Relevance of signal processing of the voice in diagnosing human tuberculosis

Giovanni Saggio
Dept. of Electronic Engineering
University of Rome “Tor Vergata”
Rome, Italy
saggio@uniroma2.it

Santosh Bothe
Dept. of MCA, IMED
Bharati Vidyapeeth University
Pune 411038, India

Abstract—The objective of this work was to reveal the tuberculosis condition of human patients by means of the voice recording and processing, in order to evidence deviations, if any, with respect to the voice parameters of healthy people. Comprehensive voice signal recording was obtained with low-cost apparatuses, and voice signal processing was realized using very low-computational effort parameters. A total of 312 patients were engaged, 284 tuberculosis diagnosed and 28 in healthy status as an age- and sex-matched control group.

We demonstrated that the voice can be used as a discriminating source of information between healthy subjects and tuberculosis patients, since evidence of alterations founded in some of the voice features.

Keywords—human voice; voice analysis; voice signal processing; tuberculosis

I. INTRODUCTION

Historically, sounds from the body have been clinically evaluated for diagnostic purposes. We can refer to the act of auscultation (in general using a stethoscope) for examining the heart, the breath and the gastrointestinal sounds. Aside from sounds revealed from traditional auscultation, in the latest years the sound of the voice has been considered [1], in particular for determining voice pathology [2-4] or voice disorder [5-7]. Since the voice is the result of sounds, produced by modulation of pushing out air flow, related basically (but not exclusively) to glottis, larynx and lungs, the voice signal processing has been useful to reveal diseases [8-10], disorders [11] or pathology [12] of larynx, inflammation of the lungs [13], and obstructive airway disease [14] or sleep apnea [15].

Within this frame, we considered the possibility that the voice analysis could lead to more intriguing possibilities in diagnosing [16], so that we performed the current study with the aim to determine the relevance of voice signal processing, if any, in diagnosing the laryngeal/lungs tuberculosis.

Recently, the World Health Organization referred of epidemiological burden of tuberculosis (TB) of about 7 million people every year [17], especially focused in developing countries. This really huge number has to be reduced as low as possible and, in such a view, the diagnosis plays a fundamental rule. Currently, the diagnosing tests leave millions undiagnosed and untreated [18], being the high-tech high-cost ones confined to industrialized countries and not still implemented in high-burden developing countries to any significant degree. This is why any technology or tool, useful for diagnosis purposes, which can result affordable, convenient, robust and reliable is welcome. In such a frame, the voice signal processing can potentially furnish an answer, being low-cost, easy manageable, safety and non-invasively recorded.

In general, the voice signal processing have been performed in the time domain [15, 19], in the frequency domain [6, 11], and in the quefreny domain (by means of cepstrum analysis, related to the inverse Fourier transform of the logarithmic amplitude Fourier spectrum) too [11, 20]. Different domain analysis can be convenient to evidence different aspects and parameters related to the voice sounds. In addition, a class discrimination of the speech waveform can be performed [19], and this has been done by means of classification algorithms such as Hidden Markov Models [2], Neural Networks [8], Prototype Distribution Map [8], Self-Organizing Map [8], Support Vector Machine [3, 4, 10, 12, 20], Gaussian Mixture Model [20], Linear Discriminant Analysis [4, 15], and Random Forest [12].

Regardless the so many different possibility to explore, we established to work in time and frequency domain, but limiting our attention to the most common, known and used statistical tools, in order to avoid to work with sophisticated algorithms which can be effective but can result as a step ahead towards complexity. This is to make results easily intelligible for the most part of people potentially involved in clinical treatments.

As far as we know, this is the first work addressing the relevance of the signal processing of the human voice in diagnosing tuberculosis on the basis of hundreds of evaluated subjects, with a unique prior work that considered the change in the quality of the voice of only one subject, a 28-year-old female with laryngeal and pulmonary tuberculosis [21].

II. 2. SUBJECTS, MATERIALS AND METHODS

A. Subjects

Voice data present noteworthy age-related [22], gender-related [23, 24] and ethnics-related differences. Therefore, it was fundamental to compare patients within a no meaningful difference in age, of the same gender, and members of the same ethnicity. It is reasonable to believe that also weight, height, smoking, alcohol consumption, social surrounding and culture can have some impacts on the production of voice,
nevertheless we did not consider these variables, so leading to a certain degree of uncertainty, but at this stage empirically esteemed to have a second order influence in diagnosing tuberculosis.

We had 312 male subjects who undertook this study, 28 of them were healthy and 284 of them was lung tuberculosis diagnosed. The 284 subjects diagnosed for tuberculosis claimed no other diseases differently related. The common symptoms like cough, fever, weight loss and loss of appetite were observed. The mean age of the subjects was 39.67 year, within 25-58, all subjects coming from Mumbai (India).

The voice recordings were performed at Tata Memorial Hospital (ACTREC, Sector 22, Kharghar, Navi Mumbai - 410208, India.), D.Y.Patil Hospital (Sector 5, Nerul Navi Mumbai, Maharashtra 400614), Sharma Hospital (Sector 11, Kharghar Navi Mumbai 410210 India), Navjeevan Hospital (Sector 17, Kharghar Navi Mumbai, Maharashtra 410210) and Sanjeevani Hospital (Sector 2 CBD Belapur Navi Mumbai, Maharashtra 400614). The local ethical guidelines were followed for each institution, and a written consensus was obtained for all the subjects.

Every patient was asked to phonate a predefined text in local language. In particular, we considered the Marathi and English alphabets for constructing the words of speech text. Each type of pronunciation [25-28] were related with different letter, words and phrases of spoken language. Consequently, the recruited subjects phonated: “Aai Aai Ye Ekade, Aapn Doghe Milun Jaun Tikade, Sundar Maze Gaon, Sunder Maze Ghar. Salgale Miluni Aankhi Banavu Aapn Tayala Aankhi Sunder. Swatch Asel Aapule Gaon Aani Ghar Tha Aajar Aapnun Jail dur. Tar Mag Maja Moti Karu, Mayene Mazya, Msatine Mulayam Karu. Aapli Maja Aanel Rangat, Sagele Houn Jau Aandi Aani Kamat Dang”. (The not fundamental translation is: “Mother, Mother come here, we together will go there, Beautiful home, will make beautiful village. We all together can make it more beautiful, If our village is clean, disease will automatically run away. Then we will enjoy a lot, our fun will bring joy, we all can focus on work.”).

This text was considered to have phonetics useful to the extraction of the “voice signature” for tuberculosis. For example, the words having maximum phonetic similarity with English alphabet “U” were chosen to extract the information of lungs and chest area. In addition, the text was designed to give the sensation of a poem about cleanliness and healthy, to tract the information of the recruited subjects.

B. Materials

In order to record the voice, we used a unidirectional microphone (8-28000Hz,-59dBV/μBar,-39 dBV/Pa+/-4dB, by Logitech, Lausanne, Switzerland) hold 6-8cm from the patient’s mouth (2cm of possible oscillation). The frequency range considered was limited to an audible range, in particular 200-18000Hz.

The time recording was 30sec, so that all the voice parameters could be exhaustively expressed. Every patient was asked to avoid any high-volume talk to prevent electrical saturation of the microphone.

The output signal from the microphone was fed to a pc running our home-made software termed “Audalysis”, capable to record and process the voice parameters. Audalysis is a modular, user-friendly and graphics-driven software (a block diagram reported in Figure 1) useful to avoid any special training to use it. The software allows selecting the sample rate (6000Hz, 8000Hz, 11025Hz, 22050Hz, or 44100Hz), the number of the audio channels (1 or 2) and the resolution in bits (8 or 16). For the current work we selected 44100Hz, 2 channels and 16bits. Audalysis stores the subject’s voice in a.wav-format file and can compute up to 12 voice parameters. Anyway, for further simplification purposes, we limited our attention to the 6 parameters we empirically believed as fundamentals: the peak frequency, the peak amplitude, the signal to noise distortion (SINAD), the inter modulation distortion (IMD), the signal-to-noise ratio (SNR), and total harmonic distortion (THD), all in their absolute values.

Collected data were analyzed by means of suitable statistical analysis techniques and results were stored both in excel and in notepad files.

C. Methods

Body Conditions:

Body parts play important roles in production of the voice, so that it is fundamental to avoid unnatural body postures and constrains. Particular body conditions, such as abdominal cramping, bowel disorder, musculoskeletal problems, etc. may have an impact on the voice production. In addition, the diaphragm and abdomen conditions are important in supporting the voice that is stronger when the body is in the upright position. Even the psychological conditions are important in the confidence one has in the voice, as can be seen by the fluttering of the voice when a person is nervous or anxious.

![Fig. 1. Audalysis is the home-made software used to record, store and process the signal voice parameters. Its modular architecture consists of three main blocks: the recorder/player, the analyzer and the database ones.](image-url)
possible to express all the detectable characteristics. Each patient was relaxed, seated on a comfortable chair, the elbows forming a 90° angle, the arm placed on a front table, and the microphone 6-8 cm apart from the mouth.

Statistics:

Determination of a statistically significant difference in voice variable values between the control group and the patient group was achieved using one-way analysis of variance (1-way ANOVA). If a statistically significant difference in a voice variable was detected, a threshold value was assigned as the upper limit of the 95% confidence interval (mean ± 1.96 standard deviation) of the control group value. It was then used to assess the statistical significance of the distribution of the data.

ANOVA was used as a statistical method for comparing acoustic measurements because of its useful characteristics. In particular, ANOVA was most appropriate as our sample size (312 subjects), removes random variability (so significant differences can be easily found), allows to evidence interactions between factors. In order to adopt the 1-way ANOVA we assumed that (1) the observations were dependent on the parameters (acoustic measurements) involved; (2) the sampling distribution was normally distributed; (3) the samples, and their populations, had equal variance.

III. RESULTS AND DISCUSSION

There was significant difference in the acoustic parameters between the two states (healthy and unhealthy) in all subjects. For instance, the voice peak frequency for TB conditions showed a mean value of 158.929 Hz with a 95% confidence interval in mean with upper bound of 166.869 Hz and lower bound of 150.977 Hz, rather different from that of healthy conditions, within the 472.734-427.693 Hz range. The IMD values significantly change in diseased ranging from 80.815 to 89.322 with respect to the "healthy" interval found to be 511.971 to 565.861. We revealed these as the most significant deviations in percentage for patients suffering from tuberculosis. The overall frame of the obtained results is summarized in Table 1.

### Table 1. Comparison of 6 different parameters between control healthy people and patients suffering from TB: The most relevant variations regards the peak frequency and the inter modulation distortion (IMD).

<table>
<thead>
<tr>
<th></th>
<th>Peak Freq. [Hz]</th>
<th>Peak Amp. [dB]</th>
<th>SINAD</th>
<th>IMD</th>
<th>SNR</th>
<th>THD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>427.693 to 472.714</td>
<td>65.751 to 70.041</td>
<td>1.9093</td>
<td>511.971 to 565.861</td>
<td>3.1991 to 3.5361</td>
<td>122.081 to 134.932</td>
</tr>
<tr>
<td>TB</td>
<td>50.977 to 166.869</td>
<td>52.020 to 65.917</td>
<td>1.7503 to 1.9346</td>
<td>80.815 to 89.322</td>
<td>3.1391 to 3.4694</td>
<td>124.527 to 137.635</td>
</tr>
</tbody>
</table>

Functioning of all systems and subsystem of human body are dependent on oxygen supply it receives during the breathing process and way they adapts to the changing input of the oxygen. Organ having direct role in oxygen supply and breathing pattern are lungs and heart. We are aware of the trivial fact that, breathing and heart rhythm changes if there is change in need of oxygen by various system of human body. It means lungs and heart functions as per the need of other system of human body. The detail biological explanation is beyond the scope of this paper, but we can conclude that heart and lungs can alters their functioning to suite the various conditions e.g. diseased, health, ideal etc. It means there is an association between functioning heart and lungs to normal and diseased conditions. (This does not require any proof as it is well accepted fact by Modern Medicine, Ayurveda: “Nadi Pariksha” and all branches of diagnosis & treatment.) Now, the sound production process here means converting the steady pressure of lungs into nonlinear acceleration of air flow. The capacity of lungs to exert the pressure on vocal track to produce the voice will change as per the phase and “magnitude of activity level” of various organs and systems of human body. Claims made by authors that, “changes in biochemical parameters of measuring health status can be mapped to corresponding changes in magnitude of voice parameters”. This is evident in case of lungs tuberculosis, as stated in Table 1. The reason for variation in peak frequency, peak amplitude and IMD in case of tuberculosis is probably due to a resonance of vocal tract that directly affects the peaks in the output signal.

IV. CONCLUSIONS

We evaluated the possibility that the human voice could furnish a cost-effective reliable tool in view of non-invasive safety diagnosis of human tuberculosis. Our intent was to develop a system for voice signal acquisition and processing which could be easily usable and intelligible for the most part of people potentially involved in clinical treatments.

In such a frame, we proved how the tuberculosis affects voice production to significant extents. The most significant parameters, among the six we examined, were found to be the Peak Frequency and Inter Modulation Distortion, which showed a major deviation in control group with respect to subjects suffering from tuberculosis.

The proposed technology can be made as a portable battery-operated voice sample analytics system for diagnosis purposes, which can be used in remote areas where access to robust diagnostic health care facilities is not available, and which can be adopted in high-burden developing countries.

Further investigations will be focused on revealing at what stage the tuberculosis can be usefully revealed, so to understand if the “early” stage can be revealed to help for better cure and infection control, by making subject aware about the disease.

ACKNOWLEDGMENT

The authors like to acknowledge the contribution of project students Ms. Priyanka Chauhan, Ms. Rashmi Chakraborti, Ms. Priya Garg, physicians of various hospitals who allowed us to collect the voice samples; Dr. S. B Muley for his support for statistical analysis of data; Ms. Monali Bobade for her guidelines for designing the text to extract the information; Mr. Virupaksha Bastikar and Dr. L. H. Kamble for their help in understanding the biochemistry behind voice production.
REFERENCES


