Tuberculosis Screening by Means of Speech Analysis

G. Saggio¹ and S. Bothe²

¹*Dept. of Electronic Eng., Univ. of Rome Tor Vergata, Rome, Italy*

²*Dept. of MCA, IMED, Bharati Vidyapeeth Univ., Pune, India*

E-mail: saggio@uniroma2.it; santoshbothe@gmail.com

Received 1 April 2016; Accepted 15 May 2016;

Publication 1 August 2016

Abstract

The parameters of the human voice change according to a mix of emotional, psychological and physical body conditions. However, while it is evident that this change occurs because of happiness, sadness, euphoria, depression, excitement, and so ahead, there is less evidence of a direct correlation of the change of the voice with respect to specific physical and/or pathological conditions and/or diseases. This work intends to demonstrate such a correlation in the specific case of the tuberculosis disease, evidencing differences in the voice parameters of unhealthy with respect healthy people.

Keywords: Screening, tuberculosis, voice parameters.

1 Introduction

The human voice is the communication means of information as well as moods and emotions [1, 2]. This is clearly provided by means of changes in several parameters of the voice, such as the pitch, the mel-frequency cepstral coefficients, the intensity and rate of the speech [3], and so on.

On the other end, the sounds arising from the human body have been historically evaluated for clinical diagnostic purposes. In fact, the *auscultation*

is commonly adopted by doctors as fundamental act of listening to sounds within the human body as a screening method for diagnosis.

Given the above, we argued that the voice could, in some way, embody the sounds with the information content of the physical condition of the individual, so to reveal, by means of its measurement, information of his/her healthy/unhealthy status. The underlying idea has been patented by the authors [#0001411389, 20th Oct. 2014, uibm.gov.it].

Such a patent intended to generalize the measuring of the voice already used to reveal voice pathologies [4–6] or voice disorders [7–9], also for the screening of pathologies or disorders which can be directly or indirectly related to the vocal folds, considering the voice production coming from an articulated co-ordination of different systems and sub-systems of the human body.

Moreover, different studies have been devoted to the impact of the articulation and precision of glottis function [10–11]. Other works have been carried out on speaker identification and emotion recognition using the voice signal processing [11–12]. Within this frame, the therapeutic application of voice characteristics has been also reviewed [10–13].

In addition, it has been already demonstrated changes in the voice parameters related to some neurodegenerative brain disorders, such as in the case of Parkinson's disease [14–21].

On these grounds, we intend here breaching with a research aimed screening the laryngeal/lungs tuberculosis, considering the analysis of different parameters measured from the voice of both unhealthy and healthy people as reference. In addition, at this stage our aim is also to adopt a low-cost low-effort system of analysis, so that few and well-known sounds parameters are considered, regardless more sophisticated mathematical analysis tools (such as math classifiers for instance).

2 Materials

We carried on tests by recordings the human voice of several subjects located in different Indian hospitals. The recordings were then analysed to determine the values of the voice parameters of our interest. Details on subjects/hospitals and hardware/software are in the following sub-sections.

2.1 Subjects

Our claim was to verify if the laryngeal/lungs tuberculosis produces meaningful changes in the voice of the patients, and to identify their changed/unchanged parameters of the measured voices, with respect to healthy people.

In such a view, acclaimed tuberculosis patients undertook this study, as well as healthy people as references. In particular, we recorded the voices of a total of 312 male subjects, aged 25–58 (39.67y in average), all with the same ethnicity, native of Mumbai (India). The group of tuberculosis people was made of 284 subjects (out of the 312), with common symptoms like cough, fever, loss of weight and of appetite, and without any other diseases differently related. The group of healthy people was made of the remaining 28 subjects, all declaring and appearing with no evidence of diseases.

The measurements were carried on four different Indian hospitals, in particular: 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), The local ethical guidelines were followed for each institution, and a written consensus was obtained for all the subjects who undertook the tests.

2.2 Hardware

We utilized a uni-directional microphone (by Logitech, Lausanne, Switzerland) with the following characteristics: 8–28000 Hz, –59 dBV/ μ Bar, –39 dBV/Pa+/-4 dB. The microphone was plugged into a standard personal computer running windows 7.

2.3 Software

The records of the voices and the extraction of the related parameters were obtained by means of software termed *Audalysis*, developed by the authors in Visual Basic 6. The records were realized at a sample rate of 41000 Hz and saved in .WAV format.

Audalysis is modular, user-friendly and graphics-driven software useful to avoid any special training to use it. The software allows selecting the sample rate (6000 Hz, 8000 Hz, 11025 Hz, 22050 Hz, or 44100 Hz), the number of the audio channels (1 or 2) and the resolution in bits (8 or 16). For the current work we selected 44100 Hz, 2 channels and 16bits. Audalysis allows: recording sound sources and to reproduce them; computing different acoustic parameters from the source file; storing both recorded sounds and information related to the speaker. These functions are detailed in the block diagram showed in Figure 1.

Figure 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 analyser and the database ones.

3 Methods

3.1 Voice Recording

Each subject, in turn, took part at the test. The subject under test was seated on a comfortable chair, the elbows forming a 90° angle, the arm placed on a front table, in a relaxing condition. The subject was asked to “read and speak” or “listen and speak” a specified text in local language. Specifically, we considered a text made of words related to the Marathi and English alphabets to obtain the pronunciation of specific vocals and specific consonants [23–25].

In order to establish which text, to be spoken by the subjects, was useful for our purposes, we assumed that Marathi and English words can be suitably divided into three categories depending on their phonetic signature. These three categories can be related to three different “areas” of the human body, here conventionally named *area 1*, *area 2* and *area 3*.

The *area 1* corresponds to the neck, the head and the brain, and is mostly stressed when the subject spells sounds like “ma maaa . . .”, and English words holding sounds of the consonants “m, f, j, k, n, and p”. This *area 1* presents a sort of phonetic signature as in Figure 2a.

The *area 2* corresponds to the chest, the abdomen and the back, and is mostly stressed when the subject spells sounds like “ue ueee . . .”, and English words holding sounds of the consonants “g and h” and the vowels “u, e, i and o”. This *area 2* presents a sort of phonetic signature as in Figure 2b.

The *area 3* corresponds to the upper limbs, and is mostly stressed when the subject spells sounds like “oa oaaa . . .”, and English words holding sounds of the consonants “b, c, d and l” and the vowels “o and a”. This *area 3* presents a sort of phonetic signature as in Figure 2c.

All taking into account, we establish for the recruited subjects to phonate the sentences: “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 Thr Aaajar Aapnhun Jail dur. Tar Mag Maja Mothi Karu, Mayene Mazya, Msatine Mulayam Karu. Aapli Maja Aanel Rangat, Sagele Houn Jau Aandi Aani Kamat Dang”. (The translation is not fundamental, anyway its meaning is: “Mother, Mother come

Figure 2 Examples of recordings of the voice when the subject speaks words which includes consonants and vowels which stress to the body areas including (a) head, neck and brain, (b) chest, abdomen and back, (c) upper limbs.

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 Indian subjects. In addition, the text was designed to give the sensation of a poem about cleanliness and healthy, to generate the interest of the patients. The subject under test was requested to rehearse the trial so to be able to read without fumbling.

Tests were performed maintaining the microphone 6–8 cm apart from the subject’s mouth. The time recording was 30 sec, 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.

3.2 Voice Analysis

The characteristics of the voice can be evidenced by means of its time analysis, spectral analysis and cepstral analysis, or combination of them. In addition, classification algorithms can be adopted too, such as neural networks, support vector machines, Bayes classifier, etc., or combination of them [26].

Anyway at this stage, for the sake of clarity, we preferred to simplify the mathematical analysis as much as possible, so to furnish data easily valuable for the most part of medical staff, having no particular mathematical knowledge too. That is, we considered six well-known audio parameters, such as; peak frequency (PF); peak amplitude (PA, the amplitude of the strongest spectral component in the entire span); signal to noise and distortion (SINAD, signal S plus noise N and distortion, equal to $(S+N)/N$ in our system); inter modulation distortion (IMD, the ratio of the intermodulation power to the RMS sum of the tone power); signal to noise ratio (SNR, ratio of the signal peak power level to the total noise level); total harmonic distortion (THD).

Nevertheless, in order to demonstrate that the recorded parameters are statistically meaningful for our purposes, a statistical analysis was performed using SPSS (a software package, by SPSS Inc.).

4 Results and Discussion

We found significant differences in the acoustic parameters between the two groups of healthy and unhealthy (suffering from laryngeal/lungs tuberculosis) people. The overall frame of the obtained results is summarized in Table 1.

Table 1 The ranges of values obtained for the six investigated parameters (PF, PA, SINAD, IMD, SNR and THD) for the two groups of healthy and unhealthy people

	Peak Freq. [Hz]	Peak Amp. [dB]	SINAD	IMD	SNR	THD
Healthy	427.693 to 472.714	65.751 to 70.041	1.9093 to 1.9285	511.971 to 565.861	3.1991 to 3.5361	122.081 to 134.932
TB	150.977 to 166.869	52.020 to 65.917	1.7503 to 1.9346	80.815 to 89.322	3.1391 to 3.4694	124.527 to 137.635

As it is shown, among the six selected parameters, two are particularly relevant, the peak frequency (PF) and the intermodulation distortion (IMD), because of meaningful differences on their values between healthy and unhealthy groups. In particular, the PF for TB conditions showed a mean value of 158.922 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 472.714–427.693 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, since the other four parameters did not changed so much between groups.

In order to derive the statistical meaning of our results, we performed the t-Test, considering a significance level of $p = 0.05$. According to the obtained results (Table 2), significant parameters, related to the tuberculosis disease, are PF, PA, SINAD and IMD, being their difference statistically meaningful ($p < 0.05$). Differently, SNR and THD do not show a significant statistical difference ($p > 0.05$).

The parameters we measured are strictly related to the population that undertook this study, i.e. with the characteristic to be adult, males, and from the region of Mumbai (India). Other different parameters will compete for children, females and adult males with other ethnicity. Nevertheless, the number of patients being analysed is statistically valid to furnish comparison data for the investigation of other homogeneous group of subjects (and adult males from the region of Mumbai are millions).

Table 2 The results of the t-Test performed between the two groups of healthy and unhealthy people

Peak Freq.	Peak Amp.	SINAD	IMD	SNR	THD
3.76E-88	<0.001	0.0067	<0.001	0.3507	0.3095

5 Conclusions

According to our results, voice sample analytics may be applied to as an aid to clinical screening of laryngeal/lungs tuberculosis disease, since it produces differences in the voice parameters, as a sort of *voice signature*. Within this frame, useful investigations can be performed using standard low-cost microphone and personal computer, the latter as installed in most of clinicians consulting room, and therefore be of particular value to primary care physicians, who do not have easy access to sophisticated diagnostic equipment. There must be extra advantages in devising computer based examination procedures especially for the children and babies who may be uncooperative with the standard procedures.

Computerized analysis can be ideally suited to the long term monitoring of patients either in the hospitals or in the communities, and should also be of value in less developed countries and remote communities.

The most likely area can be the generation of personalised voice signatures database, so to have a person specific signature and remote monitoring of diseases. An exciting prospect for the future would the implementation of clinically useful analysis procedures, simply using a mobile phone communication (as far as the phone bandwidth can permit) to a specialist centre in a local hospital.

In future, we aim to investigate other pathologies or disorders (not necessarily related to the vocal folds) and other subjects differing in age, sex and ethnicity, to build reference databases useful for screening purposes of new patients.

Acknowledgment

The authors like to acknowledge the contribution of project students Ms. Vidhu Sharma, Ms. Kanishta Vyas and Group 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

- [1] C. Gobl, N. Ailbhe, 'The role of voice quality in communicating emotion, mood and attitude', *Speech communication*, 40(1), pp. 189–212, 2003.

- [2] K. R. Scherer, 'Vocal communication of emotion: A review of research paradigms', *Speech communication*, 40(1), pp. 227–256, 2003.
- [3] D. Ververidis, C. Kotropoulos, 'Emotional speech recognition: Resources, features, and methods', *Speech communication*, 48(9), pp. 1162–1181, 2006.
- [4] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón et al., 'An improved method for voice pathology detection by means of a HMM-based feature space transformation', *Pattern recognition*, 43(9), pp. 3100–3112, 2010.
- [5] V. Uloza, A. Verikas, M. Bacauskiene et al., 'Categorizing normal and pathological voices: automated and perceptual categorization', *Journal of Voice*, 25(6), pp. 700–708, 2011.
- [6] M. K. Arjmandi, M. Pooyan, 'An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine', *Biomedical Signal Processing and Control*, 7(1), pp. 3–19, 2012.
- [7] J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz et al., 'An integrated tool for the diagnosis of voice disorders', *Medical engineering & physics*, 28(3), pp. 276–289, 2006.
- [8] G. Niedzielska, 'Acoustic analysis in the diagnosis of voice disorders in children', *International journal of pediatric otorhinolaryngology*, 57(3), pp. 189–193, 2001.
- [9] A. Alpan, J. Schoentgen, Y. Maryn et al., 'Assessment of disordered voice via the first rahmonic', *Speech communication*, 54(5), pp. 655–663, 2012.
- [10] Teija Waaramaa et al., 'Perception of emotional valences and activity levels from vowel segments of continuous speech', *Journal of voice*, 24(1), pp. 30–38, 2010.
- [11] Rajka Smiljanićb and Ann R. Bradlowc J, 'Temporal organization of English clear and conversational speech', *Journal of the acoustical society of America*, 124(5), pp. 3171–3182, 2008.
- [12] Scherer et al., 'Emotion inferences from vocal expression correlate across languages and cultures' *Journal of cross-cultural psychology*, 32(1), pp. 76–92, 2001.
- [13] Gloria S. Watersa et al., 'Task demands and sentence comprehension in patients with dementia of the Alzheimer's', *Journal of brain and language*, 62(3), pp. 361–397, 1998.
- [14] A. Tsanas, M.A. Little, P.E. McSharry, J. Spielman, L.O. Ramig, 'Novel speech signal processing algorithms for high-accuracy

- classification of Parkinson's disease', *IEEE Transactions on biomedical engineering*, 59(5), pp. 1264–1271, 2012.
- [15] K. J. Kappiarukudil and M. V. Ramesh, 'Real-time monitoring and detection of "Heart Attack" using wireless sensor networks', *IEEE Proceedings of the fourth international conference on Sensor technologies and applications at Venice*, pp. 632–636, 2010.
- [16] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, 'Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity', *Journal of the royal society interface*, 8(59), pp. 842–855, 2010.
- [17] A. Tsanas, M.A. Little, Patrick E. McSharry, Lorraine O. Ramig, 'Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests', *IEEE Transactions on Biomedical Engineering*, 57(4) pp. 884–893, 2009.
- [18] M.A. Little, Patrick E. McSharry, Eric J. Hunter, Jennifer Spielman, Lorraine O. Ramig, 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', *IEEE Transactions on Biomedical Engineering*, 56(4) pp. 1015–1022, 2009.
- [19] M.A. Little, 'Biomechanically informed nonlinear speech signal processing D. Phil. Thesis, Oxford University, Oxford, UK, pp. 60–101, 2011.
- [20] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, I.M. Moroz, 'Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection', *Bio Medical Engineering On Line*, pp. 6–23, 2007.
- [21] M. Little, P. McSharry, I. Moroz, S. Roberts, 'Nonlinear, biophysically-informed speech pathology detection', *IEEE conference proceeding International Conference on Acoustics, Speech and Signal Processing at Toulouse, France*. pp. 1080–1083, 2006.
- [22] A. Yadollahi, Z. M. Moussavi, 'Acoustical respiratory flow', *Engineering in Medicine and Biology Magazine, IEEE*, 26(1), pp. 56–61, 2007.
- [23] S. S. Kraman, G. A. Pressler, H. Pasterkamp et al., 'Design, construction, and evaluation of a bioacoustic transducer testing (BATT) system for respiratory sounds', *Biomedical Engineering, IEEE Transactions on*, 53(8), pp. 1711–1715, 2006.
- [24] A. Mendes, M. Alves-Pereira, N. A. Castelo Branco, 'Voice acoustic patterns of patients diagnosed with vibroacoustic disease', *Revista Portuguesa de Pneumologia (English Edition)*, 12(4), pp. 375–382, 2006.

- [25] R. Tahamiler, D. T. Edizer, S. Canakcioglu, ‘Nasal expiratory sound analysis in healthy people’, *Otolaryngology–Head and Neck Surgery*, 134(4), pp. 605–608, 2006.
- [26] F. Cavrini, L.R. Quitadamo, L. Bianchi, G. Saggio, ‘Combination of classifiers using the fuzzy integral for uncertainty identification and subject specific optimization: application to brain-computer interface’, *Proceedings of the International Joint Conference on Fuzzy Computation Theory and Applications*, pp. 14–24, 2014.

Biographies

G. Saggio (GS) received the M.Sc. and the Ph.D. degrees in Electronic Engineering from the University of Tor Vergata, Rome, Italy, in, respectively, 1991 and 1997.

GS spent working periods at the Dept. of Electronics and Electrical Engineering, Glasgow University, at the Microelectronics Research Center, Cambridge University, at the Rutherford Appleton Laboratory, Oxford University.

Since 1997 GS is Assistant Professor at the Department of Electronic Engineering, University of Rome “Tor Vergata”, where he holds chairs in electronics at the engineering faculty (Dept. of Electronics, Dept. of DICII, Master of Sound, and Master of CBRN Protection) and at the medical faculty (Departments of Neurophysiology, Cardiovascular Medicine, Orthopedics, and Audiology).

His research interests include analog electronics, sensors, biomedical engineering, wearable devices, human-machine interface, brain-computer interfaces.

GS patented 5 new inventions (PCT/IB2011/000077, ITMI20102270, PCT/IB2011/000645, RM2014Z000036, WO/2012/131554).

GS co-founded two Spin-Offs, that are Captiks Srl (in 2012) and Seeti Srl (in 2015).

CF has been or is the scientific responsible for projects founded by ASI (Italian Space Agency), by the avionic service of the Italian Defense Department (Armaereo), by the Italian Workers' Compensation Authority (INAIL).

He is the author or co-author of more than 100 papers on international Journals/Conferences, and author of 4 books about analog electronics (3 in Italian and 1 in English edited by CRC Press).

His work has been cited in several national and international newspapers/magazines/agencies/TVs (New York Times, Reuters, Repubblica, Rai,.) and the Italian Presidency of the Council Ministers.

S. Bothe received two master degrees M.Sc. and MPM from Skikkim Maipal University and University of Pune respectively. He received his Ph.D. in Alternative Medicine from IBAM, Kolkatta and carried out his Post-Doctoral research at HiTEG Research Group University of Rome, Tor Vergata, Rome, Italy in 2010–2011.

Santosh is a winner of Young Scientist Competition 2014 organised by Inno Indigo Project Funded by European Commission and he has worked as Visiting Professor for Ambient Assisted Living, at University of Rome. Since 2012 he is working as Professor in MCA at Bharati Vidyapeeth University, Pune. His research interest includes Cognitive science, Knowledge Engineering, ICT in health care and Voice Sample Based Disease Diagnosis.

Santosh has authored more than 50 research articles, 02 books in IoT and Cognitive Science. He is the co-founder and president of NGO Daksh Foundation (2008).

He has scientific collaboration with various research group in Europe and India. His work has been cited in several national and international newspapers/magazines/agencies and Inno Indigo project of European Commission.