Role of Upper Airway Dimensions in Snore Production: Acoustical and Perceptual Findings

ANDREW KEONG NG,[1] TONG SAN KOH,[1] EUGENE BAEY,[2] and KATHIRAVELU PUVANENDRAN[3]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Republic of Singapore; [2]Philips Respironics Incorporated, Singapore 199591, Republic of Singapore; and [3]Sleep Disorders Unit, Singapore General Hospital, Singapore 169608, Republic of Singapore

Abstract—While considerable efforts have been expended to develop snore-driven markers for detecting obstructive sleep apnea (OSA), there is little emphasis on the relationship between the human upper airway (UA) dimensions and the attributes of snores. This paper aims to investigate the acoustical and perceptual impacts of changing the cross-sectional areas (CSA) of the pharynx and oral cavity on the production of snores. Synthetic snores were generated based on the source-filter theory, whereas natural snores were recorded from 40 snorers during nocturnal polysomnography. First formant frequency (F1), spectral peak frequency (PF), and psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) of CSA perturbations were examined, completed with diagnostic appraisal of F1 and PF for single- and mixed-gender groupings using the receiver operating characteristic curve analysis. Results show that (1) narrowing the pharyngeal airway consistently increases F1, but not for PF; and (2) altering the airway dimensions yield no considerable differences in perception of snore sounds, but indirectly affect the psychoacoustics by changing the dynamics of snore source flow. Diagnostic outcomes for all groupings (*p*-value < 0.0001) demonstrate that F1 is more capable of distinguishing apneic and benign snorers than PF due to the close association of F1 with the UA anatomical structures. Correlation exists between the UA anatomy and the properties of snores; there is a promising future for developing snore-driven screening tools for OSA.

Keywords—Obstructive sleep apnea, Snoring, Upper airway cross-sectional areas, Snore synthesis, Snore signals, Acoustical properties, Snore sounds, Psychoacoustic metrics.

INTRODUCTION

With the rising public awareness of obstructive sleep apnea (OSA)—a typical respiratory disorder, researchers have attempted various imaging methods (e.g., acoustic reflection technology,[19,26] computed tomography,[33] and magnetic resonance imaging[18]) to assess the upper airway (UA) anatomy for better understanding of OSA pathophysiology and to enhance treatment success.[39] These studies, in general, noted that patients with OSA tend to possess smaller pharyngeal cross-sectional area than those without OSA.

According to the Bernoulli's principle, when a constant airflow passes through the narrow pharynx, the airflow velocity increases because of mass conservation, while the intraluminal pressure decreases. This phenomenon further diminishes the pharyngeal airway size and promotes its occlusion.[9,40] Any imbalance of forces between the UA dilating muscle activity and the inspiration negative intraluminal pressure will cause soft tissues (e.g., soft palate, uvula, tongue base, and lateral pharyngeal walls) in the UA to vibrate and/or trigger an OSA attack.[16] In the light of snoring (a hallmark of OSA), fluttering vibrations of soft tissues and/or noise-like turbulent airflow at constrictions inevitably produce acoustic waves, which are then spectrally modified by the UA anatomical structures (e.g., cross-sectional airway dimensions) to create distinct sounds before reaching the listener.[10,30]

In recent years, multiple acoustic markers of snore signals have been proposed to discriminate between apneic and benign patients, with a common interest to develop a non-invasive, inexpensive, and rapid screening tool for OSA. These markers include, but are not limited to, first formant frequency (F1) in the linear prediction (LP) spectrum[30]; peak frequency (PF) in the fast Fourier transformation curve[25]; frequency modes in the bispectrum[32]; mean and standard deviation of the coefficient of variation in a short signal frame[7]; soft phonation index and noise-to-harmonics ratio[17]; and even psychoacoustic metrics in terms of loudness, sharpness, roughness, fluctuation strength, and annoyance.[29] Although the snore-driven markers

appear to shed light on OSA detection, there is little research on correlation between the UA dimensions and the properties of snores. Such correlation studies are undeniably crucial because they not only provide insights into different perspectives of snores from physiological to acoustical and perceptual domains, but also warrant the feasibility of using snore signals as an alternative to overnight polysomnography in diagnosing OSA.

The present paper aims to address this concern by investigating the acoustical and perceptual influences of changing the cross-sectional areas (CSA) of the pharynx (PX) and oral cavity (OC) on the generation of snores. To accomplish this task, we estimated the source flow for snore production, modeled the human UA, and synthesized snores by perturbing the CSA of PX and OC prior to determining the frequency markers (F1 and PF) and the psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) of these perturbations. To validate the acoustical findings, we evaluated the diagnostic accuracy of F1 and PF of natural snores for distinguishing apneic snorers from benign snorers in single-gender (males and females separately) and mixed-gender (both males and females combined) groupings. All analysis in this study was executed within MATLAB^{TM} (version 2006a), unless otherwise stated.

METHODS

Snore Source Flow Estimation

The soft tissue in the UA is a flow-induced self-excited biomechanical oscillator[6]; it acts as the main excitation source (ES) during the production of snores. Direct measuring of source flow (SF, airflow at snore ES) is a complicated process attributable to its difficulty of accessing the snoring site during natural (without anesthesia) nocturnal snoring. Nevertheless, if one applies the well-known source-filter theory[8,14,34] in snore generation by presuming that the ES and the UA are linearly separable, and that the SF is acoustically filtered by the UA and lip radiation to generate snoring sounds, a possible technique for estimating SF is the inverse filtering—a technique extensively applied in speech science to explore the nature of vocal fold excitation, indirectly and non-invasively.[1,8,14,34]

Figure 1 (left block diagram) depicts a flow chart of an iterative adaptive inverse filtering technique[1] for approximating SF. The function of this technique is to remove the effects of ES and UA from a snore signal through an iterative structure that repeats twice, with an aim to acquire better SF estimate as compared to a direct inverse filtering method. The technique robustness has been demonstrated in voice source analysis for the estimation of glottal flow,[1] and it consists of the following two phases. First, an initial estimate of SF is computed from a noise-free snore by eliminating the contributions of ES, UA, and lip radiation, correspondingly represented by a first-order and a 14th-order discrete all-pole (DAP) model, and a fixed differentiator. The first-order DAP model approximates the combined effect of −6 dB/octave of the ES (−12 dB/octave) and the lip radiation (+6 dB/octave), and it is the first to remove from the input signal through inverse filtering. Subsequently, the effect of UA, estimated by a 14th-order DAP model, is canceled from the input signal by inverse filtering prior to eliminating the lip radiation effect via integration (i.e., the resulting signal is filtered with the inverse of the differentiator). This concludes the first phase of inverse filtering and yields the initial SF estimate. Second, the SF is refined by repeating the steps in the first phase with the initial SF estimate as input and a second-order DAP model as the approximated contribution of ES.[1] Lip radiation effect is also removed by integrating the output of the previous step. Eventually, the final SF estimate is determined after removing the contributions of UA and lip radiation by inverse filtering the input signal with a 14th-order DAP model obtained in the second phase and by integrating the resulting signal, respectively.
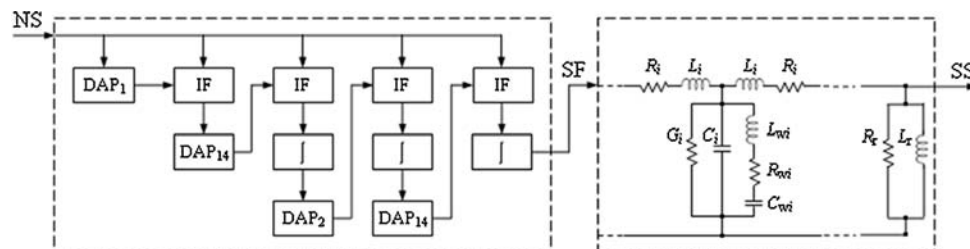
FIGURE 1. Flow charts of an iterative adaptive inverse filtering technique (left) and a transmission-line model of the upper airway for *i*th section and lip radiation load (right). NS refers to natural snore; SF, source flow; SS, synthetic snore; DAP$_p$, *p*th-order discrete all-pole model; IF, inverse filtering; ∫, integration; R_i, resistance; L_i, inertance; C_i, compliance; G_i, conductance; L_{wi}, wall inertance; R_{wi}, wall resistance; C_{wi}, wall compliance; R_r, radiation resistance; L_r, radiation inertance.

To provide simple yet reasonable approximations for the filter coefficients, the DAP modeling[13] is implemented in the iterative adaptive inverse filtering technique rather than the autoregressive moving average (ARMA) or the LP modeling. The estimation of ARMA parameters is tedious and involves nonlinear equations,[8] while the spectral peaks of LP are highly biased toward the pitch harmonics.[23] The order of DAP model for the UA defines the UA resonances, which is not possibly known owing to the complexity of the UA configuration. However, it can be estimated as follows[24]:

$$p \approx \frac{f_\mathrm{S}}{c/(2l)} + (2 \text{ to } 3) \qquad (1)$$

where p denotes the model order, and f_s denotes the signal sampling frequency in Hertz. The term c is the speed of sound, and l is the length of the UA. The extra 2 to 3 poles account for spectral tilt and provide spectral balance. With $f_\mathrm{s} = 11025$ Hz, $c \approx 35400$ cm/s for moist air at 37 °C, and possible UA lengths = 18.7 cm (from larynx to lips in adults),[10] 19.6 cm, and 17.7 cm (from mid-trachea portion to upper incisor in males and females, respectively),[35] we obtained the best estimate of p from Eq. (1) as 14, which justifies the use of a 14th-order DAP model to represent the UA.

Besides using Eq. (1), we evaluated several criteria[2,3,8,22,37] to select the DAP model order. These criteria include the final prediction error[2]

$$\mathrm{FPE}(p) = \sigma^2 \left(\frac{N + (p+1)}{N - (p+1)} \right), \qquad (2)$$

the Akaike information criterion[3]

$$\mathrm{AIC}(p) = N \log_\mathrm{e}(\sigma^2) + 2p, \qquad (3)$$

and the minimum description length[37]

$$\mathrm{MDL}(p) = N \log_\mathrm{e}(\sigma^2) + p \log_\mathrm{e}(N). \qquad (4)$$

In a nutshell, these criteria weight the prediction error variance σ^2, accompanied by the sample size N, and determine the model order that gives the minimum error value. Figure 2 plots the criterion values of three typical forms of snore signals observed in our database at $N = 512$ samples (≈ 46 ms); the snore waveforms are shown in Fig. 3 (top row). As can be seen from Fig. 2, the earlier derived $p = 14$ remains a favorable choice to model the UA attributed to its relatively low error values for all the criteria.

Upper Airway Acoustic Modeling

Since both snore and speech sounds propagate through the same medium, an UA model that can establish the anatomical–acoustical relationships is likely comparable to the transmission-line model of the lossy vocal tract, where wave propagation in an acoustic tube resembles plane wave propagation along
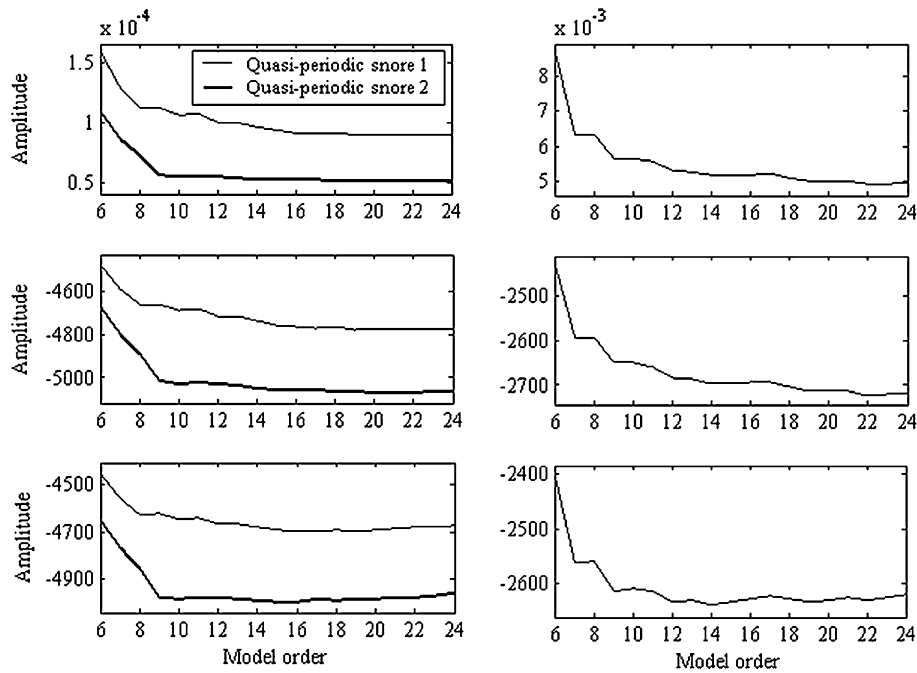

FIGURE 2. Performance of different model order selection criteria: final prediction error (top), Akaike information criterion (middle), and minimum description length (bottom) for two quasi-periodic snore signals (left) and an aperiodic snore signal (right).

an electrical transmission line.[8,14,34,41] A transmission-line model of the UA is rendered in Fig. 1 (right block diagram). The UA can be regarded as a concatenation of cylindrical sections whose lengths are much shorter than the acoustic wavelength of interest. Losses owing to viscous friction, heat conduction, and tissue vibration in each section, whose length was considered to be 0.1 cm, are represented by frequency-dependent lumped parameters,[8,14,34] as tabulated in Table 1. The acoustic resistance R_i arises from viscous and thermal losses at the boundaries, and the conductance G_i from heat conduction on the *i*th section walls. The inertance L_i is associated with the mass of air, and the compliance C_i is with the ability of air to expand and compress.

The series combination of mechanical wall impedance comprising a resistance R_{wi}, inertance L_{wi}, and compliance C_{wi} model the effect of wall vibration. In addition, radiation load at the lips is configured by a parallel connection of resistance R_r and inertance L_r to account for energy loss and mass inertia of air. The UA acoustic transfer function can be ultimately described by the product of all section transfer matrices.

Synthetic Snore Production

Figure 3 (top row) exhibits the common forms of snore signals; these signals can be mathematically formulated[4] and broadly characterized as
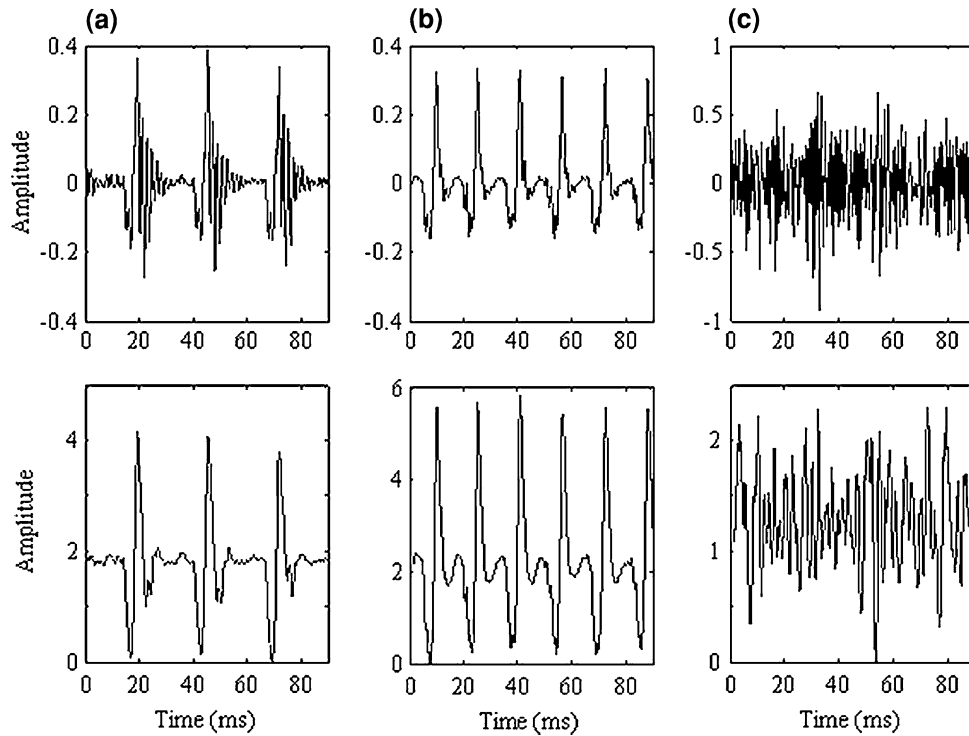

FIGURE 3. Snore signals (top) and their correspondingly source flow waveforms (bottom). (a) and (b) are quasi-periodic signals, while (c) is an aperiodic signal.

TABLE 1. Lumped parameters for *i*th section of UA model and lip radiation load.

Parameter	Expression	Parameter	Expression	Parameter	Expression
Resistance	$R_i = \frac{l_i S_i}{2A_i^2}\sqrt{\frac{\rho\mu\omega}{2}}$	Conductance	$G_i = l_i S_i \frac{\eta-1}{\rho c^2}\sqrt{\frac{\lambda\omega}{2\xi\rho}}$	Wall compliance	$C_{wi} = \frac{l_i S_i}{k}$
Inertance	$L_i = \frac{\rho l_i}{2A_i}$	Wall inertance	$L_{wi} = \frac{m}{l_i S_i}$	Radiation resistance	$R_r = \frac{128\rho c}{9\pi^2 A_m}$
Compliance	$C_i = \frac{l_i A_i}{\rho c^2}$	Wall resistance	$R_{wi} = \frac{b}{l_i S_i}$	Radiation inertance	$L_r = \frac{8\rho}{3\pi\sqrt{\pi A_m}}$

l_i, S_i, and A_i refer to length, circumference, and cross-sectional area of the *i*th cylindrical section, respectively; ω, radian frequency; ρ, density of air $= 1.14 \times 10^{-3}$ g cm^{-3} (moist air 37 °C); c, sound velocity $= 3.54 \times 10^4$ cm s^{-1} (moist air 37 °C); μ, shear viscosity $= 1.86 \times 10^{-4}$ dyne s cm^{-2} (20 °C, 1 atm); λ, heat conduction coefficient of air $= 5.5 \times 10^{-5}$ cal cm^{-1} s^{-1} °C^{-1} (0 °C); η, adiabatic gas constant $= 1.4$; ξ, specific heat of air at constant pressure $= 0.24$ cal·g^{-1} °C^{-1} (0 °C, 1 atm); m, b, and k are the mass $= 1.5$ g, mechanical resistance $= 1.6 \times 10^3$ g s^{-1}, and stiffness $= 3 \times 10^4$ dyne cm^{-1} of wall per unit area, respectively; A_m, cross-sectional area of mouth, which was deemed as the area of the last section. The length of each section was considered to be 0.1 cm.

quasi-periodic or aperiodic. Upon the iterative adaptive inverse filtering, the quasi-periodic signals yield rhythmically repeating SF waveforms (Figs. 3a and 3b, bottom row). Strong peaks in the waveforms reveal the periodic nature of the rate at which the ES maneuvers (fundamental frequency). The first quasi-periodic SF, designated as SF1 (Fig. 3a, bottom row), has a longer plateau of airflow and a lower fundamental frequency computed from the YIN algorithm[11] than the other quasi-periodic SF (SF2): 38 vs. 64 Hz. Conversely, the SF of the aperiodic snore signal, designated as SF3 (Fig. 3c, bottom row), is random with high-frequency chaotic oscillations and no fundamental frequency. These SFs were selected as excitations for three different types of synthetic snores, one for each snore type.

Apart from the SF selections, two UA area-distance profiles were chosen from the works of Jung et al.[19] and Mohsenin[26] for acoustic modeling. The profiles were constructed by the acoustic reflection technology and intended as reference UA model 1 and 2, respectively. The former has a shorter UA (from glottis to incisors) length than the latter (around 16 vs. 18.5 cm), possibly due to different subjects and measurement procedures. Alterations in the CSA of PX (from glottis to soft palate) and OC (from anterior margin of oropharyngeal junction to incisors) were made in ± 0.2 and ± 0.4 cm^2 from the reference models as displayed in Fig. 4. For instance, PX $+0.4$ cm^2 and OC -0.4 cm^2 specify a CSA increment of PX by 0.4 cm^2 and a decrement of OC by 0.4 cm^2 with respect to the reference, which in turn illustrates a subject with a smaller pharyngeal airway but wider mouth opening during snoring.

Using the above resources, snores were synthesized by convolving the SF and the UA transfer function (with lip radiation load) based on the source-filter theory.[8,14,34] For each UA model, 12 CSA perturbations were performed, thereby giving 39 synthetic snores (including reference-generated snores) of roughly 1.5 s each for the subsequent analysis.

Natural Snore Recording and Preprocessing

Forty snorers referred to the sleep clinic (Sleep Disorders Unit, Singapore General Hospital) with suspected OSA were studied. They underwent a nocturnal polysomnography and were scored manually following the Rechtschaffen and Kales criteria.[36] Thirty of the snorers (24 males; 6 females; mean $\pm$ standard deviation of age = 44 $\pm$ 13 years; body mass index, BMI = 29.3 $\pm$ 6.9 kg/m^2) were diagnosed having an apnea-hypopnea index (AHI) of at least 10 events/h (AHI = 46.9 $\pm$ 25.7, range 11.6–101.9 events/h), whereas the rest (6 males; 4 females; age = 41 $\pm$ 12 years; BMI = 26.9 $\pm$ 5.6 kg/m^2) have an AHI less than 10 events/h (AHI = 4.6 $\pm$ 3.4, range 0.2–8.9 events/h). The former was categorized as apneic snorers, and the latter as benign snorers. Written informed consent was acquired from the subjects, and ethical approval was obtained from the local Institutional Review Board.

Snoring sounds were recorded concurrently with PSG using a robust sound acquisition system in the sleep clinic.[28] A unidirectional microphone (20–20000 Hz, model SM81, Shure Incorporated) was hung from a microphone boom stand about 0.3 m above the subject's mouth[27] to capture sleep sounds (e.g., snoring, somniloquy, and body movements). The sounds were then amplified via a low-noise preamplifier (20–22000 Hz, model FP23, Shure Incorporated) and transmitted through a double-shielded coaxial cable to
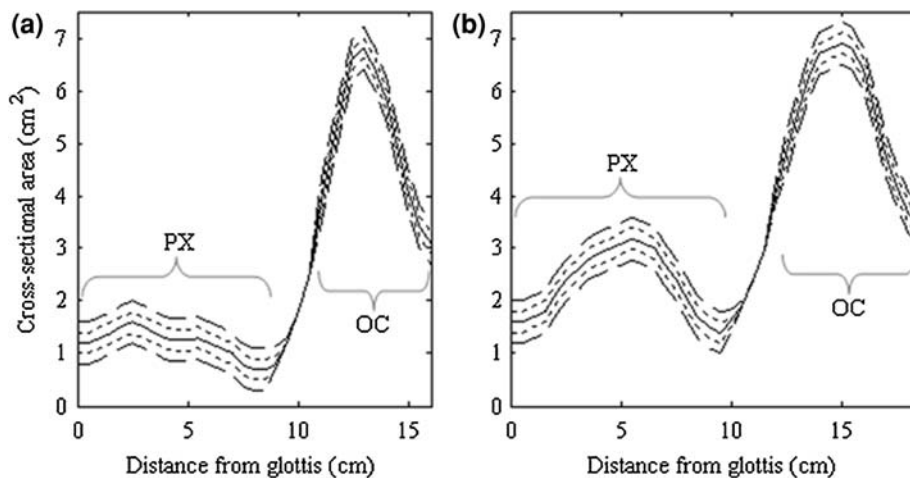
FIGURE 4. Upper airway (UA) area-distance profiles with changes of cross-sectional areas of pharynx (PX) and oral cavity (OC) at ± 0.2 cm^2 (dotted lines) and ± 0.4 cm^2 (dashed lines) from (a) reference UA model 1 (solid line); (b) reference UA model 2 (solid line).

avoid electromagnetic interferences. Eventually, the collected signals were sampled by a data acquisition card (44100 Hz sampling rate, 16-bit resolution, model NI4552, National Instruments) and stored in a database for preprocessing.

A wavelet-based preprocessing system was utilized to further enhance the signal quality and intelligibility, along with the detection of sleep sounds onset within a translation-invariant wavelet transform domain.[31] The system effectively cancels background acoustical noise (e.g., whistling sound from air conditioner) embedded in the raw signals by an improved wavelet thresholding approach (a level-correlation-dependent threshold with hard thresholding) and simultaneously discerns sleep sounds from silence or noise through a sound activity detector.

After the preprocessing stage, 40 inspiratory snores of an average root-mean-square power of -27 ± 6 dBFS (decibels, full-scale sine wave) were selected from each subject over a mean of 6.5 h continuous recording via a digital audio software package (Cool Edit Pro^{TM} version 1.2). To lessen computational cost of the subsequent analysis, the snores were low-pass filtered by means of an eighth-order zero-phase Butterworth filter with a cutoff frequency of 5000 Hz and then downsampled to a sampling rate of 11025 Hz. This new sampling rate should not introduce any aliasing problem since the maximum frequency of interest for snore signals is approximately 5000 Hz.[10]

Acoustical and Perceptual Analysis

Snore signals, both synthetic and natural, were analyzed using the DAP[13] and the Welch's averaged modified periodogram methods.[38] F1 and PF were respectively extracted from the 14th-order DAP spectrum and the Welch power spectrum. To avoid spectral leakage attributed to edge discontinuity, the signals were Hanning windowed using 512-sample (≈ 46 ms) frames with 75% overlap between successive frames.

Besides the acoustical analysis, six pairs of synthetic snores from each UA model were subjectively experimented by a group of 16 polysomnographic (sleep studies) technicians and signal processing specialists through a paired comparison approach,[15] one at a time. For each pair of snore sounds, the listener was requested to compare one sound (comparison sample) with the other (reference sample), and then rate the comparison sample in terms of loudness, sharpness, roughness, fluctuation strength, and annoyance on a 5-point scale with bipolar adjective pairs before proceeding to the next pair. For example, in the case of loudness, the value '1' in the scale symbolises very soft, '2': somewhat soft, '3': equal loudness, '4': somewhat loud, and '5': very loud. If the listener perceived the comparison sample as somewhat soft, a value of '2' will be assigned to it, and automatically a value of '4' will be assigned to the reference sample, implying that it is somewhat loud. Thus, the degree of loudness for the comparison sample and the reference sample are '2' and '4,' correspondingly. In contrast, if the listener judged both the samples as equal loudness, a value of '3' will be allocated to the samples, and the degree of loudness for the samples is '3.'

Statistical Analysis

Among the 40 natural snores from each subject, the first 30 were considered as training data to calculate F1 and PF, as well as to derive optimal cutoff values of F1 and PF using the receiver operating characteristic (ROC) curve analysis[42] in a statistical software package (MedCalc^{TM} version 9.3.6.0). Furthermore, area under the ROC curve (AUC), standard error of the AUC (SE), and *p*-value (two-tailed *z*-test) were computed, with *p*-value < 0.05 considered statistically significant. The remaining 10 snores were considered as test data to yield sensitivity and specificity of the derived cutoff values. In total, there were 900 apneic snores (AS, snores from apneic subjects) and 300 benign snores (BS, snores from benign subjects) for training, together with 300 AS and 100 BS for testing.

RESULTS

Acoustical Influences of Area Perturbations

Table 2 summarizes the values of F1 and PF for perturbations that lie within ± 0.4 cm^2 from the reference models. The key influence of decreasing the CSA of PX, while keeping constant or widening the CSA of OC, is an increase of F1. Correspondingly, an increase in the CSA of PX, while maintaining constant or narrowing the CSA of OC, can reduce F1. The variations of F1 with changing CSA are not as drastic as those of PF; however, they are more consistent and predictable for rise or fall regardless of SF types and UA models. In model 1 for SF2, when PX alters from -0.4 to $+0.4$ cm^2 in parallel with OC from $+0.4$ to -0.4 cm^2, both F1 and PF decline progressively from 708 to 659 Hz and from 701 to 655 Hz, respectively. However, under the same CSA configuration for SF2 in model 2, F1 drops from 620 to 577 Hz, but PF fluctuates between 638 and 717 Hz rather than decreases as before. This undeterminable behavior of PF is also found in snore signals synthesized using SF1 but not those using SF3. A close inspection of Table 2 also highlights that increasing the CSA of OC or shortening the UA length may boost the values of F1.

Acoustical Analysis of Natural Snores

To affirm the above findings, diagnostic performances of F1 and PF of the natural snores for three different groupings (males, females, and both males and females combined) are presented in Table 3. For all the groups, results persistently suggest that the F1 can better identify AS from BS than the PF (sensitivity 88.0–100 vs. 62.5–91.7%, specificity 83.3–92.5 vs. 78.0–97.5%). While all statistical tests of F1 and PF achieve significant associations (p-value < 0.0001), the optimal cutoff value of F1 for every group delivers larger AUC and smaller SE, emphasizing its competency in classifying apneic and benign snorers whose UA anatomical structures are distinctively different, as regularly reported in the literatures.[18,19,26,33,39] For mixed-gender grouping: cutoff value of F1 = 497 Hz, AUC = 0.8826, and SE = 0.0095; cutoff value of PF = 243 Hz, AUC = 0.8537, and SE = 0.0108. Moreover, the values

of F1 are typically higher in AS than BS and are closer to mean values, whereas the values of PF are widely spread out, particularly those in AS.

Perceptual Influences of Area Perturbations

Table 4 lists the degree (range 1–5) of each psychoacoustic metric (loudness, sharpness, roughness, fluctuation strength, and annoyance) of snore sounds involved in the paired comparison experiments. Most listeners feel that the sound quality metrics in response to the CSA perturbations are not pronounced. The degree ratings for each pair of snore sounds with same SF, when compared across different SFs, are confounded, except for those tailored to investigate the effects of varying CSA of OC, which indicate no sound change. On the contrary, when we compared snore sounds of different SFs with the same CSA

TABLE 2. First formant frequency (F1) and spectral peak frequency (PF) of snore signals synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC) for upper airway (UA) model 1 and 2.

		UA model 1						UA model 2					
ΔCSA (cm^2)		SF1		SF2		SF3		SF1		SF2		SF3	
PX	OC	F1	PF	F1	PF	F1	PF	F1	PF	F1	PF	F1	PF
0	0	557	572	691	687	686	680	565	560	600	639	586	624
−0.4	0	595	572	703	694	703	698	601	582	609	654	604	786
−0.2	0	564	589	700	691	698	693	576	569	605	637	593	675
+0.2	0	556	580	677	673	679	669	556	557	595	678	578	668
+0.4	0	538	611	663	660	667	654	550	553	585	676	565	553
0	−0.4	557	575	683	679	681	676	554	558	601	644	580	607
0	−0.2	560	571	689	685	683	679	562	559	601	650	587	610
0	+0.2	553	583	694	687	690	687	566	562	602	644	589	624
0	+0.4	562	595	698	692	694	692	568	566	603	623	596	621
−0.4	+0.4	597	593	708	701	708	701	628	612	620	650	609	790
−0.2	+0.2	569	582	702	691	702	697	578	572	608	638	601	786
+0.2	−0.2	548	588	671	668	677	665	552	558	583	717	566	556
+0.4	−0.4	534	576	659	655	660	647	550	545	577	652	560	546

Values are presented as mean. First row indicates F1 and PF in Hertz for reference UA model 1 and 2 at their respective CSA, while others indicate F1 and PF under various changes in CSA from the reference. ΔCSA refers to change in CSA perturbations (cm^2).

TABLE 3. Diagnostic performance of first formant frequency (F1) and spectral peak frequency (PF) of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

		F1				PF			
Type	AHI	V	CV (AUC; SE)	Sens	Spec	V	CV (AUC; SE)	Sens	Spec
M:A	52.8 ± 24.7	652 ± 115	513 (0.8407; 0.0114)*	88.3	83.3	604 ± 413	333 (0.8163; 0.0124)*	62.5	85.0
M:B	5.3 ± 3.5	440 ± 170				248 ± 159			
F:A	23.6 ± 15.1	623 ± 112	327 (0.9478; 0.0059)*	100	92.5	561 ± 554	213 (0.9316; 0.0069)*	91.7	97.5
F:B	3.4 ± 3.2	231 ± 151				173 ± 100			
C:A	46.9 ± 25.7	646 ± 115	497 (0.8826; 0.0095)*	88.0	84.0	596 ± 445	243 (0.8537; 0.0108)*	77.7	78.0
C:B	4.6 ± 3.4	356 ± 192				218 ± 143			

Values (V) are presented as mean ± standard deviation in Hertz (Hz). AHI refers to apnea-hypopnea index (events/h); CV, cutoff value (Hz); AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity (%); Spec, specificity (%); * p-value < 0.0001.

TABLE 4. Psychoacoustic metrics in terms of loudness (L), sharpness (S), roughness (R), fluctuation strength (F), and annoyance (PA) of snore sounds synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC).

ΔCSA (cm²)		SF1					SF2					SF3				
PX	OC	L	S	R	F	PA	L	S	R	F	PA	L	S	R	F	PA
0	0	1.4	1.5	1.7	1.7	1.5	3.2	2.8	3.1	3.4	2.9	4.4	4.6	4.2	3.9	4.6
−0.4	0	2.8	2.9	3.1	3.1	3.1	3.0	3.3	3.3	2.8	3.0	3.3	3.1	3.4	3.1	3.3
+0.4	0	3.2	3.1	2.9	2.9	2.9	3.0	2.7	2.7	3.2	3.0	2.7	2.9	2.6	2.9	2.7
0	−0.4	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
0	+0.4	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
−0.4	+0.4	2.7	2.8	3.0	2.8	2.9	3.0	2.7	3.0	3.0	2.8	3.3	3.0	3.3	3.3	3.2
+0.4	−0.4	3.3	3.2	3.0	3.2	3.1	3.0	3.3	3.0	3.0	3.2	2.7	3.0	2.7	2.7	2.8

Values are presented as mean for upper airway model 1 and 2 combined. First row indicates the degree (range 1 to 5, with '1' being least and '5' being most) of the metrics after comparing snore sounds with different SFs, while others indicate the degree of the metrics for each pair of snore sounds (row second and third, fourth and fifth, sixth and seventh) with same SF. ΔCSA refers to change in CSA perturbations (cm²).

configuration, the degree ratings are conclusive. Sounds produced by SF3, an aperiodic waveform, yield highest ratings for all the metrics (degree range 3.9–4.6), followed by the two quasi-periodic waveforms, SF2 (degree range 2.8–3.4), and SF1 (degree range 1.4–1.7).

DISCUSSION

This study demonstrates that changes in the CSA of PX and OC have different implications for generating snores, more acoustically than perceptually: (1) narrowing of the pharyngeal airway with mouth opening can increase the values of F1 and occasionally PF of snore signals, being higher for larger mouth opening; and (2) increasing or decreasing the dimensions of pharyngeal airway and mouth opening can vaguely alter the snore sound quality metrics, but these CSA changes may indirectly influence the metrics by altering the waveforms of SF. Hence, one can deduce that the UA anatomical structures are the aggravating factors in snore production.

The frequency marker F1 of snore signals has considerably proven to be more sensitive to changes in UA geometry than PF, which is in agreement with the previous speech research studies[8,12,21] recognizing the associations between F1 and constrictions in PX and OC: the greater the pharyngeal constriction or the lower the oral constriction, the higher is F1.[12] For instance, the PX is more constricted and the lips are more broadly opened for the vowel /a/ than /i/; therefore, the value of F1 for /a/ is higher than that of /i/, 710 vs. 280 Hz.[21] In addition, a substantial reduction of F1 for /i/ was found after increasing the oropharyngeal cavity space by tongue advancement (anterior–posterior position) and enlarging the palatal arch by uvulopalatopharyngoplasty and/or tonsillectomy.[5]

With the earlier establishment of anatomical–acoustical relationships, one can infer that F1 is superior to PF in detecting OSA. Apneic snorers will likely yield snores of higher F1 than that of benign snorers since they have smaller pharyngeal airway and spend more time on oral or oro-nasal breathing than benign ones,[20] thus further reducing the airway size as a result of jaw opening. The earlier diagnostic appraisal of F1 and PF lend support to the claims. It is also noteworthy that the values of F1 calculated by the DAP algorithm in this study are probably more reliable than those computed by the LP algorithm[30] because the DAP modeling is capable of determining an optimal all-pole filter, providing better spectral estimates that are less biased toward the pitch harmonics, by means of the Itakura-Saito distortion criterion.[13]

The integrity of snore sound qualities (loudness, sharpness, roughness, fluctuation strength, and psychoacoustic annoyance), on the other hand, are not easily affected by the changes of cross-sectional airway dimensions. Apart from the limited availability of listeners, another possible explanation for the perceptual findings is that the spectral contents (e.g., envelope, magnitude, and formant bandwidth) of the acoustic transfer functions of UA models with or without CSA perturbations yield no considerable discrepancy. As evident in Fig. 5, the transfer functions of reference UA models and those after CSA perturbations are somewhat similar, except for their spectral envelopes that are slightly shifted sideways. Conversely, a change in SF types can bring about a vast change in the perception of snore sounds.

The dynamics of SF (e.g., amplitude envelope, fundamental frequency, and harmonicity) contribute an important part to the psychoacoustics of snore sounds since each sound quality metrics can be anticipated from acoustic measurements.[15] Loudness correlates with amplitude; sharpness depends on center
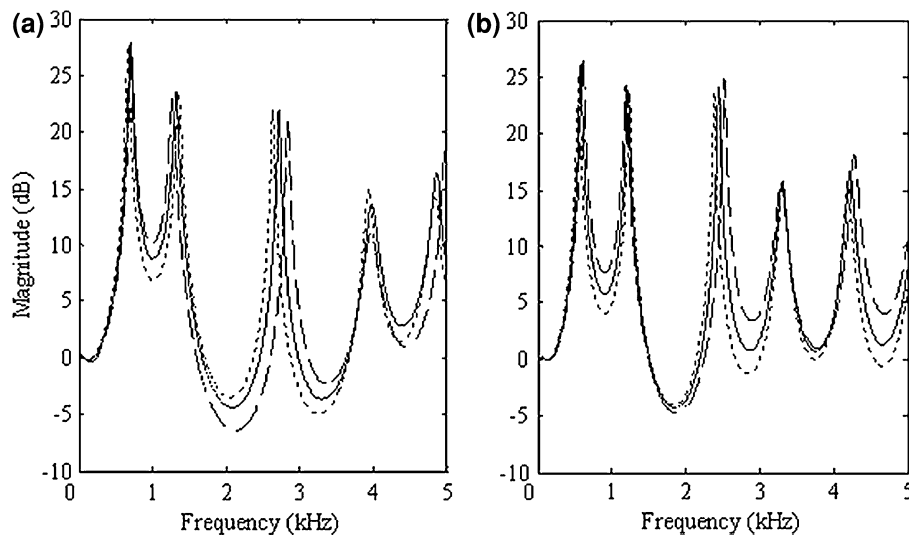
FIGURE 5. Acoustic transfer functions of (a) reference upper airway (UA) model 1 (solid line); (b) reference UA model 2 (solid line), as well as transfer functions of models with cross-sectional areas (CSA) of pharynx (PX) at -0.4 cm^2 and oral cavity (OC) at $+0.4$ cm^2 from the references (dashed lines), and models with CSA of PX at $+0.4$ cm^2 and OC at -0.4 cm^2 from the references (dotted lines).

frequency and bandwidth; roughness and fluctuation strength govern by temporal variations of high-frequency and low-frequency, respectively; and annoyance relies on the above four metrics. Relating this to the mechanisms of snoring,[4,10,16] a narrower airway escalates the negative intraluminal pressure, and hence enhances the driving forces on the vibrating soft tissues or the rate of pharyngeal airway closure and reopening. Spontaneously, the complexity of SF dynamics rises (e.g., amplifying the amplitude, increasing the fundamental frequency, and introducing higher harmonics), and the resulting sound waves are spectrally modified by the acoustic transfer function of UA, producing snoring sounds of build-up loudness, sharpness, roughness, fluctuation strength, and/or psychoacoustic annoyance. This perhaps explains why AS (mostly aperiodic signals originated from tongue base and below) and BS (mostly quasi-periodic signals originated from soft palate and/or uvula) could be differentiated using the psychoacoustics of snore sounds.[29]

The biophysics of snore production is fundamentally complex due to the involvements of different soft tissues with inhomogeneity compositions and the dynamic UA anatomical structures. Furthermore, it is difficult to apprehend the fluid–structure interactions[9,40] and interplay features[16] leading to snoring with no assumption or simplification. Consequently, limitations exist in this study. The analytical results from the synthetic snores require further justification on the methodology for snore synthesis and its relationship with physiological and physical reality. Explicitly, we assumed that the snore ES and the UA

are linearly separable, in accordance with the well-established source-filter theory.[8,14,34] This theory serves as a basic principle underlying numerous applications in speech technology because a non-interactive source-filter model is simple and the errors introduced by the assumption is approximately negligible in many cases. However, the assumption may not strictly hold for the case of snore sounds. The vibration of soft tissue can be affected by the sound pressure within the UA, entailing a certain degree of coupling between the ES and the UA when the flabby tissue reopens from its shut position. Nevertheless, the coupling may be weakened when the tissue is at its closing position, thereby favoring the independence assumption of the ES and the UA, which makes the processing of snore signals tractable. Further study should consider the effect of source-filter interaction caused by the coupling and model the UA using an ARMA filter.

Apart from the assumption made, we changed the CSA of PX and OC, without taking into account that such changes would also lead to modifications of UA dimensions in other anatomical regions. We also stimulated the acoustic waves propagating through the entire UA, but neglected the fact that these waves could be initiated in any locations where there is an imbalance of dilating and collapsing forces within the UA. However, based on the results obtained from the three SFs and the two clinical UA profiles with different UA lengths, one may fairly well predict that there is an association between the UA anatomy and the acoustical and perceptual properties of snores, and that F1 is more sensitive to changes in CSA than PF.

To further improve naturalness of snore synthesis, one could also incorporate the following components into the existing UA model: (1) acoustic parameters to account for tissue biomechanical properties and UA neuromuscular activities; (2) one or more snore ES circuits to generate SF (e.g., a self-oscillating soft palate circuit); and (3) an acoustic side branch to signify the nasal cavity. Moreover, further study should involve a larger sample of subjects, possibly with matched BMI and AHI, as well as absolutely independent training and test datasets comprising of dissimilar subjects in order to better appraise the diagnostic accuracy of F1 and PF.

Despite the limitations, this preliminary study demonstrates the relationships between the underlying UA anatomy and the attributes of snores, warranting the likelihood of using acoustical signatures in snores to classify snorers with and without OSA. We believe that further research in this direction would significantly contribute to the understanding of the pathophysiology of snoring, the development of reliable snore-driven screening tools for OSA, and the implementation of new therapies for snoring.

ACKNOWLEDGMENTS

REFERENCES

[1] Airas, M. TKK Aparat: an environment for voice inverse filtering and parameterization. *Logoped. Phoniatr. Vocol.* 33(1):49–64, 2008.

[2] Akaike, H. Power spectrum estimation through autoregression model fitting. *Ann. Inst. Stat. Math.* 21:407–419, 1969.

[3] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:719–723, 1974.

[4] Beck, R., M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *Eur. Respir. J.* 8(12):2120–2128, 1995.

[5] Bertino, G., E. Matti, S. Migliazzi, F. Pagella, C. Tinelli, and M. Benazzo. Acoustic changes in voice after surgery for snoring: preliminary results. *Acta Otorhinolaryngol. Ital.* 26(2):110–114, 2006.

[6] Bertram, C. D. Flow-induced oscillation of collapsed tubes and airway structures. *Respir. Physiol. Neurobiol.* 163(1–3):256–265, 2008.

[7] Cavusoglu, M., T. Ciloglu, Y. Serinagaoglu, M. Kamasak, O. Erogul, and T. Akcam. Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification. *Physiol. Meas.* 29(8):879–898, 2008.

[8] Childers, D. G. Speech Processing and Synthesis Toolboxes. New Yrok: Wiley, 1999.

[9] Chouly, F., A. Van Hirtum, P. Y. Lagree, X. Pelorson, and Y. Payan. Modelling the human pharyngeal airway: validation of numerical simulations using in vitro experiments. *Med. Biol. Eng. Comput.* 47(1):49–58, 2009.

[10] Dalmasso, F., and R. Prota. Snoring: analysis, measurement, clinical implications and applications. *Eur. Respir. J.* 9(1):146–159, 1996.

[11] de Cheveigne, A., and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *JAS* 111(4):1917–1930, 2002.

[12] Deller, J. R., J. H. L. Hansen, and J. G. Proakis. Discrete-Time Processing of Speech Signals. NJ: Wiley-IEEE Press, 1999.

[13] El-Jaroudi, A., and J. Makhoul. Discrete all-pole modeling. *IEEE Trans. Signal Process.* 39(2):411–423, 1991.

[14] Fant, G. Acoustic Theory of Speech Production. The Hague: Mouton, 1960.

[15] Fastl, H., and E. Zwicker. Psychoacoustics: Facts and Models. New York: Springer, 2007.

[16] Gavriely, N., and O. Jensen. Theory and measurements of snores. *J. Appl. Physiol.* 74(6):2828–2837, 1993.

[17] Hara, H., N. Murakami, Y. Miyauchi, and H. Yamashita. Acoustic analysis of snoring sounds by a multidimensional voice program. *Laryngoscope* 116(3):379–381, 2006.

[18] Ikeda, K., M. Ogura, T. Oshima, H. Suzuki, S. Higano, S. Takahashi, H. Kurosawa, W. Hida, H. Matsuoka, and T. Takasaka. Quantitative assessment of the pharyngeal airway by dynamic magnetic resonance imaging in obstructive sleep apnea syndrome. *Ann. Otol. Rhinol. Laryngol.* 110(2):183–189, 2001.

[19] Jung, D. G., H. Y. Cho, R. R. Grunstein, and B. Yee. Predictive value of kushida index and acoustic pharyngometry for the evaluation of upper airway in subjects with or without obstructive sleep apnea. *J. Korean Med. Sci.* 19(5):662–667, 2004.

[20] Koutsourelakis, I., E. Vagiakis, C. Roussos, S. Zakynthinos, and C. Roussos. Obstructive sleep apnoea and oral breathing in patients free of nasal obstruction. *Eur. Respir. J.* 28(6):1222–1228, 2006.

[21] Ladefoged, P. A Course in Phonetics. Belmont, CA: Wadsworth Publishing, 2005.

[22] Ljung, L. System Identification: Theory for the User. Upper Saddle River, NJ: Prentice Hall PTR, 1998.

[23] Makhoul, J. Linear prediction: a tutorial review. *Proc. IEEE* 63(4):561–580, 1975.

[24] Markel, J. D., and A. H. Gray. Linear Prediction of Speech. New York: Springer-Verlag, 1976.

[25] Michael, H., S. Andreas, B. Thomas, H. Beatrice, H. Werner, and K. Holger. Analysed snoring sounds correlate to obstructive sleep disordered breathing. *Eur. Arch. Otorhinolaryngol.* 265(1):105–113, 2008.

[26] Mohsenin, V. Effects of gender on upper airway collapsibility and severity of obstructive sleep apnea. *Sleep Med.* 4(6):523–529, 2003.

[27] Ng, A. K., Z. J. Ho, and T. S. Koh. Effects of microphone position on snore signal quality and patient comfort. In: Proceedings of the International Conference on Signal Processing, pp. 2130–2133, 2008.

[28] Ng, A. K. and T. S. Koh. A high-fidelity acquisition system for snore signals: design and implementation. In: Proceedings of the IEEE Instrumentation and Measurement Technology Conference, pp. 189–193, 2007, art. no. 4258108.

[29]Ng, A. K. and T. S. Koh. Using psychoacoustics of snoring sounds to screen for obstructive sleep apnea. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 1647–1650, 2008.

[30]Ng, A. K., 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 Med.* 9(8):894–898, 2008.

[31]Ng, A. K., T. S. Koh, K. Puvanendran, and U. R. Abeyratne. Snore signal enhancement and activity detection via translation-invariant wavelet transform. *IEEE Trans. Biomed. Eng.* 55(10):2332–2342, 2008.

[32]Ng, A. K., K. Y. Wong, C. H. Tan, and T. S. Koh. Bispectral analysis of snore signals for obstructive sleep apnea detection. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 6196–6199, 2007.

[33]Ogawa, T., R. Enciso, W. H. Shintaku, and G. T. Clark. Evaluation of cross-section airway configuration of obstructive sleep apnea. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 103(1):102–108, 2007.

[34]Owens, F. J. Signal Processing of Speech. London: Macmillan, 1993.

[35]Park, S., C. Kim, and I. Jung. Metric study of upper airway and trachea in normal Korean adults using fiberoptic bronchoscopy: study of endotracheal tube fixation positioning in adults. *Korean J. Anesthesiol.* 31(6):733–738, 1996.

[36]Rechtschaffen, A. and A. Kales. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Los Angeles: Brain Information Service/Brain Research Institutes, University of California, 1968.

[37]Rissanen, J. A universal prior for integers and estimation by minimum description length. *Ann. Stat.* 11(2):416–431, 1983.

[38]Stoica, P., and R. L. Moses. Spectral Analysis of Signals. Upper Saddle River, NJ: Prentice Hall, 2005.

[39]Stuck, B. A., and J. T. Maurer. Airway evaluation in obstructive sleep apnea. *Sleep Med. Rev.* 12(6):411–436, 2008.

[40]Sun, X., C. Yu, Y. Wang, and Y. Liu. Numerical simulation of soft palate movement and airflow in human upper airway by fluid-structure interaction method. *Acta Mechanica Sinica/Lixue Xuebao* 23(4):359–367, 2007.

[41]Zhang, Z., and C. Y. Espy-Wilson. A vocal-tract model of American English /l/. *JAS* 115(3):1274–1280, 2004.

[42]Zweig, M. H., and G. Campbell. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39(4):561–577, 1993.