An Analysis-by-Synthesis Approach to the Estimation of Vocal Cord Nodule Features

Takuya Koizumi, Shuji Taniguchi, Mikio Mori, and Akemi Imazawa
Dept. of Information Science, Fukui University
3-9-1 Bunkyo, Fukui 910, Japan

ABSTRACT

This paper deals with a new non-invasive method of estimating vocal cord nodule features through hoarse voice analysis. A noteworthy feature of this procedure is that it enables us to estimate vocal cord nodule features such as the mass and dimensions of nodules through the use of a novel model of pathological vocal cords which has been devised to simulate the subtle movement of the vocal cords with nodules.

INTRODUCTION

A number of studies on acoustic analysis of hoarse voice which have been reported in the last decade may be divided into two groups by their objectives. Some studies tried to establish techniques of discriminating voices caused by pathological vocal cords from normal voices, while others aimed at developing a procedure for the classification of laryngeal disease only through the analysis of hoarse voice. However, few of them have ever attempted to develop a non-invasive procedure for estimating the pathological condition of the larynx [1].

An attempt toward this end has recently been made, resulting in the development of a new noninvasive procedure for the estimation of vocal cord nodule features through hoarse voice analysis. A noteworthy feature of this procedure is that it enables us to estimate vocal cord nodule features such as the mass and dimensions of nodules through the use of a novel model of pathological vocal cords which has been devised to simulate the subtle movement of the vocal cords with nodules. This method, which might well be called an analysis-by-synthesis approach, is characterized by a fact that a hoarse-voice synthesizer comprising the model of pathological vocal cords and a vocal tract model is used to estimate the nodule features along with an acvoustic distance measure for hoarse voices defined as a function of glottal volume flow waveform and power spectral density of glottal turbulent noise.

A synthetic hoarse voice produced with this hoarse-voice synthesizer is compared with a natural hoarse voice produced by pathological vocal cords with nodules in terms of the distance measure, and values of the nodule features that minimize the distance measure are chosen as reasonable estimates of them. The effectiveness of this method can be examined by comparing the estimates of dimensions of nodules with actual dimensions of those nodules. Some estimates of nodule dimensions that have been obtained by applying the procedure to hoarse voices of patients who have laryngeal nodules are found to compare favorably with actual nodule dimensions, demonstrating that the procedutre is effective.

THE HOARSE-VOICE SYNTHESIZER

An Eight-mass Model of Pathological Vocal Cords

The hoarse-voice synthesizer cosists of a vocal tract model and a model of pathological vocal cords. The pathological vocal cords can be modeled as a mechanical vibration system made up of four independent masses coupled by nonlinear springs and dampers. To properly model the vocal cords with a couple of nodules, however, it is necessary to add a couple of split masses representing the nodules to the four mass model. This results in an eight-mass model of the vocal cords with nodules as shown in Figure1. A brief description of this model will be given below. The upper and lower edges of the vocal cords are represented by the masses m_1, m_2 and M_1, M_2, respectively.

The split masses m_{iN} and M_{iN} ($i=1,2$, $m_{1N}=m_{2N}$, $M_{1N}=M_{2N}$) representing the nodules are attached to the upper and lower masses, respectively, and therefore move synchronously with the upper and lower masses as their integal parts. Most of laryngeal nodules seem to have a paraboloid-like shape, however, for the

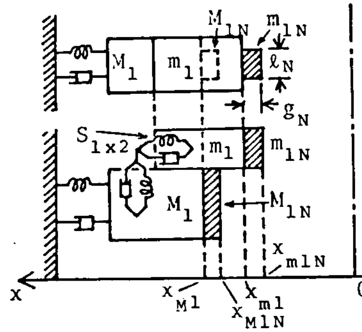

Figure 1. The structure of the eight-mass model of vocal cords with nodules.

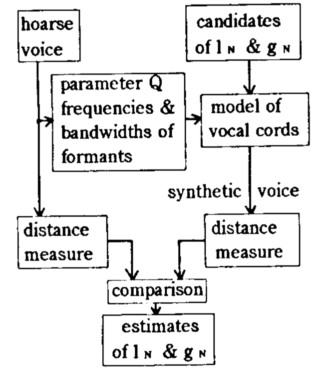

Figure 2. The method of estimating the nodule features.

simplicity of analysis it is assumed that the masses m_{iN} and M_{iN} are in the form of a rectangular parallelpiped and that those masses and their counterparts on the opposing vocal cord are placed in a proper position to collide with each other, when the model is in motion. The nodules are considered to cause a hindrance in one way or another to the motion of vocal cords known as the mucosal surface wave. This effect can be expressed as an increase in the stiffness K_{x2} of the spring S_{ix2} coupling the upper and lower masses, which is denoted as ΔK_{x2} in the following. ΔK_{x2} is thought of as a function of the width l_N of the nodules.

The Hoarse-voice Synthesizer

The hoarse-voice synthesizer consists of a glottal volume flow source and a vocal tract model. The glottal volume flow source comprises a subglottal system, a glottal impedance controlled by the eight-mass model of the vocal cords, an aspiration noise source, and a first formant load. An incomplete closure of vocal cords caused by the nodules gives rise to an aspiration noise due to turbulent glottal flow during the closed-glottis condition. To take this phenomenon into account the glottal volume flow source is provided with a noise source controlled by the vocal cord model, which supplies the glottal volume flow source with the aspiration noise when the Reynold's number for the slit formed by the opposing masses of the model exceeds a critical value. The vocal tract model that is made up of five formant filters and a simple radiation load produces the output P_0 which is a synthetic hoarse voice, when excited by the output U_G of the glottal volume flow source.

THE METHOD OF ESTIMATING THE NODULE FEATURES

The procedure for estimating vocal cord nodule features is shown in Figure 2. For an input voice sample of, say, vowel /a/, which has been uttered by a patient who has pathological vocal cords with nodules, the nodule features, i.e., dimensions of the nodules are estimated through the use of an optimization technique to be described below.

A Feature Vector for Hoarse Voice

A specific feature vector is used to describe ascoustic properties of hoarse voice. A part of this vector represents an estimated noise power spectrum of hoarse voice, which is calculated by subtracting fundamental and higher harmonic frequency components of vocal cord oscillation from power spectrum of hoarse voice. This noise power spectrum is represented as a 32-dimensional vector. This vector is augmented by a 10-dimensional glottal volume flow waveform vector to yield a 42-dimensional feature vector for hoarse voice. The glottal volume flow waveform vector has as its components real and imaginary parts of the short-time discrete Fourier transform (DFT) of hoarse voice which has been filtered using a low-pass filter with cutoff frequency of 1.5 kHz. Those real and imaginary parts of the DFT of the hoarse voice have been calculated using amplitudes and phases relative to the phase of the fundamental frequency component of the first five harmonic frequency components of the hoarse voice.

Estimation of the nodule features

It is possible to estimate the nodule features from a given hoarse voice uttered by a patient who has pathological vocal cords with nodules by maximizing a similarity in terms of the distance measure between the hoarse voice and a synthetic hoarse voice produced with the hoarse-voice synthesizer. The method of estimating the nodule features in this work is based on this analysis-by-synthesis approach. It is illustrated in Figure 2. For a stationary portion of the input voice a frequency spectral analysis is performed by means of the FFT and comb filtering to obtain the aforementioned augmented vector. This spectral analysis is necessary to produce a synthetic hoarse voice which is similar to the input voice in the sense described above. In addition to the frequency spectra several other important acoustic parameters have to be extracted from the input voice. First the input voice is subjected to the linear predictive analysis, and frequencies and bandwidths of the first five formants are estimated. Then an average fundamental period is calculated from a sequence of the fundamental period to determine a pitch control parameter Q. Using these parameters and a set of m_{iN}, M_{iN}, K_{x2}, and the width of the nodules l_N specified somehow in the hoarse-voice synthesizer will result in a synthetic hoarse voice which may or may not be similar to the input voice.

If the distance measure which will be defined in the next section is considered as a function of the mass $m_{iN}+M_{iN}=M_N$ and width l_N of nodule, then the problem of estimating the nodule features will be equivalent to that of finding a minimum of a surface representing the distance measure over the l_N-g_N plane. Here g_N represent the thickness of nodule. To find the minimum requires the evaluation of the distance measure at a number of points in the l_N-g_N plane. Finding the position of the minimum provides us with the desired nodule dimensions.

The Distance Measure

The distance measure used for the aforementioned purpose comprises the components of the feature vector defined previously, i.e., the estimated noise power spectrum and glottal volume flow waveform spectrum. It is given by

$$C=W_1 \sum_{i=1}^{32} (P_{i1}-P_{i2}) + W_2 \sum_{j=1}^{5} \{(R_{j1}-R_{j2})^2 + (I_{j1}-I_{j2})^2\}, \quad (1)$$

where P_{i1} and P_{i2} denote estimated noise power spectra of input and synthetic voices, respectively, and R_j and I_j are real and imaginary parts of the jth harmonic frequency component of the input or synthetic voice, respectively. R_j and I_j are calculated using amplitude and phase relative to that of the fundamental frequency component of the jth harmonic frequency component. The subscript 1 denotes the input voice, and the subscript 2 the synthetic voice. The weights W_1 and W_2 have been chosen as follows: $W_1=3$, $W_2=40$.

RESULTS OF EXPERIMENT

The method of estimation was applied to hoarse voices /a/ uttered by six patients who have vocal cord nodules. Since video pictures of vocal cords of those patients were available, it was possible to somewhat precisely measure dimensions of the laryngeal nodules from the pictures, taking into account a fact that the average length of the vocal cords is 14 mm for adult men and 10.5 mm for adult women. Estimates of dimensions, i.e., the width and thickness, of the nodules obtained by the method of estimation are shown in Table 1 along with measured dimensions of them.

In the model of vocal cords with nodules the incomplete glottal closure necessarily occurs, however, in actual vocal systems of patients who have vocal cord nodules it does not necessarily occur, because opposing vocal cords which are made up of a soft mucosal tissue called cover and flexible body can collide with each other even in the presence of nodules.

Some of the estimates, spesifically for voice samples 4, 5, and 6 corresponding to small nodules, are in large error because of a mismatch mentioned above between the model of vocal cords and actual vocal cords. Estimates of nodule dimensions for voice samples 1, 2, and 3

Table 1. Estimates of dimensions of the nodules obtained by the estimation scheme and corresponding measured dimensions of the nodules.

Voice Sample	Measured l_N (mm)	Estimate of l_N (mm)	Estimation Error(%)	Measured g_N (mm)	Estimate of g_N (mm)	Estimation Error(%)
1	2.96	3.15	+6.4	0.78	0.95	+21.8
2	2.11	3.15	+49.3	0.91	0.95	+4.4
3	2.96	3.15	+6.4	0.71	0.14	−80.3
4	2.87	3.15	+9.8	0.48	0.95	+97.9
5	2.16	3.15	+45.8	0.54	0.63	+16.7
6	1.63	3.15	+93.3	0.66	0.79	+19.7

corresponding to larger nodules are in agreement with actual dimensions of nodules within estimation errors less than 20 %.

Figure 3 shows waveforms of synthetic hoarse voice (sound pressure) and glottal volume flow, glottal areas, and displacements of masses for the voice sample 1. The output sound pressure and glottal volume flow waveforms are found to be rather irregular because of glottal turblent noise. The displacements of masses clearly show the occurrence of incomplete closure of the opposing upper and lower masses.

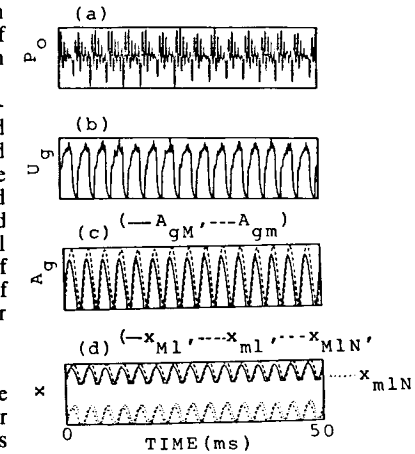

Figure 3. Waveforms derived from the estimation scheme: (a) synthetic hoarse voice (sound pressure), (b) glottal volume flow, (c) glottal areas, and (d) displacements of masses for voice sample 1.

CONCLUSIONS

In this study a new noninvasive procedure have been developed for estimating vocal cord nodule features through the use of a novel model of pathological vocal cords with a couple of nodules. This model is able to simulate the subtle movement of vocal cords with nodules in the presence of aspiration noise in the glottis. By this newly developed procedure it is possible to estimate dimensions of laryngeal nodules only through hoarse voice analysis.

This procedure has been applied to several hoarse voice samples and has been shown to be capable of estimating the state of vibration of vocal cords and dimensions of vocal cord nodules with satisfactory accuracy for large nodules. For smaller nodules, however, the estimates of dimensions of nodules were found to be in error because of the mismatch between the model and actual vocal cords.

REFERENCE

[1] Koizumi, T., Taniguchi, S., and Itakura, F. (1993), "An Analysis-by-Synthesis Approach to the Estimation of Vocal Cord Polyp Features", Laryngoscope, vol.103, pp.1035-1042.