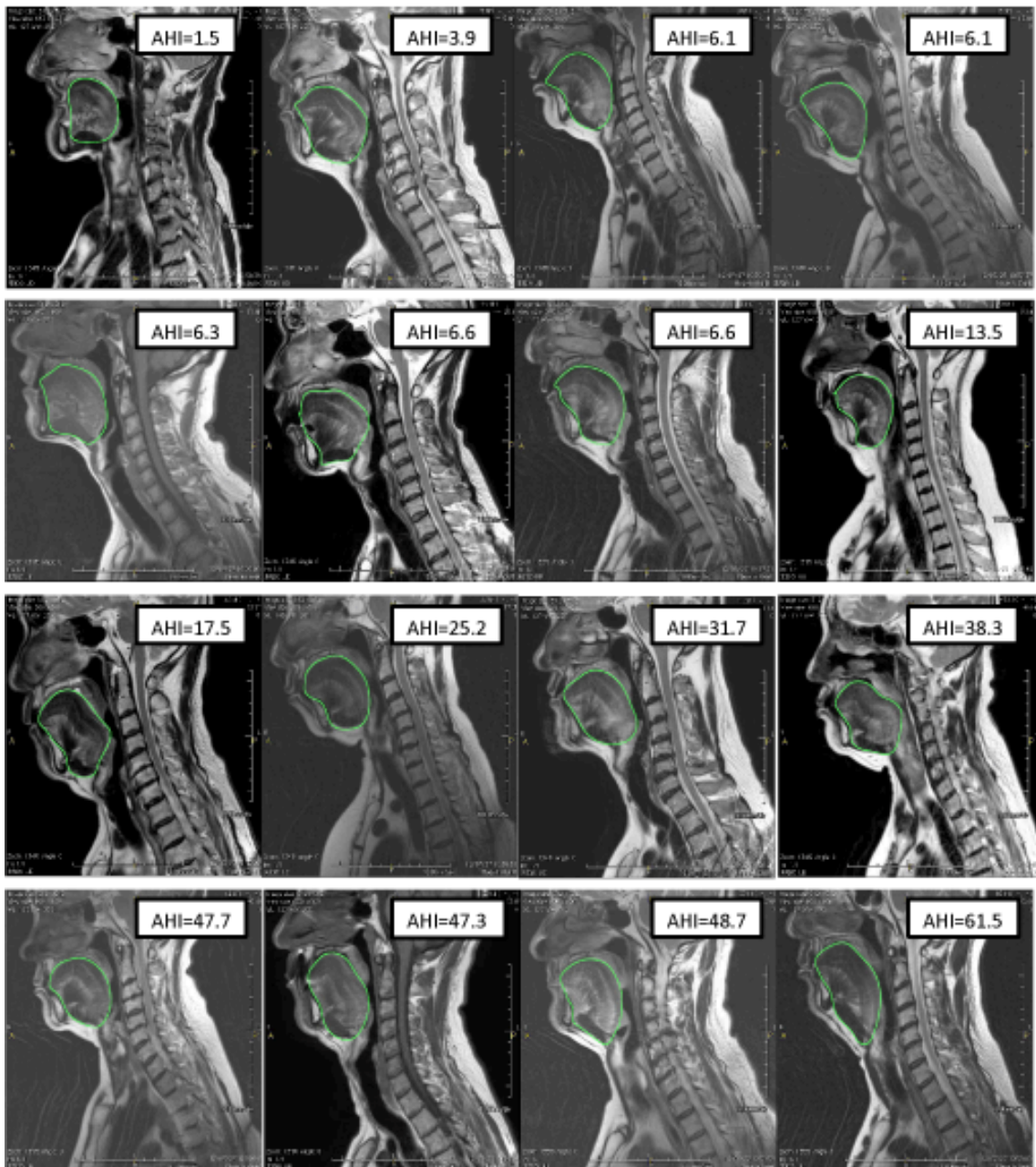


Figure 2.4: Upper airway MRIs of the subjects from benign to severe OSA

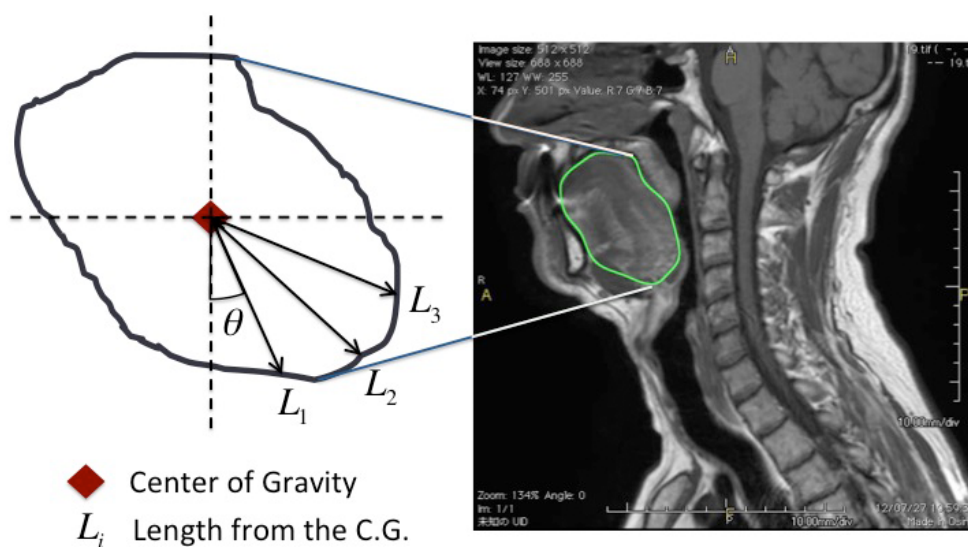


Figure 2.5: Directions whose correlation coefficients are greater than 0.5

PSG screening of the  $n$ th subject. The correlation coefficient,  $R$ , is calculated by

$$R = \frac{\sum_{n=1}^N (\text{AHI}(n) - \overline{\text{AHI}})(X(n) - \overline{X})}{\sqrt{\sum_{n=1}^N (\text{AHI}(n) - \overline{\text{AHI}})^2} \sqrt{\sum_{n=1}^N (X(n) - \overline{X})^2}} \quad (2.1)$$

where  $\overline{X}$  and  $\overline{\text{AHI}}$  are, respectively, calculated by

$$\overline{X} = \frac{1}{N} \sum_{n=1}^N X(n) \quad (2.2)$$

$$\overline{\text{AHI}} = \frac{1}{N} \sum_{n=1}^N \text{AHI}(n) \quad (2.3)$$

where  $N$  is the number of subjects ( $N = 43$ ).

#### 2.4.2 Classification of the Severe OSAS using Multi-Regression Analysis

The severity of OSA is predicted by the use of multi-regression analysis on the tongue morphology and the narrowness of the upper airway. Let the explanatory



variables be  $Y_i$ , and the predicted severity is calculated by

$$\text{AHI}' = \sum_i w_i Y_i + w_0 \quad (2.4)$$

where  $\text{AHI}'$  is a prediction value of the severity. Notice that the explanatory variables  $Y_i$  are selected from the feature values highly correlated with AHI. Then, the severe OSA is detected as follows:

If  $\text{AHI}' > \theta$  then the patient is evaluated as *Severe* else *Non-severe*

In this study, the threshold  $\theta$  is set to 30, a criterion of severe SAS defined by American Association of Sleep Medicine (AASM). Finally, the detection results are compared with the judgement from the actual AHI values, and evaluated using three values: sensitivity, specificity, and efficiency. Each measures are calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.5)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (2.6)$$

$$\text{Efficiency} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.7)$$

where  $TP, FN$  are the number of severe subjects classified to the severe and the non-severe category, and  $FP, TN$  are the number of non-severe subjects classified to the severe and the non-severe category (see table 2.2).

Namely, if the sensitivity is high, many severe patients are correctly classified to the severe category, while if the specificity is high, many non-severe patients are correctly classified to the non-severe category. Also, if the efficiency is high, many of both severe and non-severe patients are correctly classified to the respective category.

Table 2.2: Contingency table made by comparing the classification results and the actual severity

Classification Results	Actual Severity	
	Severe (Positive)	Non-Severe (Negative)
Severe (Positive)	True Positive ( $TP$ )	False Positive ( $FP$ )
Non-Severe (Negative)	False Negative ( $FN$ )	True Negative ( $TN$ )

### 2.4.3 Comparison with the Other Features in Conventional Studies

As is described in section 2.1, some morphological and volumetric features of the oropharyngeal tissues have been focused on to clear the relation with sleep apnea. Although the detailed morphological features of the soft tissues have not sufficiently been analyzed so far, it is necessary to compare those features with the method in this chapter.

Table 2.3 shows the overview of morphological and volumetric features analyzed in conventional studies. Since a technical method to determine the region of interest (ROI) has not been described in many papers, it should be noticed that conventional methods are not sufficiently objective. Because some of the soft tissues are not clearly photographed in MR images very often and there are individual differences in morphological features, a precise definition to determine the ROI should be needed. So, in order to compare with those conventional features, a detailed method to determine the ROI is firstly defined in this subsection.

Some of the conventional features should be omitted for some reasons. Since the abdominal circumference cannot be estimated from the upper airway MRI, it is left out from the comparison targets. In addition, some features around the Mandibular plane proposed in [24] are excluded because this method is so complex and troublesome that it could not be useful to realize an automatic classification method. In order to estimate the tongue and soft palate volume, it is necessary to precisely determine the ROIs of the tongue and soft palate in transverse planes. But in some slices the boundary of them are not clearly recognized, so the objective

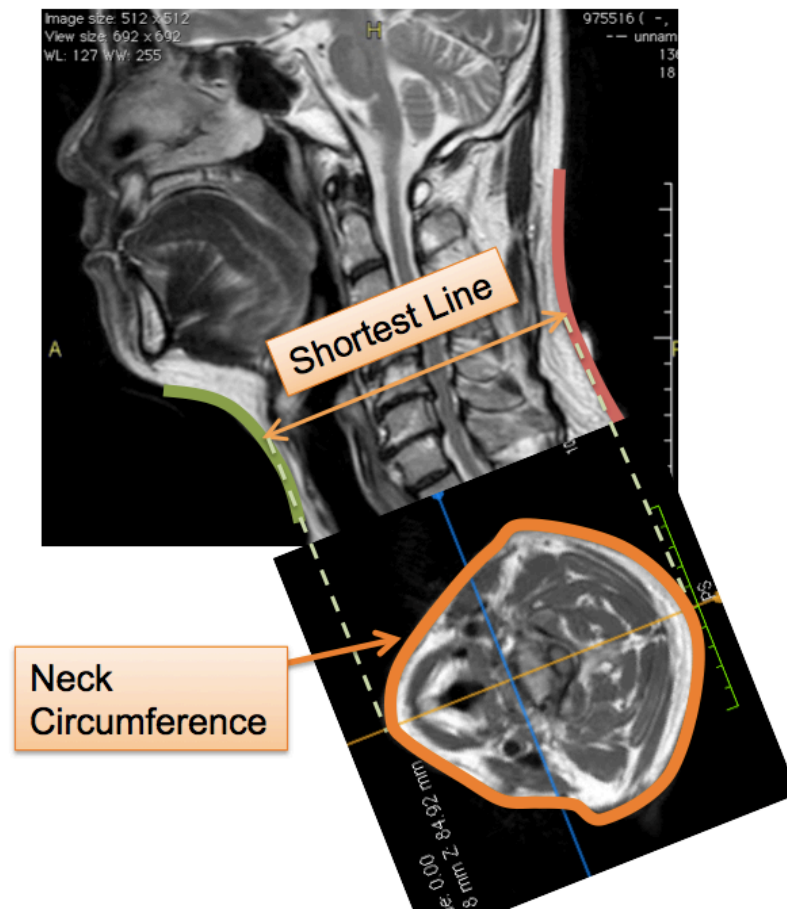


Figure 2.6: Procedure to determine the neck circumference

evaluation cannot be done without the sufficient anatomical knowledge. According to this, comparison targets in this section are (1).tongue area in sagittal plane, (2).tongue circumference in sagittal plane, (3).upper airway volume, and (4).neck circumference.

The tongue circumference and tongue area in sagittal plane are easy to be determined at hand according to the method in section 2.3.1.

The neck circumference is determined by slight complex procedures. Firstly, in sagittal plane, the boundary from the jaw to the *Adam's apple* and that around the nape of the neck are respectively marked with green and red using a mouse on the screen. Next, all distances between a point on the red line to the one on

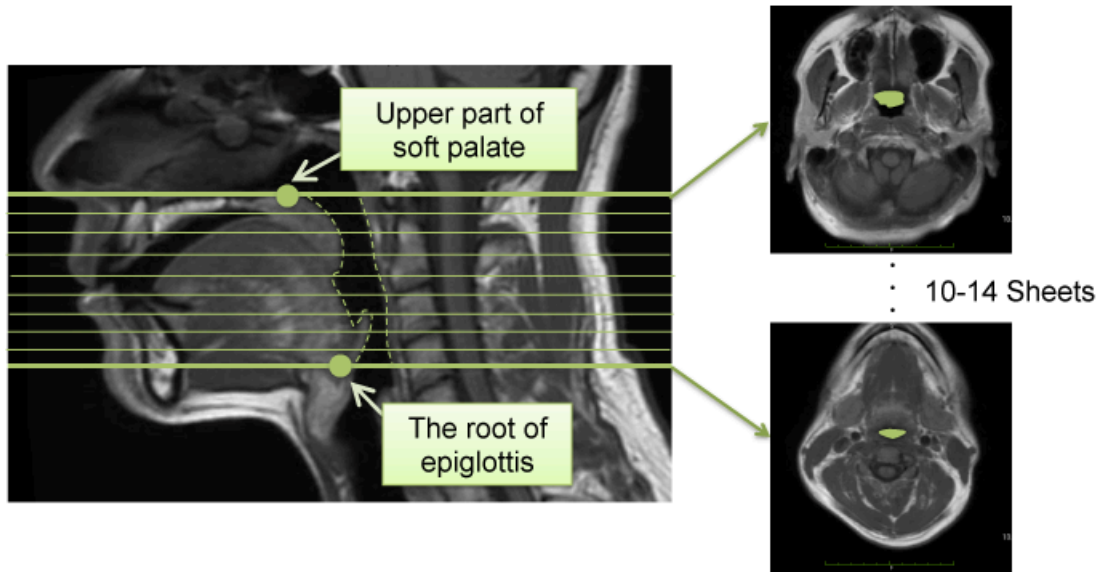


Figure 2.7: Procedure to determine the transverse planes for tongue volume estimation

the green line are calculated, and the shortest distance is determined. Then, the cross sectional plane of the neck along with the line with the shortest distance is obtained using a medical imaging software, OsiriX. Finally, the circumference of the cross sectional plane is calculated.

The airway volume is estimated based on transverse planes. Firstly, the transverse plane which comes contact with the back of the tongue is determined. Similarly, the one which crosses the root of the epiglottis is also determined. All transverse planes from the first plane to the last one are used to estimate the airway volume. In each plane, the cross sectional area of the airway is manually painted out. Since the hollow is photographed in a black or darker color and the surrounding tissues are in much lighter gray, there is no problem about such manual coloring for objective evaluation. Finally, these colored regions in transverse planes are reconstructed in 3D images using the OsiriX. The volume is automatically calculated by this software.

Table 2.3: Overview of the oropharyngeal features having been focused on in conventional studies

Features	Literature
Tongue Area in SP	B. L. Herrmann, et al, <i>Eur. J. Endocrinol.</i> , 151, 2004
Tongue Circum. in SP	B. L. Herrmann, <i>Eur. J. Endocrinol.</i> , 151, 2004
Tongue Volume	R. J. Scwab, et al, <i>AJRCCM</i> , 2003
Soft Palate Volume	M. Okubo, et al, <i>Sleep</i> , 2006
Abdominal Circum.	V. Hoffstein, <i>Eur. Respir. J.</i> , 5, 1992
Upper Airway Volume	C. Iida-Kondo, et al, <i>J. Med. Dent. Sci.</i> , 53, 2006
Neck Circumference	R. J. Schwab, et al, <i>AJRCCM</i> , 2003
Features around the MP	V. Hoffstein, et al, <i>Eur. Respir. J.</i> , 5, 1992
	M. Okubo, et al, <i>Sleep</i> , 2006

SP: Sagittal Plane, MP: Mandibular Plane, Circum.: Circumference

## 2.5 Results

Figure 2.8 shows the correlation coefficients of  $L_1 - L_{48}$  with AHI. The correlation value is over 0.5 if  $i = 3, 4$  and  $i = 24, 25, \dots, 28$ . Each of these directions are, respectively, called A ( $i = 3, 4$ ) and B ( $i = 24, 25, \dots, 28$ ). In the region A, the maximum correlation is  $R = 0.526$  ( $p < 0.001$ ) when  $i = 3$ , while in the region B the maximum correlation is  $R = 0.601$  ( $p < 0.001$ ) when  $i = 26$ .

Also, the correlation of the sectional area of the upper airway ( $S$ ) with the severity AHI is calculated to  $R = -0.354$  ( $p = 0.02 < 0.05$ ). In table 2.4, the mean and standard deviation of the feature values ( $L_3, L_{26}, S$ ) in the mild to moderate OSA patients ( $\text{AHI} < 30$ ) and in the severe patients ( $\text{AHI} \geq 30$ ). According to this result, the values of  $L_3$  and  $L_{26}$  in the severe patients are relatively higher, and  $S$  tends to be lower than those of the other patients.

Table 2.6 shows the average, standard deviation, and correlation coefficient with AHI. In conventional features (upper four), the correlation coefficients with the severity are relatively good, since more than 0.4 value can be obtained. But in the case of  $L_3$  and  $L_{26}$  the coefficients are still better. These features are also

based on the tongue morphology, but the conventional methods are based on the volumetric features. Thus, it becomes clear that the tongue morphology has a more important role to estimate the severity of sleep apnea than the volumetric feature of the tongue. Since P-value of all features except for airway volume are less than 0.05, these correlation with AHI are sufficiently significant. Unexpectedly, the airway volume has no relation to the severity of OSAS at all.

Next, the most correlated features  $L_3$ ,  $L_{26}$  and  $S$  are adopted as the explanatory variables as the multi-regression equation in equation 2.4, and the author obtained the prediction equation as follows:

- (1) Using only morphological features of the tongue region ( $L_3, L_{26}$ ) as explanatory variables ( $Y_1, Y_2$ )

$$\text{AHI}' = 0.192 \times L_3 + 1.511 \times L_{26} - 93.866 \quad (R^2 = 0.375)$$

- (2) Using only cross sectional areas of the upper airway ( $S$ ) as an explanatory variable ( $Y_1$ )

$$\text{AHI}' = -0.093 \times S + 46.987 \quad (R^2 = 0.125)$$

- (3) Using both features ( $L_3, L_{26}, S$ ) as explanatory variables ( $Y_1, Y_2, Y_3$ )

$$\begin{aligned} \text{AHI}' = -0.052 \times L_3 + 1.593 \times L_{26} - 0.045 \times S - 94.117 \\ (R^2 = 0.398) \end{aligned}$$

The specificity, sensitivity, and efficiency of three multi-regression equations are shown in table 2.5. The sensitivity of the no.2 equation is higher than that of no.1 equation and the specificity and efficiency of no.3 equation are the best among them.

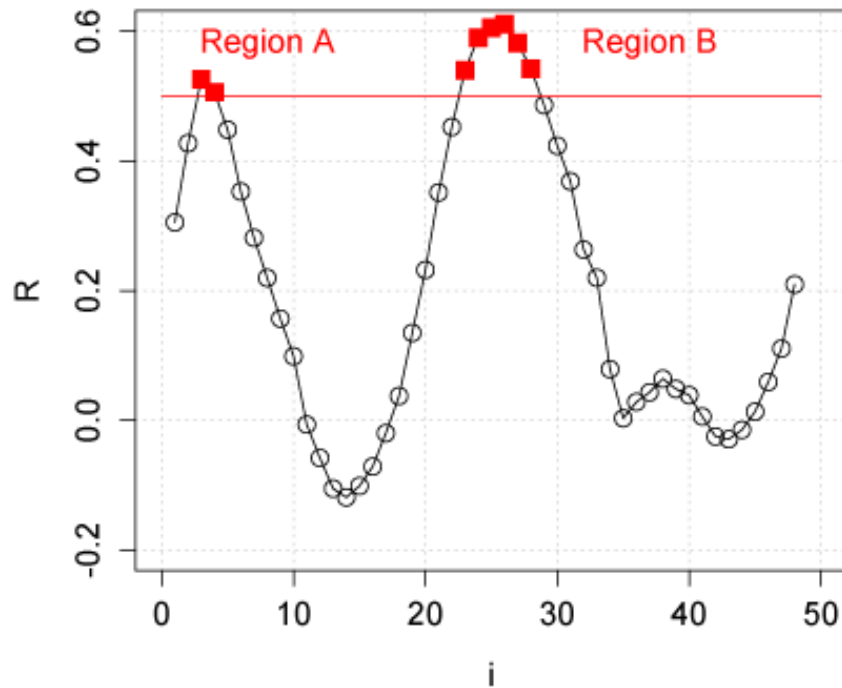


Figure 2.8: Correlation coefficients of  $L_i$  ( $i = 1, 2, \dots, 48$ ) with AHI (Left) and relationship between  $F_5$  and AHI (Right)

## 2.6 Discussion

### 2.6.1 Correlation with the Severity

The directions with which the correlation coefficients are over 0.5 can be separated into region A and B.

The region A ( $L_{24}, L_{27}$ ) is considered to indicate an enlargement of the back of the tongue, which is also known to be typical in OSA patients[3]. Under the present situation, a visual examination of the oral cavity is the most common method of assessing the enlargement. But according to table 2.7, MMP (modified Mallampati grade), which is determined by the visual examination of the oral cavity (See figure 2.10), is not so much correlated with AHI ( $R = 0.316$ ) as our feature is ( $R = 0.611$ ). This is because the inferior positioning of the soft palate

Table 2.4: The mean and standard deviation of feature values in patients with AHI<30 (from benign to moderate) and with AHI $\geq$  30 (serious)

Features	AHI< 30	AHI $\geq$ 30
Tongue ( $L_3$ ) [pixels]	96.06 ( $\pm$ 9.11)	103.51 ( $\pm$ 9.61)
Tongue ( $L_{26}$ ) [pixels]	86.21 ( $\pm$ 6.79)	92.56 ( $\pm$ 8.48)
SA ( $S$ ) [pixels]	124.33 ( $\pm$ 107.99)	53.14 ( $\pm$ 40.94)

Table 2.5: Sensitivity, specificity, and efficiency of multi-regression equations with the cut-off AHI determined to 30

Equ.	Sensitivity	Specificity	Efficiency
(1)	0.864	0.429	0.651
(2)	<b>0.955</b>	0.238	0.605
(3)	0.909	<b>0.524</b>	<b>0.721</b>

is categorized to one of the four grades by the visual examination qualitatively, while in our method the inferior positioning can be evaluated more precisely and quantitatively than MMP’s four grades.

The region B ( $L_5, L_6$ ) is nearly diagonal to the region A ( $L_{24} - L_{28}$ ) and slightly toward the airway. According to figure 2.2, the hyoid bone exists in this direction. Kitamura [2] and Yucel [6] indicated that the lower positioning of the hyoid bone is associated with OSA using cephalometric techniques. The inferior positioning makes the tongue elongated.

In addition, since cephalometry does not provide any information of the tongue enlargement, our method has an advantage in comparison with conventional methods. Table 2.7 shows some morphological features of the soft tissue in the upper airway (but BMI is not) and the correlation coefficients with AHI, which have been reported in conventional studies[2][4][5][6]. Naturally, it is difficult to simply compare these results with each other, because the number of subjects and subjects’ AHI are different among these reports. But, in so far as this table indicates, we can say that our result is far better than the other features ( $R = 0.611$ ) and significant



Table 2.6: Average, standard deviation, and correlation coefficient with AHI of the proposed and the conventional features both estimated from the upper airway MR images

Feature	AHI<30	AHI≥30	Corr. Coeff.	P-value
TA in SP [pixels]	168389(±14670)	174547(±21950)	0.441	0.01
TC in SP [pixels]	441(±21)	454(±20)	0.446	<0.01
Neck Circum. [cm]	41.6(±2.5)	44.1(±4.8)	0.482	<0.01
Airway Vol. [cm <sup>3</sup> ]	6.82(±2.73)	7.60(±3.05)	0.098	0.58
$L_3$ [pixels]	96.1(±9.1)	103.5(±9.6)	0.526	<0.01
$L_{26}$ [pixels]	86.2(±6.8)	92.6(±8.5)	0.611	<0.01
$S$ [pixels]	124.3(±108.0)	53.1(±40.9)	-0.354	0.02

TA: Tongue Area, SP: Sagittal Plane, TC: Tongue Circumference

as well ( $p < 0.001$ ).

### 2.6.2 Classification Ability of the Severe OSAS

In the case of equation (1) using only the tongue features, the sensitivity is 0.86 at most, but the specificity is 0.43. Otherwise, the sensitivity reaches 0.955 in the case of equation (2) using the cross sectional area, but the sensitivity is relatively low: 0.238. Accordingly, the false positive ratio tends to be high when using the cross sectional area. In table 2.4, the standard deviation of the cross sectional area in patients with AHI<30 is higher than that in patients with AHI≥30. Many of severe OSAS patients generally tend to have narrower airway and the results here also support it, but it is also found that the eight moderate patients have also narrower airway. This is the reason why the sensitivity becomes low using equation (2).

Also, the sensitivity and specificity of equation (3) are relatively better than those of equation (1) and (2). In this study, the cross sectional area is estimated from the transverse plane around the uvula. This area becomes narrower due to the enlargement of the tongue, soft palate, and tonsils. Especially, tonsils cannot be photographed in the sagittal plane. This is because the sensitivity becomes

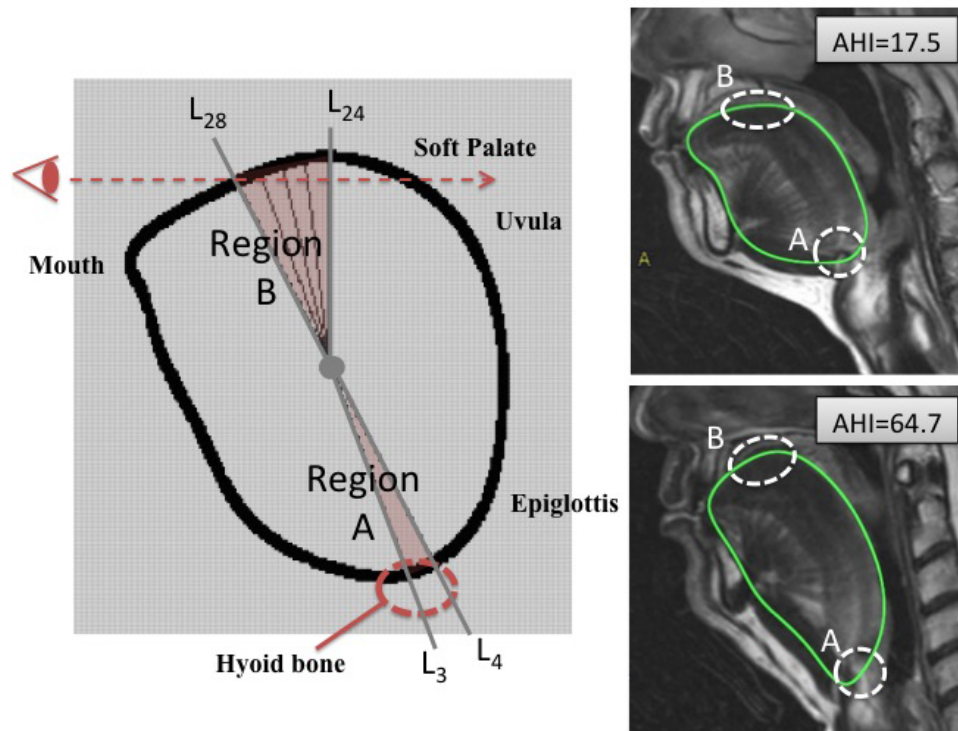


Figure 2.9: Directions whose correlation is higher than 0.5 (region A and B) and the corresponding regions in MR images of moderate and severe OSAS patients.



Figure 2.10: Modified Mallampati Grade; Class I (soft palate, uvula, faces, pillars visible), Class II (soft palate, uvula, faces visible), Class III (soft palate, base of uvula visible), Class IV (only hard palate visible), (cited from Wikimedia Commons, [http:// commons.wikimedia.org/ wiki/ File:Mallampati.svg](http://commons.wikimedia.org/wiki/File:Mallampati.svg))

Table 2.7: Correlation between AHI and Morphologic Properties in Conventional Studies

Feature	Method	Corr. Coef.	$p$ -value	Literature
PNS-P	Cephalogram	0.194	0.011	Kitamura, et al [2]
		0.240		Yu, et al [4]
MP-H	Cephalogram	0.449	< 0.001	Kitamura, et al [2]
		0.310	< 0.01	Sakakibara, et al [5]
		0.420	< 0.001	Yucel, et al [6]
TGL	Cephalogram	0.320	< 0.01	Yu, et al [4]
		0.213	< 0.05	Sakakibara, et al [5]
H-VL	Cephalogram	0.430	< 0.01	Yu, et al [4]
		0.425	< 0.001	Sakakibara, et al [5]
BMI	Other	0.310	< 0.001	Kitamura, et al [2]
		0.525		Sakakibara, et al [5]
MMP	Visual Exam.	0.316	< 0.001	Kitamura, et al [2]
Tonsil Size	Visual Exam.	0.212	0.005	Kitamura, et al [2]
Tongue Area in SP	2D MRI	0.441	< 0.01	Herrman, et al [25]
Tongue Circum.		0.446	< 0.01	Herrman, et al [25]
Neck Circum.		0.482	< 0.01	Hoffstein, et al[26]
Airway Vol.	3D MRI	0.098	0.58	Iida-Kondo, et al[10]
$L_3$	2D MRI	<b>0.526</b>	< 0.001	This study
$L_{26}$		<b>0.611</b>	< 0.001	This study
SA		<b>-0.354</b>	0.02	This study

better by using the cross sectional area estimated from the transverse plane.

## 2.7 Conclusion

The direction to the back of tongue from the center of gravity is the most correlated with AHI. In addition, the diagonal direction to the hyoid bone is also correlated. The length of these two directions are useful information about the severity of sleep apnea, and the correlation coefficient with the severity is higher than the other morphological features proposed in conventional studies. Thus, it is useful to focus on our proposed features to evaluate the severity of sleep apnea.

Multi-regression analysis using those tongue morphological features and the

cross sectional area in the narrowest upper airway has been done to predict the severity of OSAS. By thresholding the predicted severity, severe OSAS can be detected with 0.909 sensitivity and 0.524 specificity.

In the future, the other morphological features such as soft palate, epiglottis, or tonsils should be focused on to boost up the sensitivity. Moreover, 3-dimensional morphological features of the soft tissue are also promising to obtain useful information about the severity of OSA.

This study was supported in part by the Grant-in-Aid for General Research Program from the Akiyama Life Science Foundation.

## References

- [1] H. W. Agnew Jr., Wilse B. Webb, and Robert L. Williams. The first night effect: An EEG study of sleep. *Psychophysiology*, Vol. 2, No. 3, pp. 263–266, 1996.
- [2] T. Kitamura, A. Sakabe, N. Ueda, T. Shiomori, T. Udaka, T. Ohbuchi, and H. Suzuki. Usefulness of cephalometry and pharyngeal findings in the primary diagnosis of obstructive sleep apnea syndrome. *Nippon Jibiinkoka Gakkai Kaiho*, Vol. 111, No. 11, pp. 695–700, 2008.
- [3] The Research Group of Sleep-Disordered Breathing, editor. *Guideline for Diagnosis and Treatment of Adult SAS*. Medical Review, 2008.
- [4] X. Yu, K. Fujimoto, K. Urushibata, Y. Matsuzawa, and K. Kubo. Cephalometric analysis in obese and nonobese patients with obstructive sleep apnea syndrome. *Chest*, Vol. 124, pp. 212–218, 2003.
- [5] H. Sakakibara, M. Tong, K. Matsushita, M. Hirata, Y. Konishi, and S. Suet-sugu. Cephalometric abnormalities in non-obese and obese patients with obstructive sleep apnoea. *European Respiratory Journal*, Vol. 13, pp. 403–410, 1999.
- [6] A. Yucel, M. Unlu, A. Haktanir, M. Acar, and F. Fidan. Evaluation of the upper airway cross-sectional area changes in different degrees of severity of obstructive sleep apnea syndrome: Cephalometric and dynamic ct study. *American Journal of Neuroradiology*, Vol. 26, pp. 2624–2629, 2005.

- 
- [7] T. Mochizuki. Examination of low palatal arch with long low-hanging soft palate in obstructive sleep apnea syndrome and cephalometry. *Nippon Jibi-inkoka Gakkai Kaiho*, Vol. 100, pp. 937–945, 1997.
- [8] K. A. Ferguson, T. Ono, A. A. Lowe, C. F. Ryan, and J. A. Fleetham. The relationship between obesity and craniofacial structure in obstructive sleep apnea. *Chest*, Vol. 108, pp. 375–385, 1995.
- [9] M. Partinen, C. Guilleminault, M-A. Quera-Salva, and A. Jamieson. Obstructive sleep apnea and cephalometric roentgenograms, the role of anatomic upper airway abnormalities in the definition of abnormal breathing during sleep. *Chest*, Vol. 93, pp. 1199–1205, 1988.
- [10] C. Iida-Kondo, N. Yoshino, T. Kurabayashi, S. Mataka, M. Hasegawa, and N. Kurosaki. Comparison of tongue volume/oral cavity volume ratio between obstructive sleep apnea syndrome patients and normal adults using magnetic resonance imaging. *Journal of Medical and Dental Science*, Vol. 53, pp. 119–126, 2006.
- [11] R. Schwab, M. Psirstein, R. Pierson, A. Mackley, R. Hachadoorian, R. Arenas, G. Maislin, and A. Pack. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. *American Journal of Respiratory and Critical Care of Medicine*, Vol. 168, pp. 522–530, 2003.
- [12] R. Arens, J. M. McDonough, A. M. Corbin, M. E. Hernandez, G. Maislin, R. I. Schwab, and A. I. Pack. Linear dimensions of the upper airway structure during development, assessment by magnetic resonance imaging. *American Journal of Respiratory and Critical Care Medicine*, Vol. 165, pp. 117–122, 2002.
- [13] I. L. Mortimore, I. Marshall, P. K. Wraith, R. J. Sellar, and N. Douglas. Neck and total body fat deposition in nonobese and obese patients with sleep apnea

- compared with that in control subjects. *American Journal of Respiratory and Critical Care Medicine*, Vol. 157, pp. 280–283, 1998.
- [14] D. Rodenstein, G. Dooms, Y. Thomas, G. Liistro, D. C. Stanescu, C. Culee, and G. Aubert-Tulkens. Pharyngeal shape and dimensions in healthy subjects, snorers, and patients with obstructive sleep apnea. *Thorax*, Vol. 45, pp. 722–727, 1990.
- [15] K. C. Welch, G. D. Foster, C. T. Ritter, T. A. Wadden, R. Arens, G. Maislin, and R. J. Schwab. A novel volumetric magnetic resonance imaging paradigm to study upper airway anatomy. *Sleep*, Vol. 25, No. 5, pp. 532–542, 2002.
- [16] Sleep-related breathing disorders in adults : recommendations for syndrome definition and measurement techniques in clinical research, 1999.
- [17] M. S. Atkins and B. T. Mackiewicz. Fully automatic segmentation of the brain in MRI. *IEEE Trans. Medical Imaging*, Vol. 17, No. 1, pp. 98–107, 1998.
- [18] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, Vol. 33, pp. 115–126, 2006.
- [19] T. Heimann, B. van Ginneken, M. A. Styner, Y. Arzhaeva, and et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Medical Imaging*, Vol. 28, No. 8, pp. 1251–1265, 2009.
- [20] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Medical Imaging*, Vol. 19, No. 3, pp. 203–210, 2000.
- [21] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, and et al. CURVES: Curve evolution for vessel segmentation. *Medical Image Analysis*, Vol. 5, No. 3, pp. 195–206, 2001.

- [22] S. Nakagawa. Pattern information processing. *Maruzen Pub*, 1999.
- [23] 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. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 12, pp. 1553–1562, 2008.
- [24] M. Okubo, M. Suzuki, A. Hirouchi, S. Okabe, K. Ikeda, and et al. Morphologic analysis of mandible and upper airway soft tissue by mri of patients with obstructive sleep apnea hypopnea syndrome. *Sleep*, Vol. 29, pp. 909–915, 2006.
- [25] B. L. Herrmann, T. E. Wessendorf, W. Ajaj, S. Kahlke, H. Teschler, and K. Mann. Effects of octreotide on sleep apnoea and tongue volume (magnetic resonance imaging) in patients with acromegaly. *European Journal of Endocrinology*, Vol. 151, pp. 309–315, 2004.
- [26] V. Hoffstein and S. Mateika. Differences in abdominal and neck circumferences in patients with and without obstructive sleep apnoea. *European Respiratory Journal*, Vol. 5, pp. 377–381, 1992.



## Chapter 3

# Acoustic Analysis and Classification of Oral and Nasal Snoring Sounds

### 3.1 Introduction

In this chapter, the author proposes a novel method to classify oral/nasal snores 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 Syndrome (OSAS)[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 OSAS 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 OSAS patients.

On the other hand, Herzog, et al, [9] showed that patients with primary snoring revealed peak intensities between 100 and 300 Hz, whereas OSAS 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 OSAS 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 OSAS 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 OSAS. 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 OSAS patients than in simple (benign) snorers.

These 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. The author's 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 OSAS occurrence for normal persons. Although open mouth and the corresponding loud snoring are highly related to OSAS, few researchers have focused on the acoustic properties of oral and nasal snoring sounds and their classification. Dalmasso, et al[14], 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[15], reported that the frequency of airflow and supraglottic pressure oscillations were less during mouth than during nasal simulated snoring. These studies are informative, but no concrete technical methods have been established to classify oral and nasal snoring sounds.

In this chapter, the author firstly analyzes the acoustic properties of simulated snoring sounds with oral and nasal breath and proposes a concrete method for their classification.

## **3.2 Acquisition of Classification Targets**

### **3.2.1 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 3.1 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.

While recording, the subjects are asked to simulate snoring by breathing deeply

Table 3.1: Detail of all subjects and the obtained snoring episodes and subsequences in this study

Subject	Age	# of episodes	# of subseq.	Apnea/Benign
A	32	n:9, o:9	n:69, o:66	Benign
B	55	n:8, o:9	n:110, o:109	Benign
C	63	n:7, o:8	n:47, o:80	Apnea
D	52	n:6, o:8	n:21, o:43	Apnea
E	21	n:8, o:8	n:156, o:128	Apnea
F	35	n:0, o:9	n:0, o:101	Benign
G	46	n:8, o:8	n:94, o:131	Apnea
H	40	n:9, o:9	n:95, o:161	Benign
I	67	n:9, o:9	n:122, o:81	Benign
J	63	n:9, o:9	n:90, o:127	Benign
K	62	n:11, o:8	n:57, o:39	Apnea
L	56	n:9, o:8	n:99, o:98	Benign
M	43	n:8, o:7	n:47, o:120	Benign
N	38	n:8, o:9	n:48, o:55	Benign
O	22	n:9, o:7	n:52, o:45	Benign

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 or physiological studies[15][16][17][18][19]. Similarities between simulated and natural (nocturnal) snoring are described in chapter 1 from the bibliographical point of view.

### 3.2.2 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 (figure 3.2), 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



Figure 3.1: A portable linear PCM recorder for recording snoring sounds (Olympus LS-11)

definition of snoring sounds should be considered.

According to some medical researchers[20][21], 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,[22] 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 [19], because they depend highly on the anatomical site of snoring, such as soft palate, tongue, epiglottis, tonsils, and the mixture of them [23]. Therefore, it is not easy to cut out the snoring episodes automatically[13]. In fact, many automatic extraction methods of snoring episodes have been proposed so far[24][25][26][27]. For instance, Cavusoglu, et al, [24] have tried to classify snores and non-snores according to their sub-band energy distributions in the frequency domain. Karunajeewa, et al, [25] have focused on more various features - number

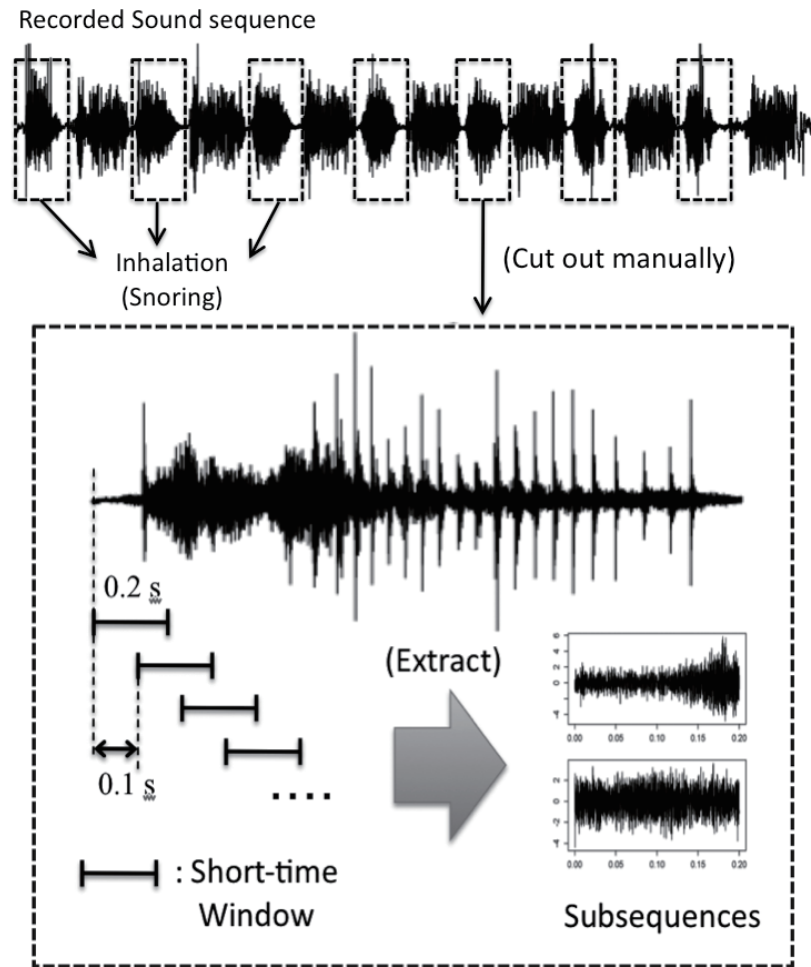


Figure 3.2: An extraction method of episodes and subsequences of snoring sounds from a recorded sound.

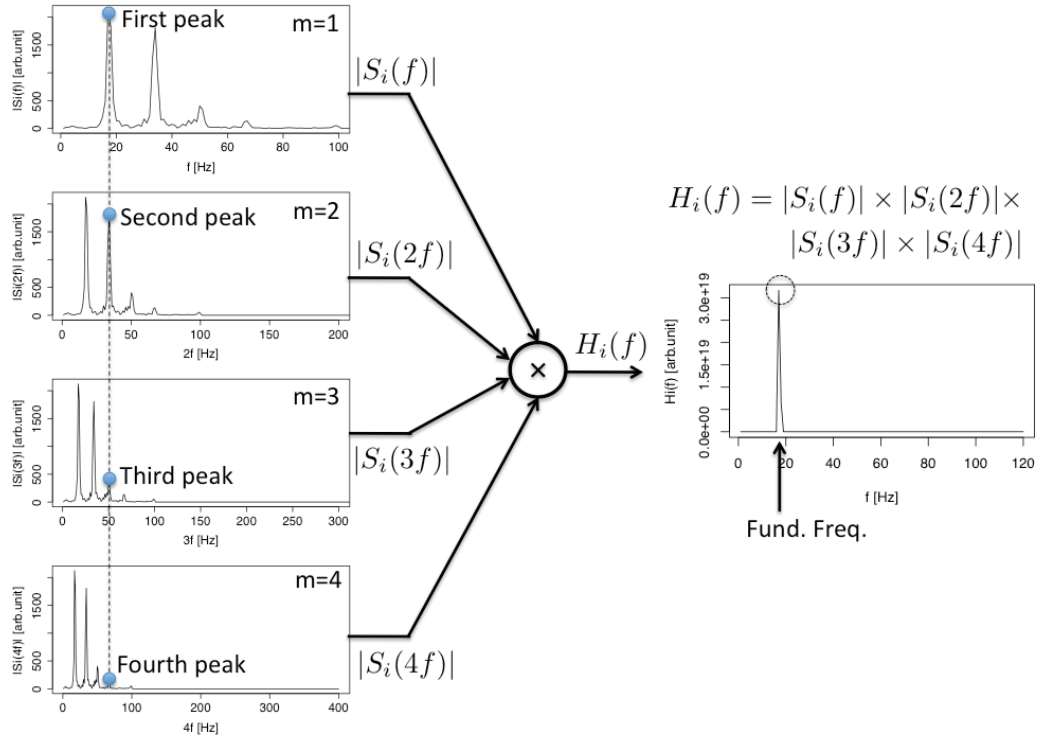


Figure 3.3: An example of the harmonic product spectrum calculated with  $r = 4$ .

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 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, ordinary people can empirically distinguish snoring sounds from simple breathing noises[13]. So, the author 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 the reference[13].

### 3.2.3 Subsequence Extraction

Since the acoustic properties of snoring sounds are nonstationarily changing as time passes even in a snoring episode[14][28], the author 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  $E[\cdot]$  is an expectation operator. As a result, 1384 oral and 1107 nasal subsequences are obtained from all subjects, and they are used as the classification targets.

## 3.3 Classification Method

### 3.3.1 Feature Extraction

#### Fundamental Frequency

Figures 3.4 and 3.5 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 [15] 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 3.4 and 3.5, the author adopts the fundamental frequency of snoring sounds as the first acoustic property for the classification. Many pitch detection algorithms



have been proposed so far, but the author used in this paper *Harmonic Product Spectrum* (HPS) method[29]. 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 3.4 and 3.5, 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:

$$H_i(f) = \prod_{m=1}^r |S_i(mf)| \quad (3.1)$$

where  $r$  is the number of harmonic peaks,  $|S_i(mf)|$  is the amplitude spectrum, and  $m$  is a scaling parameter. Figure 3.3 shows an example of HPS calculation when  $r$  is set to four. In this case, the HPS is calculated by  $H_i(f) = |S_i(f)| \cdot |S_i(2f)| \cdot |S_i(3f)| \cdot |S_i(4f)|$ . 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 [30][23]. 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 (figure 3.3). Accordingly, we can obtain the fundamental frequency  $f_b$  by using the following criterion:

$$H_i(f_b) = \max H_i(f) \quad (3.2)$$

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

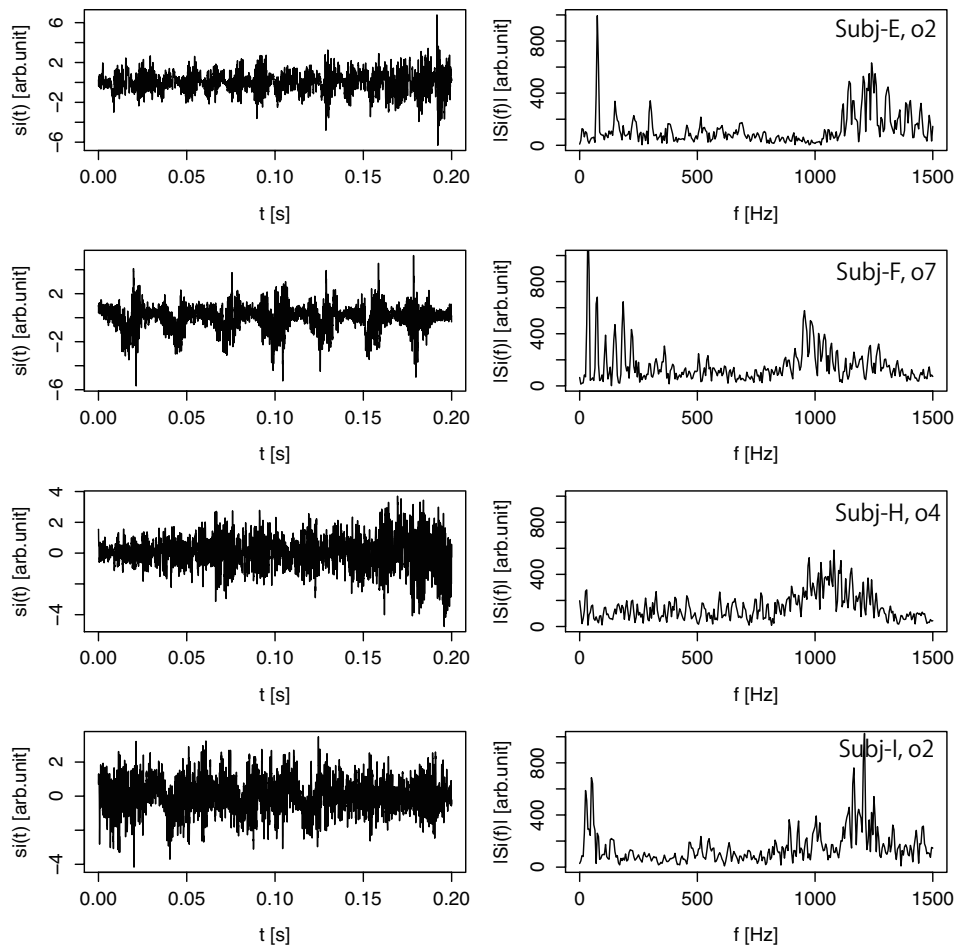


Figure 3.4: Subsequences of oral snoring and their amplitude spectra.

### The Maximum of the Amplitude Spectrum in a Specific Band

According to figures 3.4 and 3.5, there are some intensity peaks in a specific band over 1kHz, whereas no such peaks exist in nasal snoring sounds. Agrawal, et al, [23] 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 the results of oral snores in this chapter. In addition, according

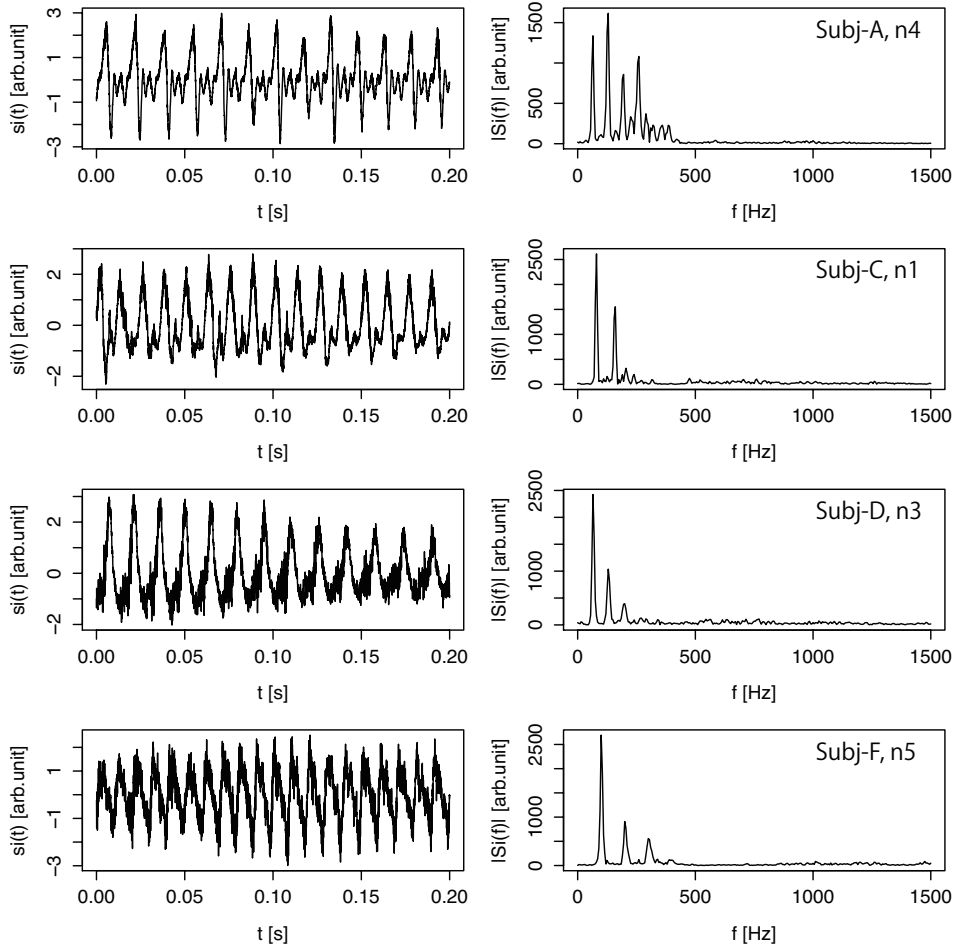


Figure 3.5: Subsequences of nasal snoring and their amplitude spectra.

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:

$$M = \max_{f_1 \leq f \leq f_2} |S_i(f)| \quad (3.3)$$

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.

### 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  $\mathbf{x}_i = (x_{i1}, x_{i2}) = (M, f_b)$  where  $M$  and  $f_b$  are estimated from the  $i$ th subsequence. But it is necessary to adjust some parameters,  $r$  in Eq.(3.1) and  $f_1, f_2$  in Eq.(3.3). In this paper, the ratio of within-class variance and between-class variance (called the variance ratio[31][32]) is used for an evaluation criterion for parameter adjustment. For a two-class problem, the variance ratio is defined as

$$\gamma = \frac{(m_1 + m_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (3.4)$$

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.

### 3.3.2 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 Net-

works, Decision Tree, Support Vector Machines, and so on[33][32]. 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[31]. 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. This is analyzed in the next chapter.

### 3.4 Performance Evaluation

#### 3.4.1 $m$ -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  $\mathbf{x}_i (i = 1, 2, \dots, N)$  into  $m$  groups  $G_1, G_2, \dots, G_m$ .
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  $R = \sum_{j=1}^m |G_j| r_j / N$  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[31]. But, three more different ways of dividing all the data into groups are considered in the next section in order to evaluate the usefulness of our method more objectively.

### 3.4.2 Leave-One-Out Test

Next the author 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.

### 3.4.3 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, the author defined the number of groups as the number of episodes and assigned the data extracted from the same episode into the same group. This evaluation method is in this study called *Leave-Episode-Out* (LEO) test.

### 3.4.4 Leave-*Subject*-Out Test

An individual difference may be what the author 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 this cannot be evaluated using 10-fold CV, LOO, or LEO test. Accordingly, the author 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. This evaluation method is called *Leave-Subject-Out* (LSO) test in this study.

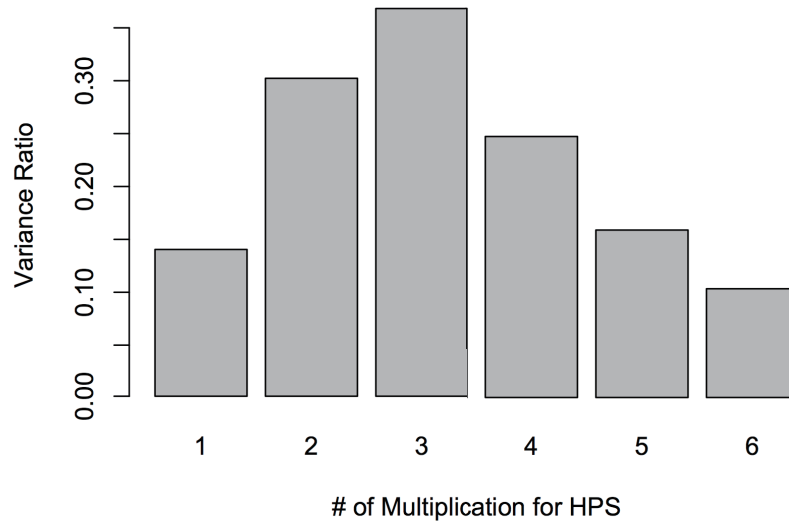


Figure 3.6: The variance ratio of the fundamental frequency calculated with the number of multiplication,  $r$ , being set to 1 to 6

## 3.5 Results and Discussion

### 3.5.1 Optimal Value of $r$ and its Relation to the Acoustic Properties

Firstly, the variance ratio of the fundamental frequency is calculated when the number of multiplication ( $r$ ) for the HPS method is adjusted to from 1 to 6 respectively. Figure 3.6 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* [19].

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 3.7, where the maximum peak does not lie at the fundamental frequency

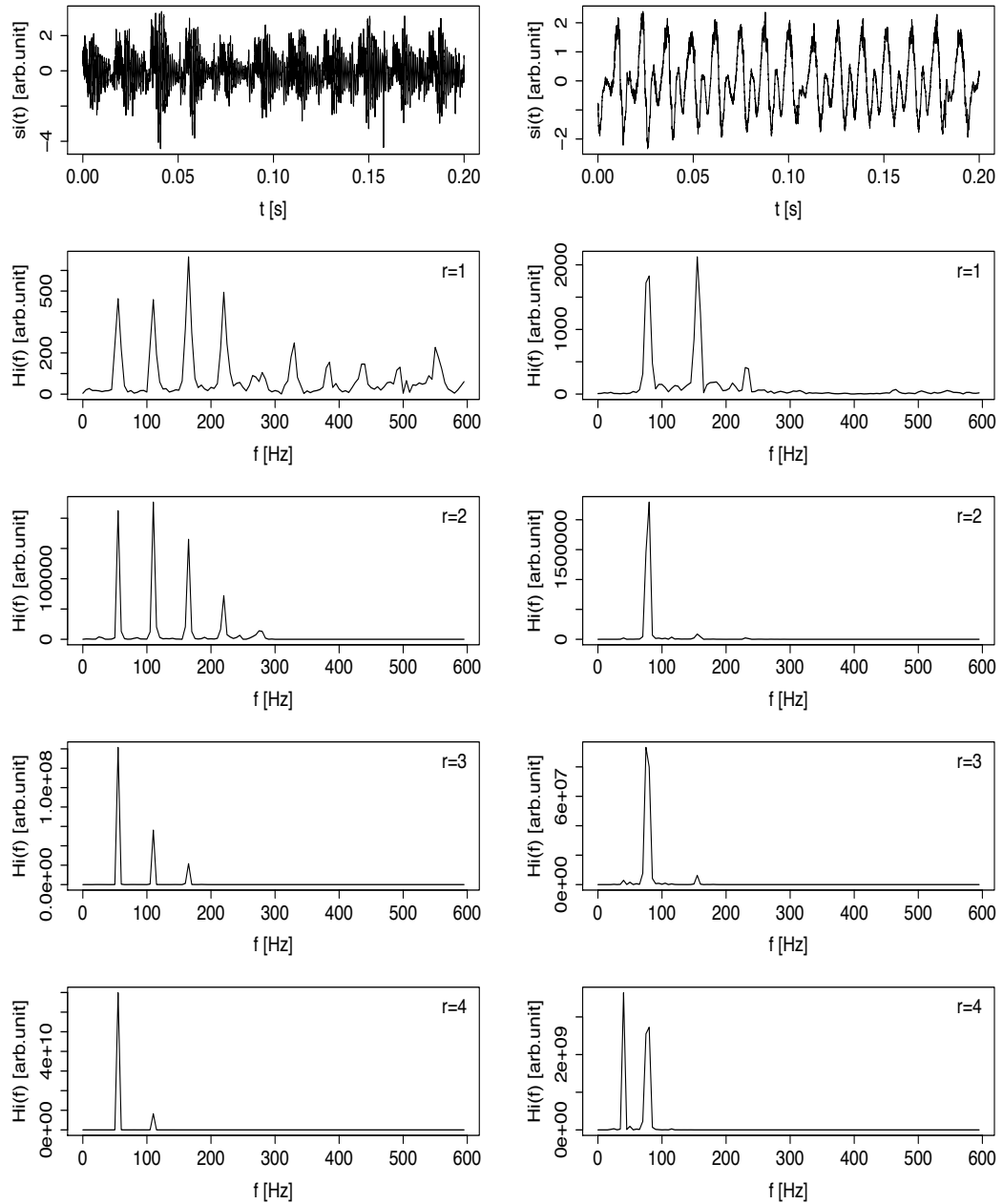


Figure 3.7: Two examples of incorrect estimation of the fundamental frequency when  $r$  is determined to less than 3 (right) and greater than 3 (left)



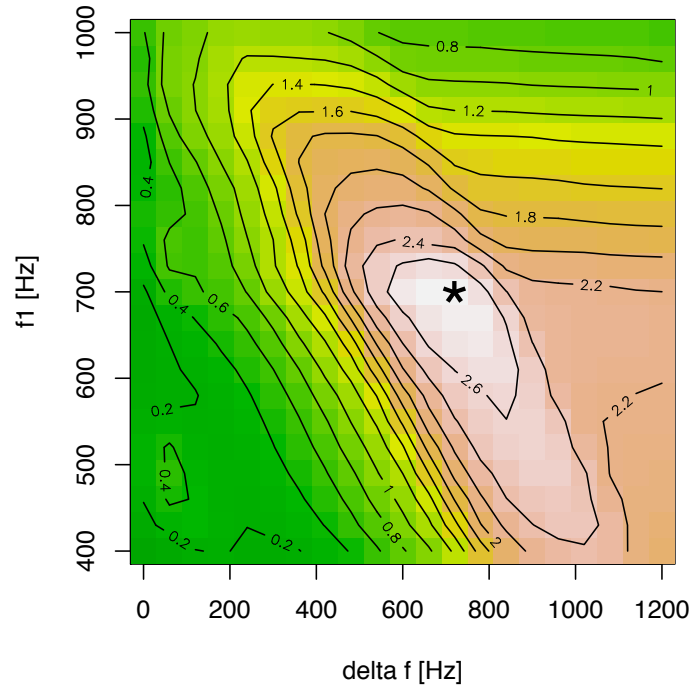


Figure 3.8: Contour plots of classification rate obtained from various  $f_1$  and  $\Delta f$  under the LSO test

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  $r = 4$ .

### 3.5.2 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.(3.3). The left panel in figure 3.8 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 3.4, 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, the author discusses what kind of phenomenon such intensity peak reflects. According to Agrawal and coworkers' report[23], 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 [30][34]. Agrawal [23] 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[30][35].

In comparison with the results in this chapter, 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).

### 3.5.3 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 3.9 shows the classification rate when  $k$  is set to from 1 to 50 with 2 steps. The classification rate is largely converged to

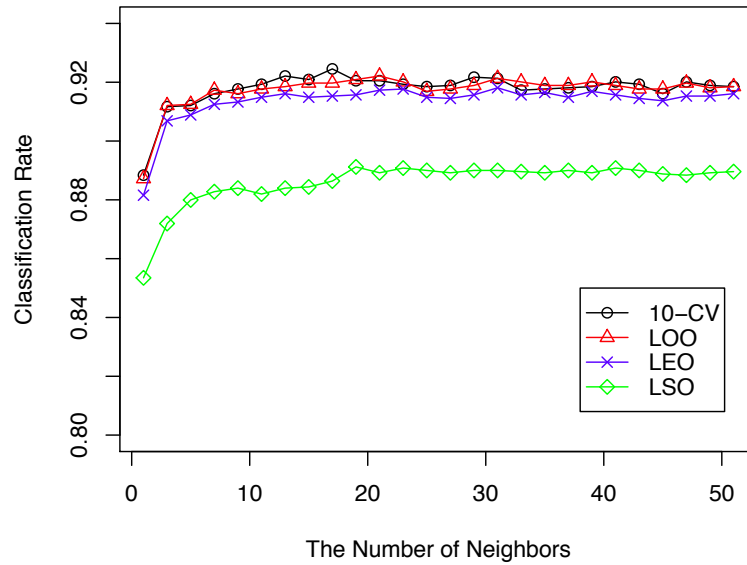


Figure 3.9: Classification rates with the number of neighbors,  $k$ , set to 1 to 50 with 2 steps under the four cross-validation tests .

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 3.10 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 3.10, 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 3.2 shows the detail result of the classification with  $k = 19$ , where the

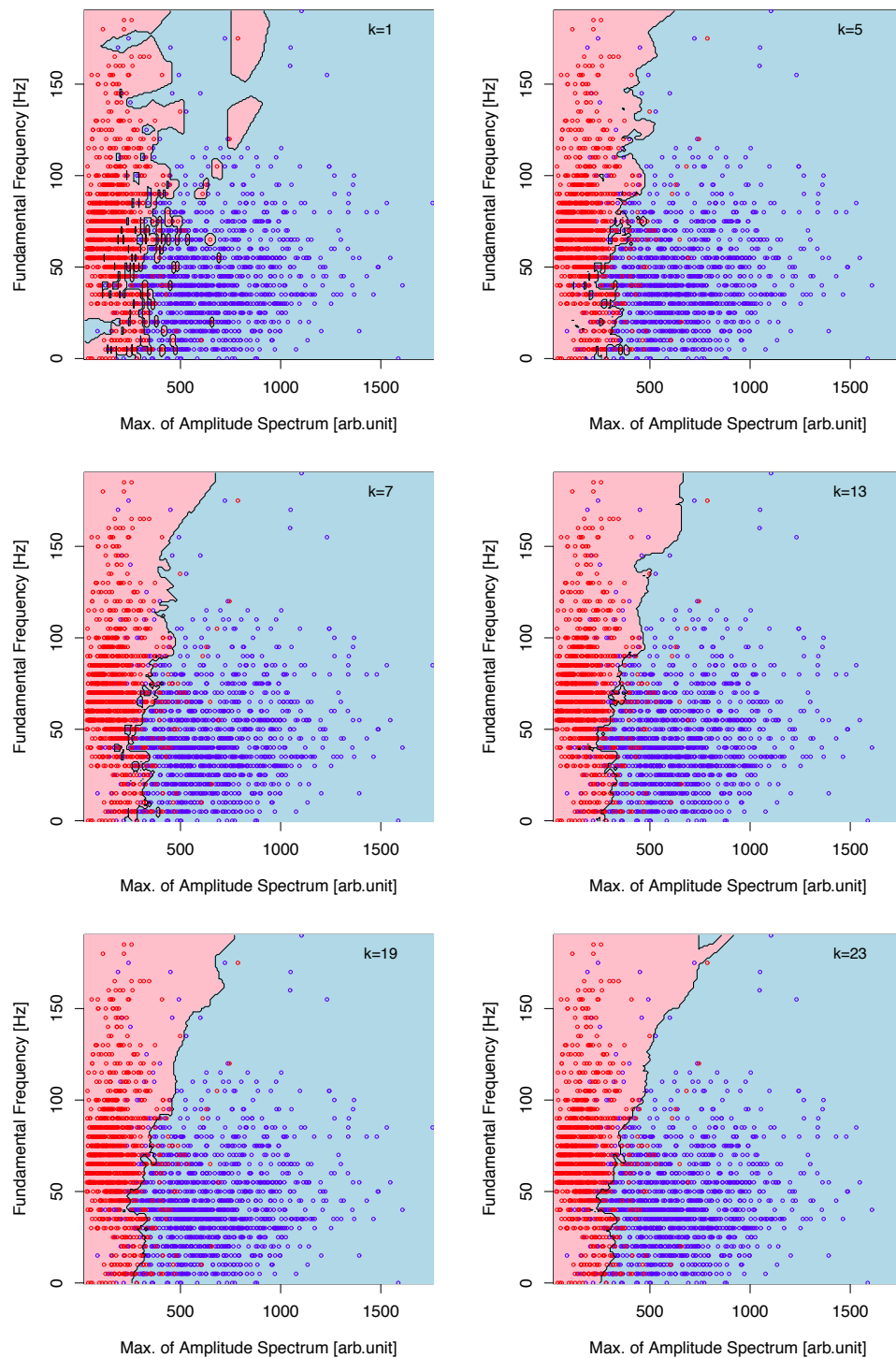


Figure 3.10: Scatter plots of nasal (red) and oral (blue) snoring sounds and the classification boundary estimated by kNN classifier with  $k = 1, 5, 7, 13, 19,$  and  $23$

Table 3.2: Classification rate with the optimal parameters ( $r = 3, f_1 = 700, f_2 = 1420, k = 19$ ) under the four different evaluation methods

Test	Classification Rate		
	Nasal (%)	Oral (%)	Total (%)
10-fold CV	1014/1107 (91.60%)	1289/1384 (93.14%)	2303/2491 (92.45%)
LOO	1015/1107 (91.69%)	1283/1384 (92.70%)	2298/2491 (92.25%)
LEO	1012/1107 (91.42%)	1281/1384 (92.56%)	2293/2491 (92.05%)
LSO	964/1107 (87.08%)	1268/1384 (91.62%)	2232/2491 (89.60%)

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[31]. 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.

#### 3.5.4 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 3.1, 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.

Figure 3.11 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 figure 3.5, whereas the other panels are similar to the oral snores in figure 3.4. In addition, figure 3.12 shows scatter plots of oronasal snores on the 2-dimensional

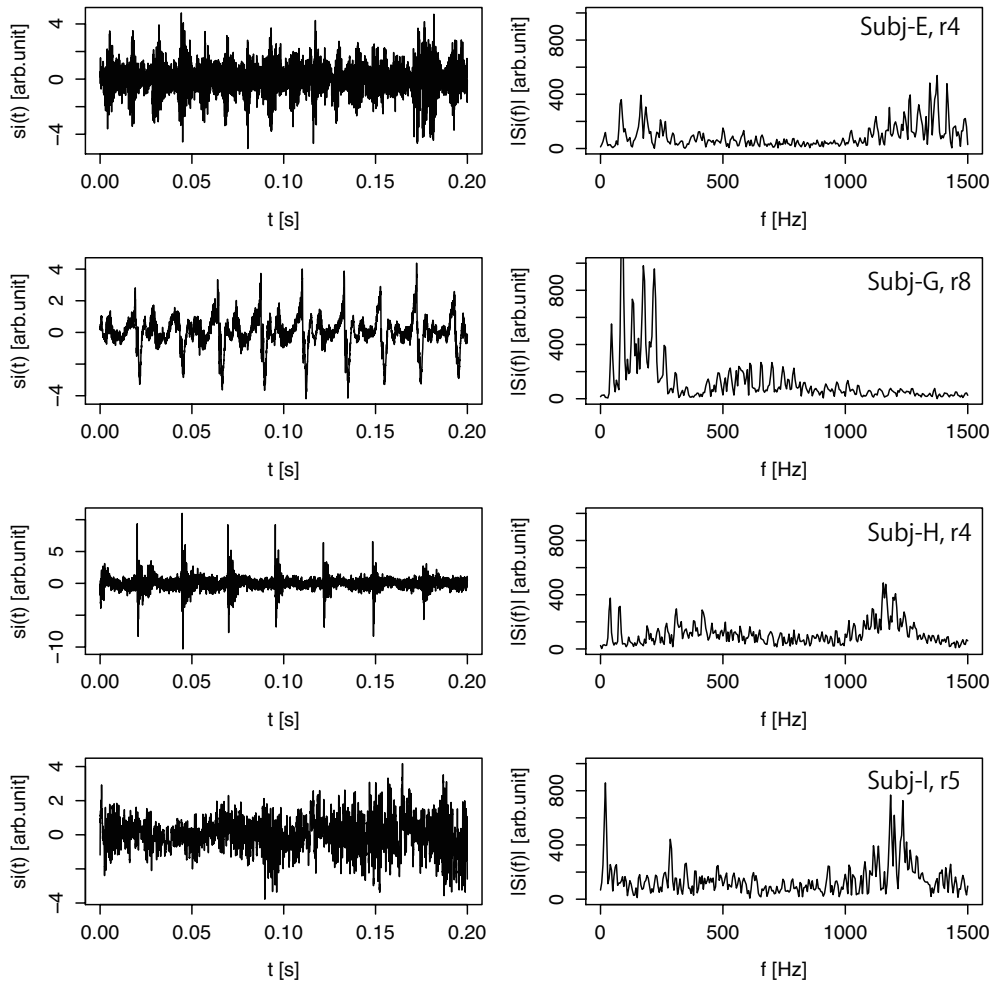


Figure 3.11: Waveforms and amplitude spectra of oronasal simulated snores

feature space. Compared with figure 3.10, 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.

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

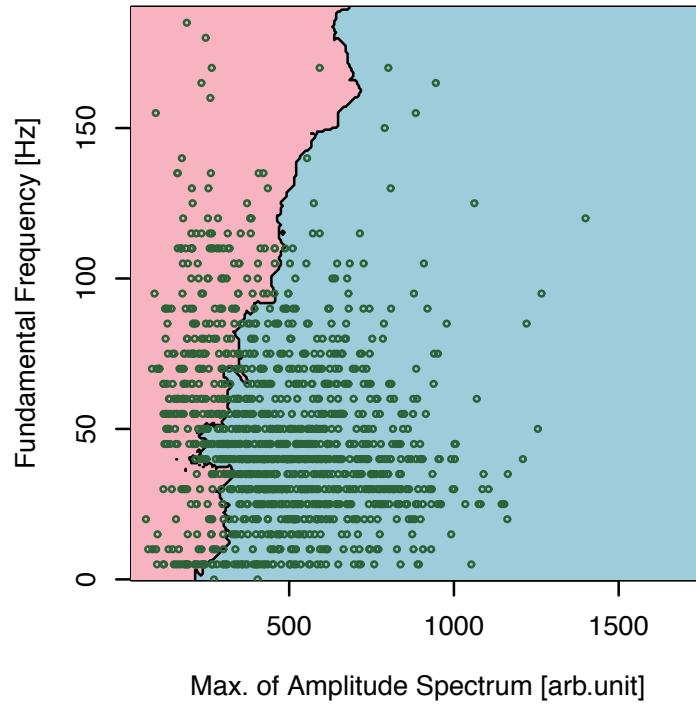


Figure 3.12: Scatter plots of oronasal simulated snores on the 2-dimensional feature space

every subject under the LSO test (shown in table 3.3). 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. Dalmaso, et al, described that the shape of cross sectional area in the upper airway are very similar during oral snoring and during free snoring (ornasal route)[14]. 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.

Table 3.3: Classification results of oronasal simulated snores

Subjects	# of Subseq.	Oral (%)	Nasal (%)	Majority
A	32	<b>31 (96.9%)</b>	1 (3.1%)	Oral
B	102	<b>88 (86.3%)</b>	14 (13.7%)	Oral
C	60	<b>60 (100%)</b>	0 (0.0%)	Oral
D	55	<b>53 (96.7%)</b>	2 (3.3%)	Oral
E	65	<b>35 (53.8%)</b>	30 (46.2%)	Oral
F	82	<b>82 (100%)</b>	0 (0.0%)	Oral
G	120	<b>90 (75.0%)</b>	30 (25.0%)	Oral
H	117	<b>111 (94.9%)</b>	6 (5.1%)	Oral
I	51	<b>51 (100%)</b>	0 (0.0%)	Oral
J	105	24 (22.9%)	<b>81 (77.1%)</b>	Nasal
K	65	<b>51 (78.5%)</b>	14 (21.5%)	Oral
L	85	<b>76 (89.4%)</b>	9 (10.6%)	Oral
M	110	<b>100 (90.9%)</b>	10 (9.1%)	Oral
L	80	<b>70 (87.5%)</b>	10 (12.5%)	Oral
O	100	<b>76 (76.0%)</b>	24 (24.0%)	Oral
Total	1229	998(81.2%)	231(18.8%)	–

### 3.6 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 OSAS, 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.

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

- [1] M. Ichioka. Respiratory organs and sleep disorders. *Folia Pharmacologica Japonica (in Japanese)*, Vol. 129, pp. 432–435, 2007.
- [2] H. A. McLean, A. M. Urton, H. S. Driver, A. K. W. Tan, A. G. Day, P. W. Munt, and M. F. Fitzpatrick. Effect of treating severe nasal obstruction on the severity of obstructive sleep apnoea. *European Respiratory Journal*, Vol. 25, pp. 521–527, 2005.
- [3] V. Taasan, J. W. Wynne, N. Cassisi, and A. J. Block. The effect of nasal packing on sleep-disordered breathing and nocturnal oxygen desaturation. *Laryngoscope*, Vol. 91, pp. 1163–1172, 1982.
- [4] C. W. Zwillich, C. Pickett, F. N. Hanson, and J. V. Weil. Disturbed sleep and prolonged apnea during nasal obstruction in normal men. *American Review of Respiratory Disease*, Vol. 124, pp. 158–160, 1981.
- [5] A. J. Mousavi, S. M. Fereshtehnejad, N. Khalili, M. Naghavi, and H. Yahyazadeh. Determination of the factors affecting duration of hospitalization in patients with chronic obstructive pulmonary disease (COPD) in Iran. *Medical Journal of the Islamic Republic of Iran*, Vol. 22, No. 1, pp. 29–36, 2008.
- [6] H. Hara, Y. Miyauchi, J. Matsumoto, and Y. Yamashita. Morphological change of the upper airway in OSAS patients with open mouth. *Stomatopharyngology (in Japanese)*, Vol. 20, No. 1, p. 45, 2007.

- 
- [7] J. R. Perez-Padilla, E. Slawinski, L. M. Difrancesco, R. Feige, J. E. Remmers, and W. A. Whitelaw. Characteristics of the snoring noise in patients with and without occlusive sleep apnea. *American Review of Respiratory Disorders*, Vol. 147, No. 3, pp. 635–644, 1993.
- [8] J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mananas, and P. Caminal. Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. *European Respiratory Journal*, Vol. 9, No. 11, pp. 2365–2370, 1996.
- [9] M. Herzog, A. Schmidt, T. Bremert, B. Herzog, W. Hosemann, and H. Kaftan. Analysed snoring sounds correlate to obstructive sleep disordered breathing. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 1, pp. 105–113, 2008.
- [10] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita. Acoustic analysis of snoring sounds by a multidimensional voice program. *Laryngoscope*, Vol. 116, No. 3, pp. 379–381, 2006.
- [11] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. *Proc. of Annual Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, Vol. 3, pp. 2527–2530, 2003.
- [12] A. K. Ng, T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, Vol. 9, No. 8, pp. 894–898, 2008.
- [13] T. Emoto, U. R. Abeyratne, T. Kusumoto, M. Akutagawa, E. Kondo, I. Kawata, T. Azuma, S. Konaka, and Y. Konouchi. Discriminating apneic snorers and benign snorers based on snoring formant extracted via a noise-

- robust linear prediction technique. *Journal of Japanese Society for Medical and Biological Engineering*, Vol. 48, No. 1, pp. 115–121, 2010.
- [14] F. Dalmaso and R. Prota. Snoring: analysis, measurement, clinical implications and applications. *European Respiratory Journal*, Vol. 9, pp. 146–159, 1996.
- [15] G. Liistro, D. Stanescu, and C. Veriter. Pattern of simulated snoring is different through mouth and nose. *Journal of Applied Physiology*, Vol. 70, No. 6, pp. 2746–2741, 1991.
- [16] 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. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 12, pp. 1553–1562, 2008.
- [17] M. Herzog, T. Metz, A. Schmidt, T. Bremert, W. Hosemann B. Venohr, and H. Kaftan. The prognostic value of simulated snoring in awake patients with suspected sleep-disordered breathing: Introduction of a new technique of examination. *Sleep*, Vol. 29, No. 11, pp. 1456–1462, 2006.
- [18] F. Lofaso, K. Leroux, M. A. Quera-Salva, G. Mroue, M. P. D’Ortho, D. Isabey, and B. Louis. Snoring detection during auto-nasal continuous positive airway pressure. *European Respiratory Journal*, Vol. 19, pp. 108–112, 2002.
- [19] R. Beck, M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *European Respiratory Journal*, Vol. 8, pp. 2120–2128, 1995.
- [20] E. J. Olson, W. R. Moore, and T. A. Staats. Obstructive sleep apnea-hypopnea. *Mayo Clinic Proceedings*, Vol. 78, No. 12, pp. 1545–1552, 2003.
- [21] Y. Inoue and Y. Yamashiro. *Sleep Disordered Breathing: update 2006 (in Japanese)*. Nippon Hyoronsha Inc., 2007.

- 
- [22] U. R. Abeyratne, A. S. Wakwella, and C. Hukins. Pitch jump probability measures for the analysis of snoring sounds in apnea. *Physiological Measurement*, Vol. 26, No. 5, pp. 779–798, 2005.
- [23] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. E. Camilleri. Sound frequency analysis and the site of snoring in natural and induced sleep. *Clinical Otolaryngology*, Vol. 27, pp. 162–166, 2002.
- [24] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam. An efficient method for snore/nonsnore classification of sleep sounds. *Physiological Measurement*, Vol. 28, No. 8, pp. 841–854, 2007.
- [25] A. Karunajeewa, U. R. Abeyratne, and C. Hukins. Silence-breathing-snore classification from snore-related sounds. *Physiological Measurement*, Vol. 29, No. 2, pp. 227–244, 2008.
- [26] A. Azarbarzin. Unsupervised classification of respiratory sound signal into snore/non-snore classes. *Proc. of Annual Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, pp. 3666–3669, 2010.
- [27] A. Yadollahi and Z. Moussavi. Automatic breath and snore sounds classification from tracheal and ambient sounds recordings. *Medical Engineering and Physics*, Vol. 32, No. 9, pp. 985–990, 2010.
- [28] T. Mikami and Y. Kojima. Automatic recognition of breathing route during sleep using snoring sounds. *IEEJ Trans. on Electronics, Information and Systems*, Vol. 131, No. 8, pp. 1516–1517, 2011.
- [29] M. R. Schroeder. Period histogram and product spectrum: New methods for fundamental frequency measurement. *Journal of the Acoustic Society of America*, Vol. 43, No. 4, pp. 829–834, 1968.

- [30] S. J. Quinn, L. Huang, P. D. M. Ellis, and J. E. F. Williams. The differentiation of snoring mechanisms using sound analysis. *Clinical Otolaryngology*, Vol. 21, pp. 119–123, 1996.
- [31] K. Ishii, N. Ueda, E. Maeda, and H. Murase. *Pattern Recognition (in Japanese)*. Ohmsha, 1998.
- [32] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [33] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
- [34] H. Hara. Snoring and sleep apnea syndrome : Clinical significance of acoustic analysis of snoring sound. *Yamaguchi Medicine (in Japanese)*, Vol. 53, No. 6, pp. 265–267, 2004.
- [35] A. K. Bieger-Farhan, N. K. Chadha, A. E. Camileri, P. Stone, and K. McGuinness. Portable method for the determination of snoring site by sound analysis. *Journal of Laryngology and Otology*, Vol. 118, No. 2, pp. 135–138, 2004.

## Chapter 4

# Spectral Classification of Oral and Nasal Snoring Sounds using a Support Vector Machine

### 4.1 Introduction

Under normal circumstances, breathing during sleep is primarily nasal rather than oral [1], but numerous investigations have shown that loud habitual snoring is due to nasal obstruction, which can have an influence on sleep disordered breathing[2][3]. Since oral breathing tends to make the upper airway more collapsible [4], such snoring caused by oral breathing is found in many sleep apnea/hypopnea patients and it should be detected in the earlier stage. But unfortunately we cannot know our own sleep condition or snoring. Thus, a simple technical method that can detect oral snoring makes it possible to develop a home monitoring device in a bedroom.

In conventional studies, the acoustic properties of snoring sounds have been analyzed so far for the purpose of discriminating apnea patients from simple snorers[5][6][7], of differentiating palatal from non-palatal snoring for medical surgery[8][9], and of detecting snoring episodes from overnight sleep sounds contaminated with non-snore artifacts[10][11]. A few researchers have focused on the difference be-

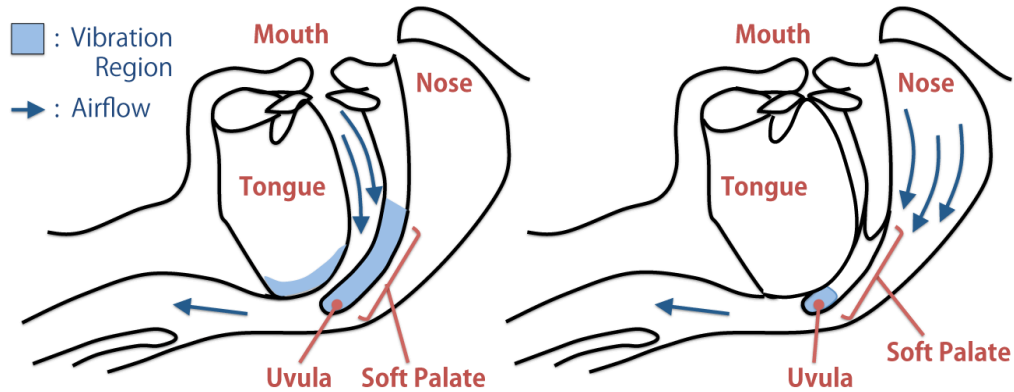


Figure 4.1: Simple chart of the snoring sites and the airflow while oral and nasal breath

tween oral and nasal snoring. Dalmasso, et al, [12] reported that the shape of cross sectional area in the upper airway is very different during oral snoring and during nasal snoring. Liistro, et al, [13] demonstrated using cineradiography that in the case of simulated snoring through the nose, the soft palate is in close contact with the back tongue and the uvula alone presents high-frequency oscillations, whereas in the case of through the mouth the whole soft palate oscillates at high frequency. These studies are informative, but they have not focused on a concrete method to classify oral and nasal snores automatically for the development of a home monitoring device.

In chapter 3, the author has tried to classify oral and nasal snoring using k-Nearest Neighbor method based on two acoustic properties (fundamental frequency and the maximum of the amplitude spectrum in a specific band) , and demonstrated that about 89% of snores are successfully classified[14][15].

In this paper, the author adopt a Support Vector Machine, which is one of the most powerful method for pattern classification. Especially, some of the kernel functions (chi-square and Kullback-Leibler) have recently been known to be effective for the classification of histogram-based feature vectors [16][17]. Moreover, Ishigaki, et al, [18] have dealt with the KL kernel as a similarity measure between frequency spectra and have numerically demonstrated the effectiveness. They also



applied the SVM with KL kernel to the early fault diagnosis of LP gas pressure regulator based on the spectral features. Accordingly, it can be expected that the SVM-based spectral classification technique outperforms our previous method reported in the ref.[15].

In order to use SVMs in general, some problem-specific parameters should be determined in advance, but SVMs have not yet been applied to the classification of oral and nasal snoring sounds. In this paper, the author adopts seven kernel functions (linear, polynomial, sigmoid, Gaussian, Laplacian, chi-square, and Kullback-Leibler) for SVM-based spectral classification of snoring sounds and examine the classification results with various values of the parameters so as to make the best choice of the kernel and to find the best values of the parameters.

## 4.2 Data Acquisition and Preprocessing

### 4.2.1 Data Acquisition

Snoring sounds the author analyzes in this paper are recorded with a portable linear PCM recorder (Olympus LS-10) with 44.1kHz sampling frequency and 16bit quantization rate. Fifteen subjects (10 benign snorers, 5 apnea patients) are asked to simulate snoring by inhaling deeply enough to produce a snoring sound in their throat with two types of breath; *oral* and *nasal*.

While producing oral snoring, subjects' nostrils are completely closed with their fingers, whereas they are asked to close their mouth while snoring nasally. Before recording, the author firstly explained the objective of this study to subjects and obtained the informed consent from all of them.

### 4.2.2 Similarity between simulated and actual snoring sounds

In this section, the author summarizes the discussion about the similarity between simulated and actual snoring sounds reported in our previous paper[15].

In the oral simulated snores, the author can find an intensity peak at over 1kHz

(figure 4), which is also found in the actual snores the site of which is known to be the tongue base[19]. The tongue base snoring tends to occur with an open oral airway[8][20]. These conventional reports about the acoustic properties of actual oral snores are consistent with those of simulated oral snores the author obtained. In addition, Herzog, et al, have reported that simulated oral snoring sounds are equivalent to the actual (nocturnal) ones according to the peaks of power spectrum [21],[22].

In the case of nasal simulated snores, the author 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[8][19]. It has been reported that the palatal snoring occurs with the oral airway closed[8], and the inspiratory nasal snores have a fundamental peak with associated harmonic peaks[23]. These conventional reports about the acoustic properties of natural nasal snores are consistent with those of simulated nasal snores the author obtained.

Based on these literature findings, it is possible that the simulated snoring sounds are similar (not necessarily equal) to the natural ones produced by snorers during sleep. Moreover, there are several merits in adopting simulated snoring: (1) most subjects were unwilling to allow us to record their actual snoring sounds during sleep, and (2) simulated snoring has also been adopted in medical studies [13], [21], [22], [24]. Therefore, the author decided to deal with simulated snoring sounds in this paper.

### 4.2.3 Episode Selection and Subsequence Extraction

Snoring sounds produced by each inhalation (called *episodes*) are cut out one by one manually, as is also done in ref. [25]. Although a snoring sound is defined as the sound of pharyngeal vibration triggered by airflow turbulence across a narrowed upper airway[26][27], some episodes sound like simple sleep noises (wheeze sounds) rather than snoring sounds (vibrational sounds). It is necessary to discriminate

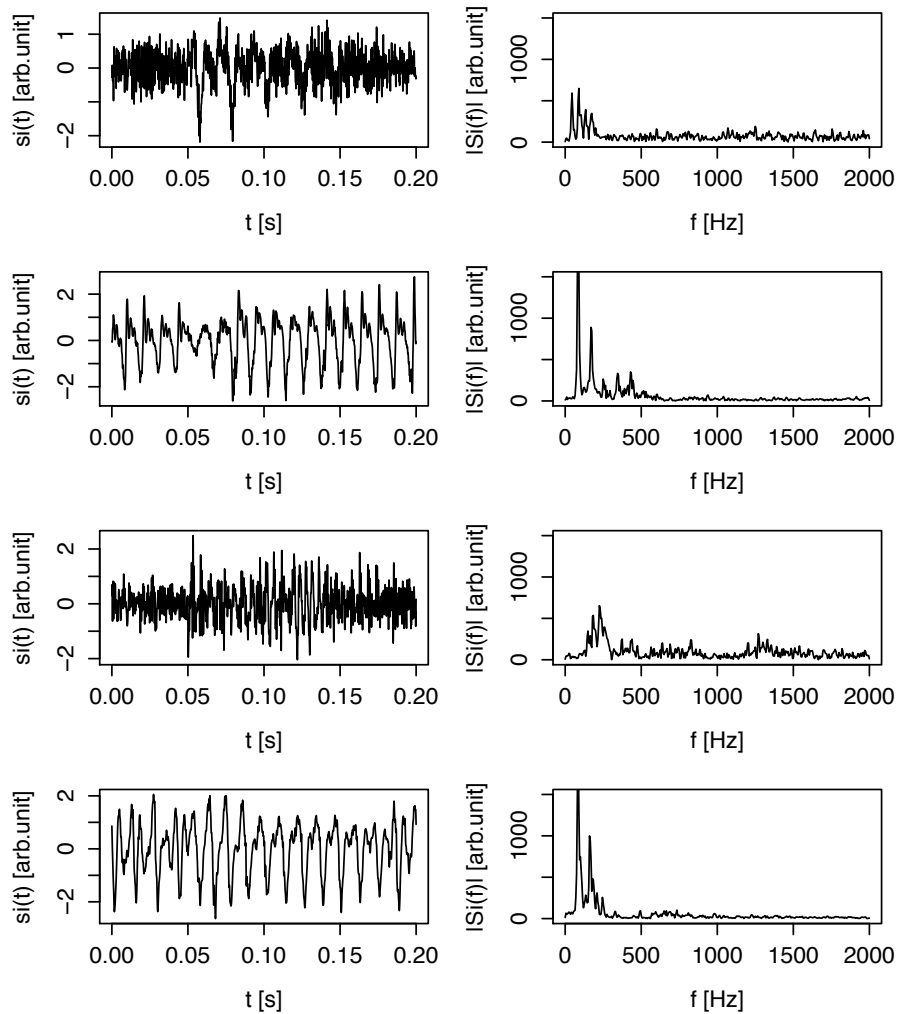


Figure 4.2: The waveforms and amplitude spectra of nasal snoring sounds

snoring sounds from simple wheeze sounds.

We can easily discriminate these sounds by listening them, but it is difficult to realize the automatic techniques [25]. So, the author thought that the manual discrimination would be more reliable. In addition, the author decided to use the episodes that *all* three persons can recognize as snores for more objective evaluation. The episodes that one or two persons rejected are not adopted as snoring episodes. This method is also adopted in ref.[25].

Since within the episode the snoring dynamics is gradually or suddenly changing

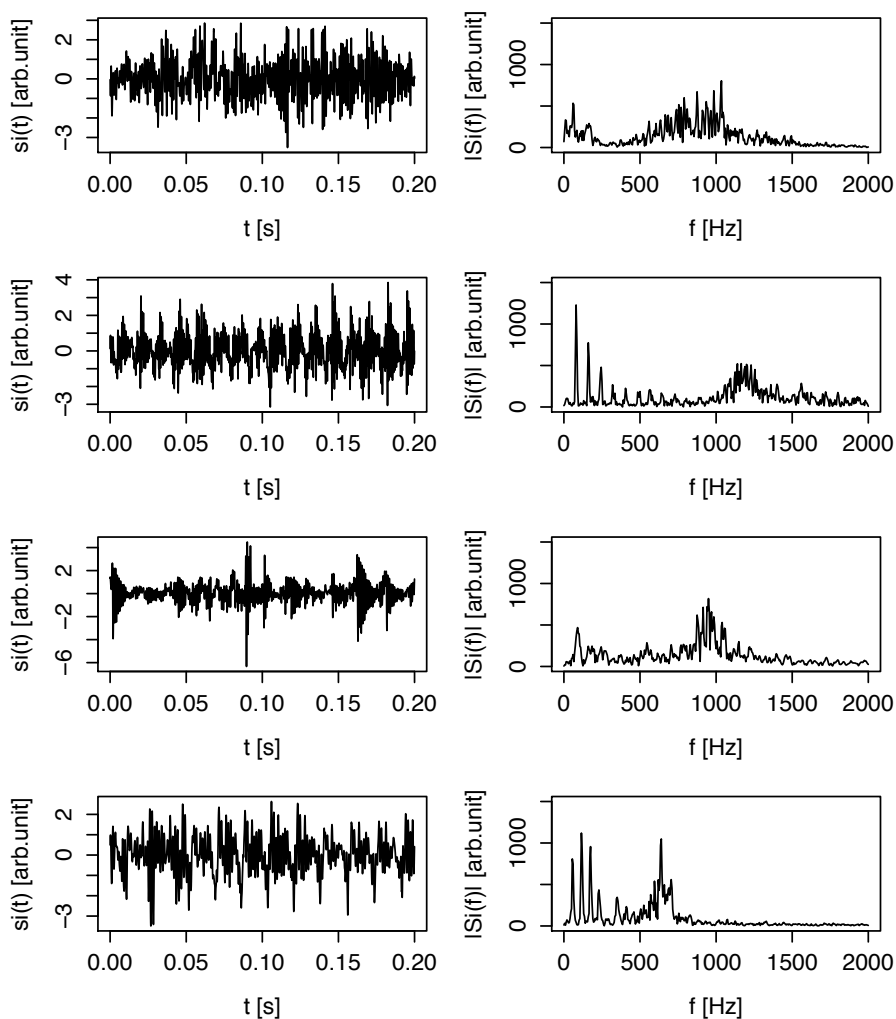


Figure 4.3: The waveforms and amplitude spectra of oral snoring sounds

as time passes[12][24], the author also cut out the short-time subsequences whose length is 0.2 seconds from the episode at every after 0.1 seconds. As a result, the author obtained 1107 nasal, and 1384 oral subsequences from all recorded data and those are defined as our classification targets. The obtained subsequences are the same as the ones used in our previous work[15].

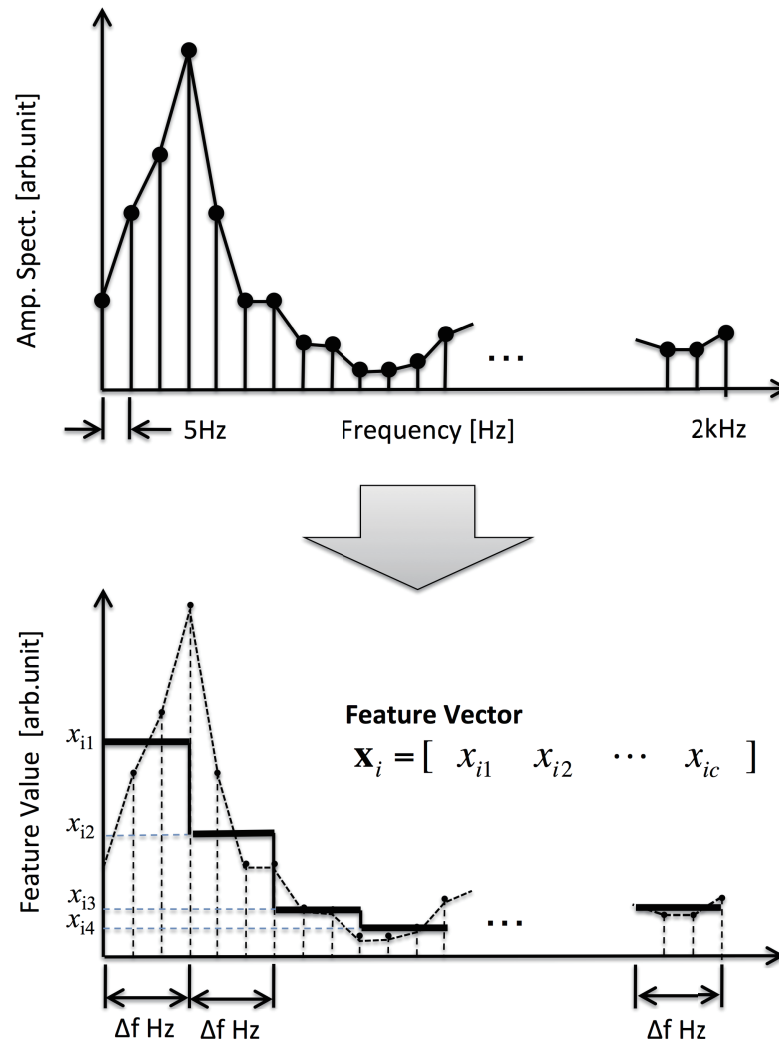


Figure 4.4: Discrete amplitude spectra and elements of the feature vector.

### 4.3 Feature Extraction Method

Firstly, all subsequences are transformed to the frequency spectra using Fast Fourier Transform (FFT) with Hanning window function. According to the digital signal processing theory, the FFT amplitude spectrum is a set of discrete spectral values at every 5Hz (the inverse of the subsequence length; 0.2 seconds) in the frequency domain. Since it has been reported that snoring sounds mainly consist of lower 2kHz components[8], the spectra are low-pass filtered with 2kHz cut off frequency.

Figures 4.2 and 4.3 show, respectively, representative subsequences and the amplitude spectra of oral and nasal snores recorded from different subjects respectively. Each spectrum is composed of  $d(= 2000/5 = 400)$  discrete points. The difference between oral and nasal snores is clearly recognizable; nasal snores are composed of lower frequency components below 500Hz whereas oral ones are composed of not only lower frequency components but also the intensity peak above 500Hz.

According to Quinn, et al, [8] lower frequency components indicate the vibration of soft palate, while the higher frequency components indicate the airflow noise which occurs around the tongue base (see figure 4.6). Since open mouth tends to make the upper airway around the tongue base more collapsible[28], oral snoring may consist of both the soft palate vibration and the tongue base noise as well. This is a biomechanical rationale for using the spectral properties to classify oral and nasal snoring sounds. But the original spectra have too many data points to represent the spectral form, so it is necessary to reduce some redundancy in these spectra.

In this paper, the amplitude spectrum is averaged over  $\Delta f$  Hz band shown in figure 4.4. The averaged spectrum is defined as  $c$ -dimensional feature vector,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ic})$  where  $c = 2000/\Delta f$  is the number of spectrum division and  $x_{ij}$  is the averaged value of the amplitude spectrum from  $j \cdot \Delta f$  Hz to  $(j + 1) \cdot \Delta f$  Hz. Naturally the value of  $c$  should be determined to the suitable value to obtain

a high performance.

## 4.4 Classification Method

### 4.4.1 Support Vector Machine

A Support Vector Machine (SVM) is a nonlinear two-class classifier that determines the unique hyper-plane by maximizing the distance from it to the nearest data point on each class.

Let  $\mathbf{x}_i$  and  $y_i \in \{+1, -1\}$  be the feature vector of the  $i$ th subsequence and its class label (+1 and -1 mean "oral" and "nasal" respectively), the dual form of this optimization problem turned out to be a quadratic convex programming as follows,

$$\text{maximize } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (4.1)$$

$$\text{subject to } 0 \leq \alpha_i \leq C \quad (4.2)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (4.3)$$

where  $\alpha_i$  is a Lagrange multiplier, and  $K(\mathbf{x}_i, \mathbf{x}_j)$  is a kernel function that means the dot-product in high-dimensional Hilbert space, and  $C$  is the penalty factor.

In this study, the kernlab package in R, a statistical software, was used to train the SVM, which is implemented by SMO (Sequential Minimum Optimization) algorithm, one of the simplest and the most effective techniques[29]. The optimal classification function is determined as follows:

$$y(\mathbf{x}) = \text{sign} \left( \sum_{i \in S} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) - b^* \right) \quad (4.4)$$

where  $S$  is a set of the indices of support vectors,  $\alpha_i^*$  is a solution of the optimization problem, and  $b^*$  is the bias parameter that can be determined by the optimal solution.

#### 4.4.2 Kernel Functions

Many kernel functions have been proposed so far, and the author firstly decided to adopt five representative kernel functions: linear, polynomial, sigmoid, Gaussian, and Laplacian. Moreover, based on the feature vectors described in section 3, the author also adopted the chi-square ( $\chi^2$ ) kernel and the Kullback-Leibler (KL) kernel. All kernel functions adopted here are as follows:

The  $\chi^2$  kernel is firstly introduced by Chapelle[16], et al, for the purpose of histogram-based image classification. They demonstrated that this kernel function is especially useful for histogram-based feature vectors and leads to a better result in comparison with Euclidean distance based kernels (namely, Gaussian and Laplacian).

The KL kernel is relatively novel and firstly introduced by Moreno, et al [17], for multimedia applications. This kernel function is based on the KL divergence, a measure of dissimilarity between two probability density functions. Ishigaki, et al, [18] applied it to the SVM-based spectral classification for early fault diagnosis of the LP gas pressure regulator and demonstrated that the SVM with KL kernel is the best among 6 kernel functions (polynomial, Gaussian, Laplacian, sublinear,  $\chi^2$ , and KL). They also reported that the KL kernel is especially robust for the shift of the spectral peak position in the frequency domain.

As for KL kernel, a *symmetric* version of KL divergence, which satisfies the axiom of distance, is generally adopted in conventional studies[17][18], so the author also used this type of KL kernel in this paper. In order to use these kernel functions, the best parameters should be selected under some restrictions to satisfy Mercer's condition and to avoid some numerical problems.

#### 4.4.3 A Criterion for Finding the Best Parameters

In the literature of the application of SVMs, the optimal value of  $C$  is determined to the value ranging in one decade steps from 0.01 to 100 in refs.[30][31]. So, the



Table 4.1: Kernel functions for the SVM used in this study

Kernel Name	Definition	Parameters
Linear	$K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$	none
Polynomial	$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + c)^d$	$c, d$
Sigmoid	$K(\mathbf{x}, \mathbf{y}) = \tanh(\beta \mathbf{x}^T \mathbf{y} + c)$	$\beta, c$
Gaussian	$K(\mathbf{x}, \mathbf{y}) = \exp(-\beta \ \mathbf{x} - \mathbf{y}\ ^2)$	$\beta$
Laplacian	$K(\mathbf{x}, \mathbf{y}) = \exp(-\beta \ \mathbf{x} - \mathbf{y}\ )$	$\beta$
$\chi^2$	$K(\mathbf{x}, \mathbf{y}) = \exp(-\beta \cdot D(\mathbf{x}, \mathbf{y}))$	$\beta$
	$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^d \frac{(x_i - y_i)^2}{x_i + y_i}$	$\beta$
Kullback-Leibler	$K(\mathbf{x}, \mathbf{y}) = \exp(-\beta \cdot I(\mathbf{x} \mathbf{y}))$	$\beta$
	$I(\mathbf{x} \mathbf{y}) = \sum_{i=1}^d (x_i - y_i) \log \frac{x_i}{y_i}$	$\beta$

author also sets  $C$ -parameter to this range. But if  $C = 0.01$  or  $C = 100$  is the best, the author additionally investigates more results when  $C = 0.001$  or  $C = 10000$ .

The parameters of polynomial kernel ( $c, d$ ) should be adjusted to positive values so as to satisfy Mercer's condition and  $d$  is theoretically a positive integer. But it took a very long time to solve the optimization problem when the author sets  $d \geq 8$ , so the author adjusted this value to be less than 8.

The parameter of  $\beta$  except for sigmoid kernel should also be set to a positive value so that a kernel matrix is positive definite. Thus, the author roughly determines  $\beta$  to the value ranging in one decade steps from very low ( $10^{-10}$ ) to very high ( $10^{10}$ ) at first. After that, the range in which the classification rate (calculated under the LSO test described in the next section) becomes greater than 0.7 is specified and the parameters are reset to the value from this range in 0.1 decade steps. The parameter of  $c$  for polynomial kernel are also determined based on the above criterion. The concrete range where the classification rate is greater than 0.7 is shown in table 2.

In the case of sigmoid kernel, the likelihood of obtaining a kernel matrix that is not positive definite is much higher if  $c$  is negative[32]. Meanwhile it is also known

Table 4.2: The concrete values of kernel specific parameters for obtaining the best classification accuracy under the LSO test when  $\Delta f$  is determined to 500.

Kernel	Parameter range for adjustment	Optimal
Linear	$C = 0.01, 0.1, 1, 10, 100$	$C = 10$
Polynom.	$c = 10^{-5}, 10^{-4.9}, 10^{-4.8}, \dots, 10^2$ $d = 1, 2, 3, 4, 5, 6, 7$ $C = 0.01, 0.1, 1, 10, 100$	$c = 10^{0.2}$ $d = 3$ $C = 1$
Sigmoid	$c = 10^{-10}, 10^{-9.9}, 10^{-9.8}, \dots, 10^0$ $\beta = 10^{-4}, 10^{-3.9}, 10^{-3.8}, \dots, 10^{-0.1}, 10^0$ $C = 0.01, 0.1, 1, 10, 100$	$c = 10^{-4.8}$ $\beta = 10^{-1.8}$ $C = 10$
Gauss	$\beta = 10^{-4}, 10^{-3.9}, 10^{-3.8}, \dots, 10^2$ $C = 0.01, 0.1, 1, 10, 100$	$\beta = 10^{-0.4}$ $C = 0.1$
Laplace	$\beta = 10^{-6}, 10^{-5.9}, 10^{-5.8}, \dots, 10^1$ $C = 0.01, 0.1, 1, 10, 100$	$\beta = 10^{0.1}$ $C = 0.1$
$\chi^2$	$\beta = 10^{-5}, 10^{-4.9}, 10^{-4.8}, \dots, 10^0$ $C = 0.01, 0.1, 1, 10, 100$	$\beta = 10^{-2.1}$ $C = 1$
KL	$\beta = 10^{-4}, 10^{-3.9}, 10^{-3.8}, \dots, 10^2$ $C = 0.01, 0.1, 1, 10, 100$	$\beta = 10^{-0.1}$ $C = 1$

that  $\beta > 0$  and  $c < 0$  are better choice in many cases [33]. So, the author decided to adjust these values to from very low negative ( $-10^{10}$ ) to very high positive ( $10^{10}$ ) at first. After that, the author precisely resets these values to the range, shown in table 2, in which the classification rate becomes greater than 0.7.

#### 4.4.4 Comparison with the Other Classifiers

In this chapter, SVM is adopted to classify the spectral properties of snoring sounds with various kernel functions, but it does not guarantee the best method for the spectral classification. For the purpose of performance comparison, the other classifiers such as  $k$ -Nearest Neighbor method ( $k$ -NN), Multilayer Perceptron (MLP), Learning Vector Quantization (LVQ) are also adopted to classify oral and nasal snoring sounds.

In the same way as the SVM, classifier-specific parameters have to be adjusted to a suitable value so as to obtain the best classification performance. Those

Table 4.3: Four classifiers adopted to discriminate oral and nasal snoring sounds for comparison

Classifier	Specific parameters	Representation
$k$ -NN	the number of neighbors	$k$
MLP	the number of neurons in the hidden layer	$N_h$
LVQ	the number of code-book vectors	$N_c$

parameters are listed on table 4.3. For comparison, the classifiers adopted here are quite general and have been widely used traditionally in various field[34][35][36].

### **$k$ -Nearest Neighbor method**

In the case of  $k$ -NN method, the number of neighbors ( $k$ ) is adjusted to from 1 to 20 by 2 steps shown in table 4.3, and the author used Euclid distance as the dissimilarity between two feature vectors. In this method, the classification result is the majority class among the  $k$  nearest training data closest to the input. This method is also adopted in section 3.

### **Multilayer Perceptron**

The structure of MLP is three layers (input, hidden, and output). Since it receives 4-dimensional feature vectors as the input, the number of neurons in input layer is four. Similarly, there is the only one neuron in the output layer whose output is 1 (if oral) or 0 (if nasal). The number of neurons in hidden layer,  $N_h$ , is adjusted to from 1 to 20 shown in table 4.3. The connection weights are updated with Back-Propagation (BP) learning. The classification rate is averaged over 20 iterations with different initial weight values.

### **Learning Vector Quantization**

In the case of LVQ, there are various LVQ algorithms such as LVQ1, LVQ2, LVQ3, and OLVQ1, which are somewhat different from each other. In this study, the au-

thor adopted LVQ1 algorithm. The number of codebook vectors,  $N_c$ , are adjusted to from 10 to 800 by 20 steps (see table 4.3).

## 4.5 Performance Evaluation Methods

### 4.5.1 10-fold Cross Validation Test

In this paper, the author evaluates our method by calculating the ratio of correctly classified data with 10-fold Cross Validation test as follows:

1. Divide all data randomly into 10 groups expressed by  $G_1, G_2, \dots, G_{10}$
2.  $j \leftarrow 1$
3.  $\forall \mathbf{x} \in G_j$  are used as *test* data, while the data belonged to the remainder groups,  $\forall \mathbf{x} \in G_{i \neq j}$ , are used as *training* data.
4. Solve the optimization problem expressed by eqs.4.1 and 4.3.
5. The SVM with the optimized parameters classifies the test data into "oral" or "nasal" categories.
6. The ratio of correctly classified test data,  $R_j$ , is calculated.
7.  $j \leftarrow j + 1$  and go to step 3 while  $j \leq 10$
8. The classification accuracy of all data is calculated by  $R = \sum_{j=1}^{10} |G_j| R_j / \sum_j |G_j|$ .

This method is generally used in order to evaluate the classification results in many problems[36]. In this paper, the classification accuracy is averaged over 20 iterations with different random values.

### 4.5.2 Leave-Subject-Out Test

The Leave-*Subject*-Out (LSO) test was also used in our previous works[14][15] so as to evaluate the effect of the individual difference to the classification performance. In this test, all subsequences obtained from the same subject are allocated to the same group in step 1. Namely, the number of groups is equal to the number of subjects. The other procedure is the same as 10-fold CV test.

If the individual difference is higher than the difference between oral and nasal snore data, the classification result calculated with this test may become lower than that with 10-fold CV test.

## 4.6 Results and Discussion

### 4.6.1 The Optimal Parameter of $\Delta f$ and the Best Choice of Kernels

In order to find the optimal value of  $\Delta f$ , the author used SVM with five kernel functions respectively under the LSO test. Figure 4.5 shows classification accuracy of oral and nasal snoring sounds against the coarse width of frequency ( $\Delta f$ ). The classification accuracy at every  $\Delta f$  shows the best result obtained with the SVM when the parameters are determined to the range shown in table 4.7.

At every  $\Delta f$  Hz, the accuracy of KL kernel is relatively higher than those of the other kernels, especially when  $\Delta f$  is lower. The highest classification accuracy is obtained at  $\Delta f = 500\text{Hz}$  in every kernel, so it is found out that the optimal dimension of feature vector is four, and the KL kernel is the best among our adopted five kernel functions.

Table 4.5 shows the classification accuracy in more detail when the parameters are determined to the optimal values shown in table 4.7 and  $\Delta f$  is 500Hz. From this result, there are no clear difference among five kernels under the 10-fold CV test (around 96%). But, in the case of LSO test, all kernel functions achieve good results (above 93% at least), and the KL kernel achieves the best result (95.74%). As far as the author considers from these results, KL kernel seems to be the best choice. But the accuracy with KL kernel is only 0.8% higher than that with  $\chi^2$  kernel, so the author cannot necessarily advocate that the KL kernel is better than  $\chi^2$  kernel. In addition, since the classification accuracy with Gaussian kernel is 0.02% higher than that with Laplacian kernel, it is natural to consider that these results are almost the same. The polynomial kernel may give the worst accuracy among them, but it is above 93% and only 0.45% lower than the Laplacian kernel.

Table 4.4: Classification accuracies of SVMs with five different kernels respectively under the 10-fold CV and LSO test when  $\Delta f$  is determined to 500.

Kernel	Classification Accuracy		
	Nasal (%)	Oral (%)	Total (%)
Linear	93.59 $\pm$ 0.07%	92.63 $\pm$ 0.09%	93.06 $\pm$ 0.06%
Polynom.	96.95 $\pm$ 0.09%	96.45 $\pm$ 0.11%	96.67 $\pm$ 0.09%
Sigmoid	96.92 $\pm$ 0.15%	96.75 $\pm$ 0.10%	96.82 $\pm$ 0.09%
Gauss	96.17 $\pm$ 0.08%	96.75 $\pm$ 0.05%	96.49 $\pm$ 0.04%
Laplace	96.11 $\pm$ 0.09%	96.88 $\pm$ 0.06%	96.53 $\pm$ 0.05%
$\chi^2$	97.10 $\pm$ 0.15%	96.80 $\pm$ 0.18%	96.93 $\pm$ 0.12%
KL	96.72 $\pm$ 0.08%	96.68 $\pm$ 0.06%	96.70 $\pm$ 0.05%

Table 4.5: Classification accuracies of SVMs with five different kernels respectively under the 10-fold CV and LSO test when  $\Delta f$  is determined to 500.

Kernel	Classification Accuracy		
	Nasal (%)	Oral (%)	Total (%)
Linear	1026/1107 (92.68%)	1236/1384 (89.30%)	2262/2491 (90.81%)
Polynom.	1046/1107 (94.49%)	1284/1384 (92.77%)	2330/2491 (93.54%)
Sigmoid	1048/1107 (94.67%)	1285/1384 (92.85%)	2333/2491 (93.66%)
Gauss	1054/1107 (95.21%)	1290/1384 (93.21%)	2344/2491 (94.10%)
Laplace	1051/1107 (94.94%)	1290/1384 (93.21%)	2341/2491 (93.98%)
$\chi^2$	1064/1107 (96.12%)	1301/1384 (94.00%)	2365/2491 (94.94%)
KL	1067/1107 (96.39%)	1318/1384 (95.23%)	2385/2491 (95.74%)

In summary, the KL kernel is the best choice if  $\Delta f$  is set to below 500Hz. But, in the case of  $\Delta f = 500\text{Hz}$ , there are no clear differences among the classification accuracies obtained with five kernel functions, and at least over 93% of snore subsequences are correctly classified by the use of any kernel functions under the LSO test.

#### 4.6.2 Individual Differences and Reliability

If there are significant individual differences in our data and the number of subjects is relatively not enough as well, the classification rate under the LSO test becomes much lower than that under the 10-fold CV test. So, the comparison between these

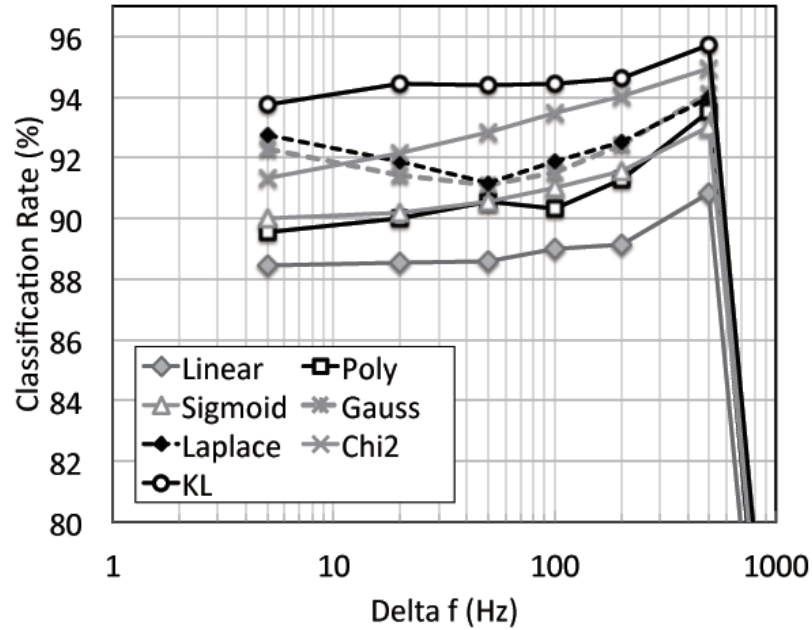


Figure 4.5: Classification rate against various  $\Delta f$  parameter estimated with SVM

tests indicates whether our results are reliable or not.

From table 3, the classification results under the LSO test are about 1-3% lower than the other. There seem to be slight individual differences in our data, but the results are fairly objective and the performance is sufficiently good: at least over 90% under the LSO test. Many subjects are in general preferable for objective evaluation, but, as far as the author investigates the classification results, 15 subjects are fairly enough to obtain the objective results. In fact, the number of subjects in some much-cited papers[5][7][8][19][25][23] is in the range from 10 to 20.

### 4.6.3 Comparison with Our Conventional Method

Finally, the author compares our results with the conventional method which has been proposed by the authors[15]. In our conventional method, two acoustic properties of snoring sounds are firstly extracted: fundamental frequency ( $f_b$ ) and the

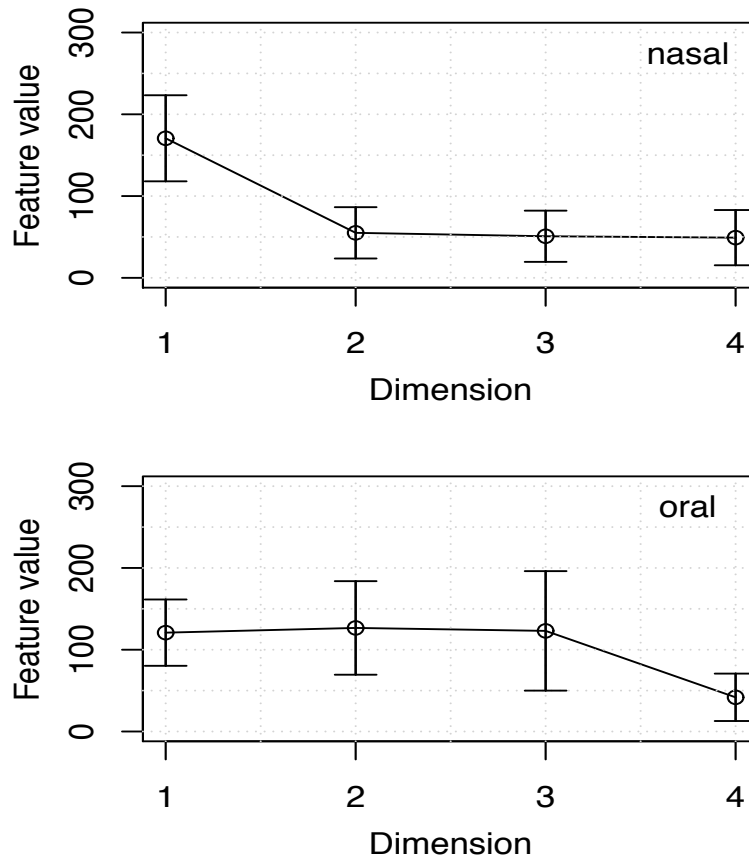


Figure 4.6: The average of feature values on each dimension when  $\Delta f$  is determined to 500.

maximum of the amplitude spectrum ( $M$ ) in a specific band from 700Hz to 1.4kHz. After that, 2-dimensional feature vectors composed of  $f_b$  and  $M$  are made for all data, and the classification accuracy is calculated using k-Nearest Neighbor method under the 10-fold CV test and the LSO tests. Table 4.6 shows the classification accuracy which has been reported in ref.[15]. Compared with the table 4.5, the best classification accuracy in this paper is about 5% higher than that reported in our previous study. So, the author analyzes the data that are miss-classified with the conventional method, but correctly classified with our method in this paper.

Figure 4.7 shows the waveforms and the amplitude spectra of such data. According to these panels, they are not periodic waveforms and there are no clear



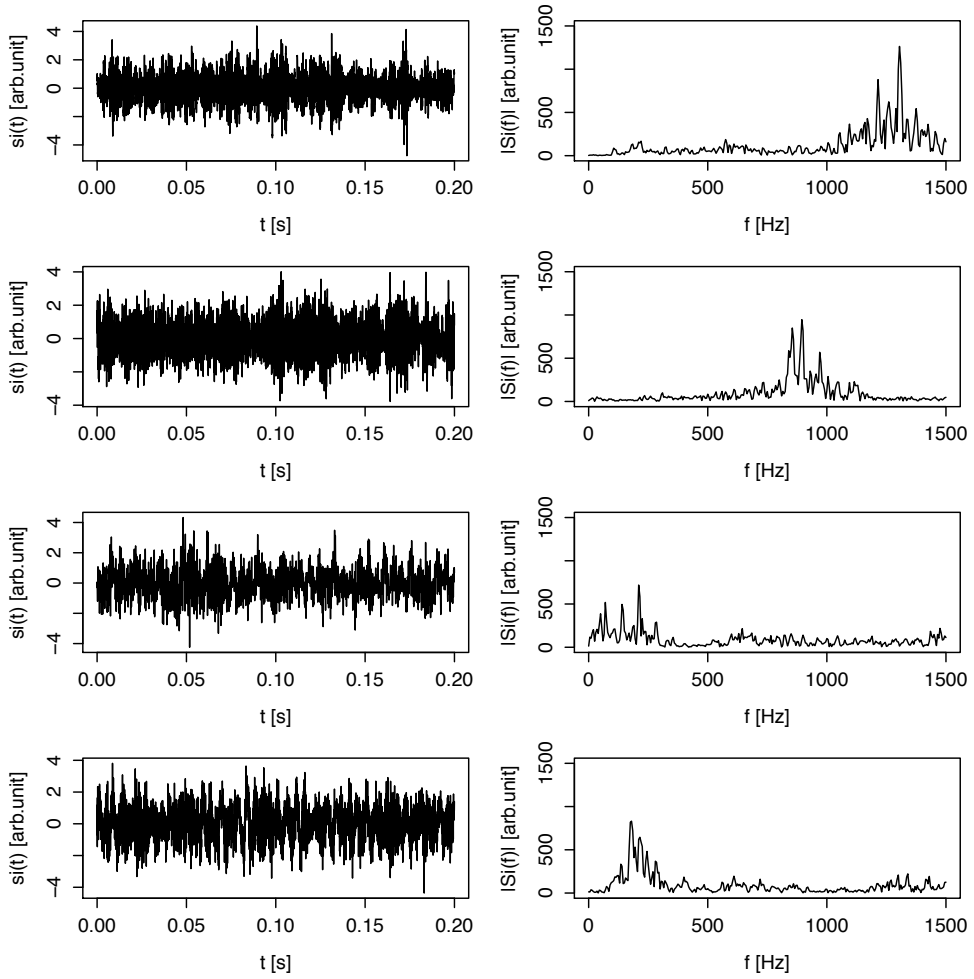


Figure 4.7: The waveforms and amplitude spectra of non-vibrational sounds in oral (upper two) and nasal (lower two) snoring

prominent peaks in the lower frequency domain as well. They are derived from the turbulence noise that occurs when the airflow pressure is not enough to generate the palatal vibration. From the biomechanism of snoring, these snores are so-called *unvoiced* sounds, whereas the vibrational snores *voiced* sounds [25]. The fundamental frequency of unvoiced sounds are not determined and becomes meaningless by the use of Harmonic Product Spectrum (HPS), a method for fundamental frequency estimation, which is adopted in our previous work.

However, by the use of our method in this paper, these unvoiced snores can

Table 4.6: Classification accuracy estimated with the conventional method in our previous work[15]

Test	Classification Accuracy		
	Nasal (%)	Oral (%)	Total (%)
10-fold CV	1014/1107 (91.60%)	1289/1384 (93.14%)	2303/2491 (92.45%)
LSO	964/1107 (87.08%)	1268/1384 (91.62%)	2232/2491 (89.60%)

also be classified correctly, because snoring sounds through nose do not have any frequency components above 500Hz even if they are not periodic, and the most components are intensively concentrated in the lower frequency domain below 500Hz. In addition, unvoiced oral snores do not have any lower frequency components below 500Hz, but have higher ones above 700Hz. Namely, the rough spectrum below 2kHz is sufficiently effective for such classification.

#### 4.6.4 Miss-classified Spectral Features

The SVM seems to be powerful for the purpose of classifying oral and nasal snoring sounds, since over 95% classification rate can be obtained at most under the LSO test. But there are still some spectral features that cannot be classified correctly. Figure 4.8 shows the representative miss-classified data that the author investigated.

In these spectra, the author can find some irregular peaks on the frequency domain. In the upper two panels that show the miss-classified oral snoring sounds, the spectral components are concentrated on the lower frequency domain which are similar to the nasal snoring sounds shown in figure 4.2. The author can find some peaks in the range above 500 Hz, but compared with the lower frequency components they are not prominent enough to be recognized as oral snores.

In the case of the miss-classified nasal snoring sounds shown in the lower two panels, the spectral components are almost equally distributed. These waveforms are not classified to nasal snoring sounds, because the lower frequency components

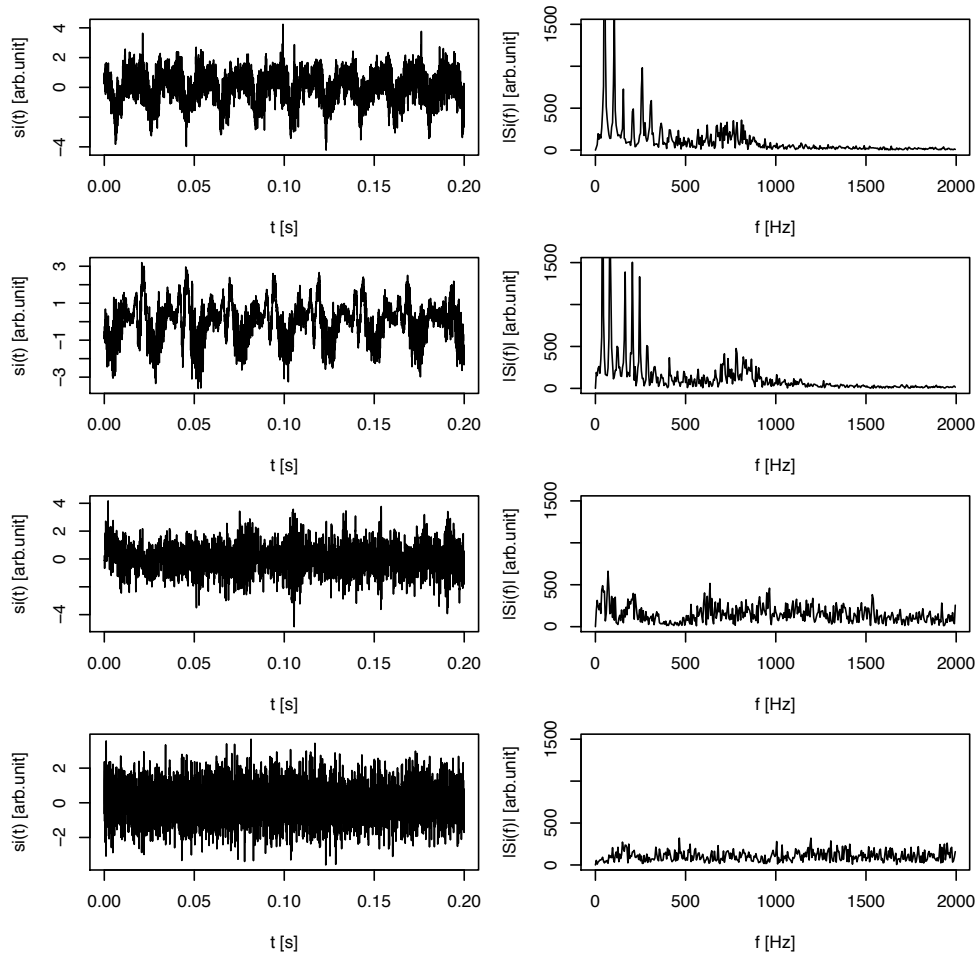


Figure 4.8: The waveforms and amplitude spectra of miss-classified oral (upper two) and nasal (lower two) snoring sounds

below 500Hz are dominant over the spectral properties of nasal snoring sounds.

#### 4.6.5 Computational Time

In this experiment, the author used the kernlab, a package for kernel-based machine learning methods in R. All programs are executed on the R software installed on the Apple's Mac mini with 2.4 GHz Core2Duo CPU and 4GB memory. Except for the polynomial kernel with higher degree greater than 2, it takes 0.94 seconds on average to solve the optimization problem defined as eqs.1 and 2. This package is implemented with the Sequential Minimum Optimization (SMO) algorithm.

Table 4.7: Specific parameters of classification methods

Classifier	Parameter range for adjustment	optimal
$k$ -NN	$k = 1, 3, 5, \dots, 37, 39$	$k = 11$
MLP	$N_h = 1, 2, 3, \dots, 19, 20$	$N_h = 4$
LVQ	$N_c = 100, 120, 140, \dots, 780, 800$	$N_c = 260$

But if once the optimization problem defined as eqs.1 and 2 is solved in advance, the computational effort is only required to calculate the classification result using eq.3. For our 2491 spectral data, it takes about 25 seconds to make the feature vectors by following the feature extraction method in section 3 (executed by interpreter) and 80 milliseconds on average to calculate the classification results using eq.3 (executed in native code). If the feature extraction process is implemented in native code, the computational time for feature extraction is drastically decreased. In any case, the computational time can be estimated at less than 10 milliseconds (= (25 seconds + 80 milliseconds) / 2491 subsequences) at most.

Actually, the computation time depends highly on the hardware performance. But this result is efficient enough for consideration of real-time processing if our method is run on the home medical device equipped with an equivalent performance CPU, because the next subsequence comes every 0.1 seconds. This is long enough to finish calculating the classification result of the current subsequence.

#### 4.6.6 Comparison with Other Classifiers

Classification performance is evaluated under leave-subject-out (LSO) test. Figure 4.12 shows the classification rate when using four classifiers as well as SVMs with three kernels. In the case of MLP and LVQ, classification rate is averaged over 20 iterations with different initial weights. Since the results are not so much different from each other, every classifier can be adopted for the purpose of oral and nasal snore classification. But, the results of SVM-based classification are relatively better than the other classification methods. In addition, the results of LVQ and

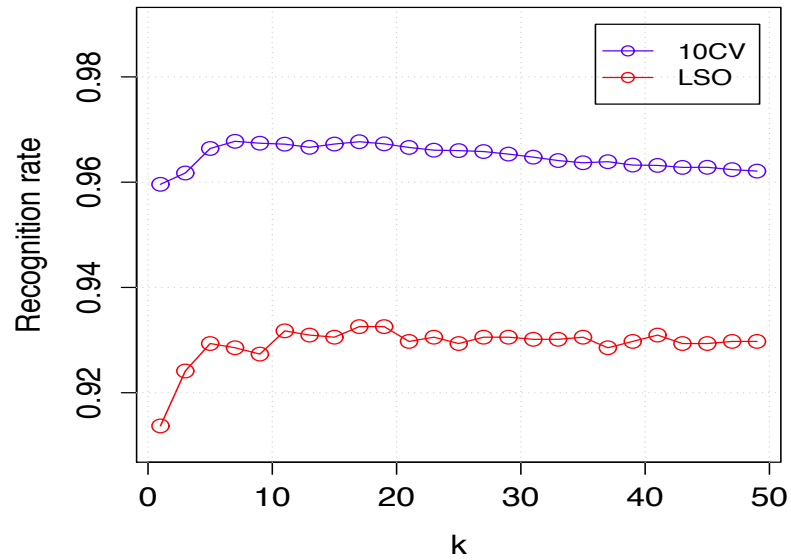


Figure 4.9: Recognition rate when using  $k$ -Nearest Neighbor method with various  $k$  numbers

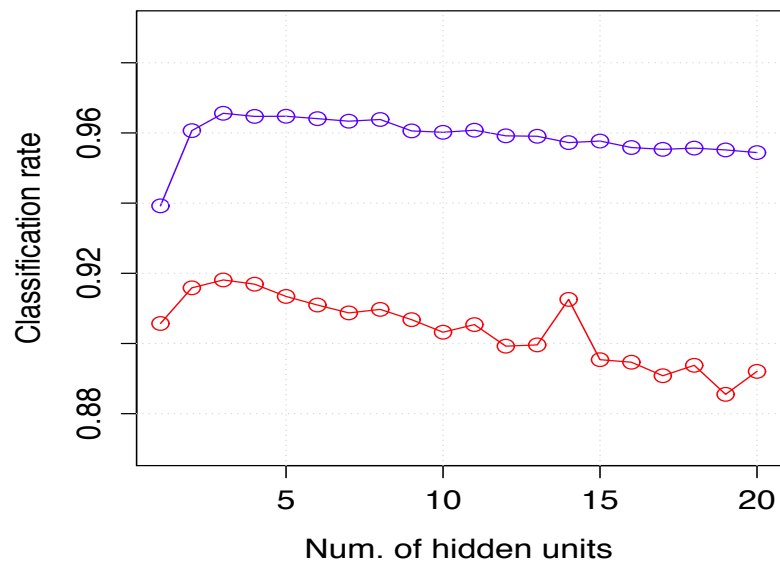


Figure 4.10: Recognition rate when using MLP with various numbers of neurons in hidden layer

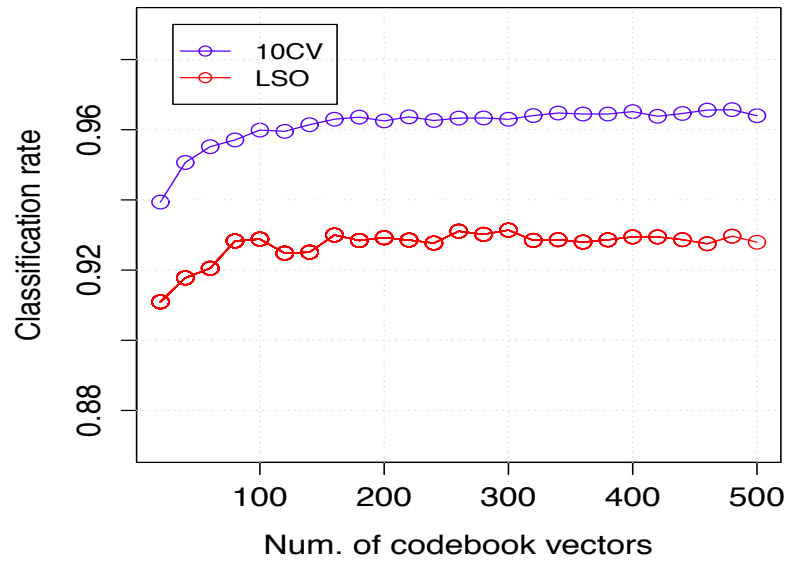


Figure 4.11: Recognition rate when using LVQ with various numbers of codebook vectors

$k$ -NN are slightly higher than those of CART and MLP.

The result of MLP is pretty surprisingly worse than those of the other methods except for CART. When using MLP, there are some annoying problems which may trap in a local minimum. Indeed, the author found that the total classification rate of MLP sometimes reached over 92%, but in many cases it could not be over 92%.

Similarly, the initial setting of LVQ is generally done at random, which leads to the convergence to one of the local minima. Even if the parameters are determined to the optimal value shown in table 4.7, it is not necessarily guaranteed that the best unique solution will be found. Therefore, in the case of MLP or LVQ, some trials should be done with different initial weights so as to obtain a good performance.

In the case of kNN method, the classification performance is relatively good. The input feature vector is compared with all of the reference data (with class labels) stored in advance in this method. This means that the computational cost becomes exponentially higher as the number of reference data is increasing. In this

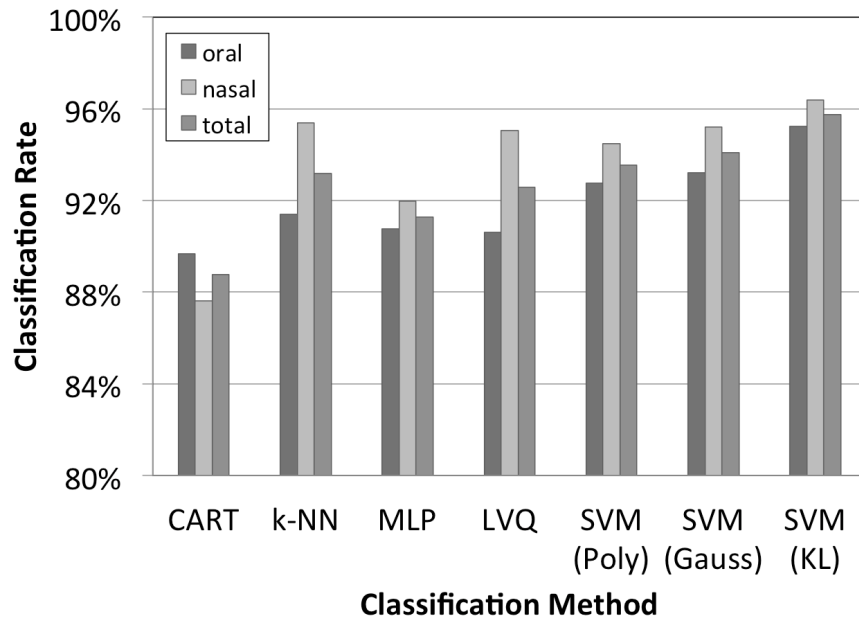


Figure 4.12: Recognition rate of all classifiers under the LSO test

experiment, the total number of subsequences is 2491 obtained from 16 subjects, but the problem of computational cost will be given rise to if more and more data are used for this classification.

From those points of view, it is said that the SVM is the best choice according to the following reasons:

1. The optimal classification boundary can uniquely be determined.
2. Once the optimal solution is determined, the classification result is easily obtained by only calculating eq.4.4
3. With KL kernel, the classification performance can reach the best compared with other methods, including other kernels.

## 4.7 Conclusion

In this paper, the author adopted an SVM with seven kernel functions (linear, polynomial, sigmoid, Gaussian, Laplacian,  $\chi^2$ , and KL) for classifying oral and nasal snoring sounds. The author can obtain around 96% classification accuracy under the 10-fold CV test and at least over 93% under the LSO test using every kernel function. The best accuracy reaches over 95% by the use of KL kernel under the LSO test, but there is not so much difference among seven kernel functions. Compared with our previous work, the classification accuracy is improved at the rate of about 5% by the use of our method.

As for computational time, SVM-based classification process has to be completed within 0.1 seconds. Even in our computation environment (Mac mini with Core2Duo CPU, 4MB memory, and implemented in R code), the classification has been finished within 10 milliseconds. Naturally, the computational time does not necessarily depend on the CPU performance but on the other factors as well. But our method is efficient enough to realize a real-time processing if the computer's performance is almost the same as our computer.

Compared with the other classification methods, the SVM with any kernel functions are a bit better than the other methods in the point of classification rate estimated under the LSO test. Moreover, the optimal classification boundary of SVM can uniquely be calculated whereas that of the MLP or LVQ cannot. In addition, the computational cost is exponentially increasing with the number of training data for kNN method.

Future works are listed bellow:

1. It is necessary to extract only snoring sounds from the recorded data contaminated with the other artifacts such as ambulance, dog's barks, taking while sleeping, linen noise, and so on.
2. The quantitative relation between the number of oral snores and the severity of OSAS should be clarified.



3. A method that can estimate the severity using the frequency of oral snores should be developed.

## References

- [1] H. A. McLean, A. M. Urton, H. S. Driver, A. K. W. Tan, A. G. Day, P. W. Munt, and M. F. Fitzpatrick. Effect of treating severe nasal obstruction on the severity of obstructive sleep apnoea. *European Respiratory Journal*, Vol. 25, pp. 521–527, 2005.
- [2] C. W. Zwillich, C. Pickett, F. N. Hanson, and J. V. Weil. Disturbed sleep and prolonged apnea during nasal obstruction in normal men. *American Review of Respiratory Disease*, Vol. 124, pp. 158–160, 1981.
- [3] J. E. Cillo Jr, R. Finn, and R. M. Dasheiff. Combined open rhinoplasty with spreader grafts and laser-assisted uvuloplasty for sleep-disordered breathing: Long-term subjective outcomes. *Journal of Oral and Maxillofacial Surgery*, Vol. 64, No. 8, pp. 1241–1247, 2006.
- [4] H. Hara, Y. Miyauchi, J. Matsumoto, and Y. Yamashita. Morphological change of the upper airway in OSAS patients with open mouth. *Stomatopharyngology (in Japanese)*, Vol. 20, No. 1, p. 45, 2007.
- [5] J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mananas, and P. Caminal. Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. *European Respiratory Journal*, Vol. 9, No. 11, pp. 2365–2370, 1996.

- 
- [6] H. Hara. Snoring and sleep apnea syndrome : Clinical significance of acoustic analysis of snoring sound. *Yamaguchi Medicine (in Japanese)*, Vol. 53, No. 6, pp. 265–267, 2004.
- [7] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. *Engineering in Medicine and Biology Society, 2003 Proc. of the 25th Annual International Conference of the IEEE*, Vol. 3, pp. 2527–2530, 2003.
- [8] S. J Quinn, L. Huang, P. D. M. Ellis, and J. E. F. Williams. The differentiation of snoring mechanisms using sound analysis. *Clinical Otolaryngology*, Vol. 21, pp. 119–123, 1996.
- [9] J. E. Osborne, E. Z. Osman, P. D. Hill, B. V. Lee, and C. Sparkes. A new acoustic method of differentiating palatal from non-palatal snoring. *Clinical Otolaryngology*, Vol. 24, No. 2, pp. 130–133, 1999.
- [10] U. R. Abeyratne, A. S. Wakwella, and C. Hukins. Pitch jump probability measures for the analysis of snoring sounds in apnea. *Physiological Measurement*, Vol. 26, No. 5, pp. 779–798, 2005.
- [11] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam. An efficient method for snore/nonsnore classification of sleep sounds. *Physiological Measurement*, Vol. 28, No. 8, pp. 841–854, 2007.
- [12] F. Dalmaso and R. Prota. Snoring: analysis, measurement, clinical implications and applications. *European Respiratory Journal*, Vol. 9, pp. 146–159, 1996.
- [13] G. Liistro, D. Stanescu, and C. Veriter. Pattern of simulated snoring is different through mouth and nose. *Journal of Applied Physiology*, Vol. 70, No. 6, pp. 2746–2741, 1991.

- 
- [14] T. Mikami, Y. Kojima, K. Yonezawa, M. Yamamoto, and M. Furukawa. Automatic classification of oral/nasal snoring sounds based on the acoustic properties. *Proc. IEEE 37th Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP)*, 2012.
- [15] T. Mikami, Y. Kojima, K. Yonezawa, M. Yamamoto, and M. Furukawa. Classification of oral/nasal simulated snores based on the acoustic properties. *Journal of Biomechanical Science and Engineering*, Vol. 7, No. 4, pp. 433–448, 2012.
- [16] O. Chapelle, P. Haffner, and V. N. Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, pp. 1055–1064, 1999.
- [17] P. J. Moreno and et al. A kullback-leibler divergence based kernel for svm classification in multimedia applications. *Proc. NIPS*, 2004.
- [18] T. Ishigaki, T. Higuchi, and K. Watanabe. Deterioration diagnosis of pressure regulator for high pressure gas by spectrum classification with the Kullback-Leibler kernel. *IEICE Transactions*, Vol. J90-D, No. 10, pp. 2787–2797, 2007.
- [19] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. E. Camilleri. Sound frequency analysis and the site of snoring in natural and induced sleep. *Clinical Otolaryngology*, Vol. 27, pp. 162–166, 2002.
- [20] A. K. Bieger-Farhan, N. K. Chadha, A. E. Camileri, P. Stone, and K. McGuinness. Portable method for the determination of snoring site by sound analysis. *Journal of Laryngology and Otology*, Vol. 118, No. 2, pp. 135–138, 2004.
- [21] M. Herzog, T. Metz, A. Schmidt, T. Bremert, W. Hosemann B. Venohr, and H. Kaftan. The prognostic value of simulated snoring in awake patients with suspected sleep-disordered breathing: Introduction of a new technique of examination. *Sleep*, Vol. 29, No. 11, pp. 1456–1462, 2006.

- 
- [22] 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. *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 12, pp. 1553–1562, 2008.
- [23] J. R. Perez-Padilla, E. Slawinski, L. M. Difrancesco, R. Feige, J. E. Remmers, and W. A. Whitelaw. Characteristics of the snoring noise in patients with and without occlusive sleep apnea. *American Review of Respiratory Disorders*, Vol. 147, No. 3, pp. 635–644, 1993.
- [24] R. Beck, M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *European Respiratory Journal*, Vol. 8, pp. 2120–2128, 1995.
- [25] T. Emoto, U. R. Abeyratne, T. Kusumoto, M. Akutagawa, E. Kondo, I. Kawata, T. Azuma, S. Konaka, and Y. Konouchi. Discriminating apneic snorers and benign snorers based on snoring formant extracted via a noise-robust linear prediction technique. *Journal of Japanese Society for Medical and Biological Engineering*, Vol. 48, No. 1, pp. 115–121, 2010.
- [26] E. J. Olson, W. R. Moore, and T. A. Staats. Obstructive sleep apnea-hypopnea. *Mayo Clinic Proceedings*, Vol. 78, No. 12, pp. 1545–1552, 2003.
- [27] Y. Inoue and Y. Yamashiro. *Sleep Disordered Breathing: update 2006 (in Japanese)*. Nippon Hyoronsha Inc., 2007.
- [28] M. Ichioka. Respiratory organs and sleep disorders. *Folia Pharmacologica Japonica (in Japanese)*, Vol. 129, pp. 432–435, 2007.
- [29] N. Cristianini and et al. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2004.
- [30] H. Murata, T. Onoda, K. Yoshimoto, Y. Nakano, and S. Kondo. Non-intrusive electric appliances load monitoring system -experiment for real household-

- 
- . *IEEJ Trans. Electronics, Information, and Systems*, Vol. 124, No. 9, pp. 1874–1880, 2004.
- [31] T. Horiuchi, T. Beppu, Y. Eujioka, and M. Hara. Comparison of pattern classification methods in discrimination of inferior shijimi clams based on acoustic signals. *Journal of Japanese Society of Fuzzy Theory and Intelligent Informatics*, Vol. 20, No. 5, pp. 817–822, 2008.
- [32] A. Karatzoglou and et al. kernlab - an S4 package for kernel methods in R. *Journal of Statistical Software*, Vol. 11, No. 9, 2004.
- [33] H-T. Lin and C-J Lin. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Technical report, National Taiwan University*, 2003.
- [34] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
- [35] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [36] K. Ishii, N. Ueda, E. Maeda, and H. Murase. *Pattern Recognition (in Japanese)*. Ohmsha, 1998.

# Chapter 5

## Conclusion

### 5.1 Contributions of this Thesis

This thesis analyzed the upper airway MR images where the enlargement of the upper airway soft tissues is identified, and the snoring sounds which occur when the inhalation airflow passes through the narrowed upper airway, so as to realize the automatic classification of OSAS.

Firstly, the upper airway MR images of OSAS patients are focused on and the usefulness of the morphological feature of the tongue region and the cross sectional area of the narrowest upper airway is indicated for the classification of severe OSAS. Concretely, it is clarified that the directions from the center of tongue region to the back of the tongue and to the hyoid bone are the most correlated features with AHI. As a result of classifying the severe OSAS using multi-regression analysis with those features, the true positive ratio is 0.909 and the false positive ratio is 0.476. Those features have an advantage for us to be able to evaluate OSAS more simply and more objectively than traditional methods such as visual examination of the oropharynx and X-ray image based cephalogram.

Next, the author focused on oral snoring found in many OSAS patients, pointed out the difference between the acoustic properties of oral snores and those of nasal ones based on the FFT amplitude spectra, and clarified the relation to the physiological mechanism of breathing from the bibliographical point of view. By using

those acoustic properties, about 90% of oral snoring sounds are successfully classified with  $k$ -Nearest Neighbor method. But there are some non-vibrational sounds which do not have a fundamental frequency, and those data cannot be classified to the correct category.

According to the acoustic properties of miss-classified data, the entire spectral shape below 2kHz is adopted as a multi-dimensional feature vector instead. As a result of using a SVM, the classification accuracy of oral snoring sounds is about 5% improved compared with the previous result. The result is discussed in the context of individual difference, reliability, computational time, and comparison with other classifiers and the good result is obtained in every point.

Based on the above result, the author quantified the morphological properties of the tongue region, the cross sectional area of the narrowest airway in the MR images, and the acoustic properties of oral snoring sounds which are typical symptoms of OSAS. In addition, he proposed the classification method and indicated the usefulness of the proposed method.

## 5.2 Future Works

In future works, some pre-processing before feature extraction is done manually in this study, and thus it should be improved to be automatic. Moreover, the enlargement of the soft palate or the uvula are also useful for OSAS classification. It is expected that the classification accuracy will further be improved if a concrete method to quantify the enlargement is also proposed.

As for snoring sounds, it is necessary to clarify actual snoring sounds in detail. Although Herzog and colleagues indicated that the acoustic properties of simulated snores are not so different from those of actual ones, this difference should be analyzed in more detail.

A hybrid method which combines image-based classification and sound-based classification will probably become a powerful technique compared with the meth-



ods proposed respectively in each chapter.