Biomedical Signal Processing and Control 55 (2020) 101670

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Voice quality assessment and monitoring: Understanding vocal folds dynamics and its relationships to neurocognition

This large Special Issue of the journal Biomedical Signal Processing and Control collects twelve contributions that are extended version of papers presented at the 10th MAVEBA International Workshop, held in Firenze, Italy, December 15–17, 2017.

The MAVEBA Workshop: Models and Analysis of Vocal Emissions for Biomedical Applications

The quite long title of the Workshop refers both to the study of our voice from the technical point of view and to its biomedical applications: therefore, rigor, methodology, devices, along with their essential connection to the real complex world of the human being and related clinical applications.

The lucky series of the MAVEBA International Workshops was born in 1999 and is proposed every two years as a multidisciplinary meeting for researchers to share their knowledge and recent results with anyone who is interested in this multifaceted subject that involves bioengineering, medicine, psychology, linguistics, singing and related fields, with applications ranging from the infant to the elder. Though multidisciplinary, MAVEBA is indeed a specialized meeting that over the years unveiled the thousand shades of this increasingly fascinating field of research.

In the same spirit that since almost twenty years animates the MAVEBA International Workshops, the aim of the 10th meeting was that of providing a methodological support for biomedical applications in the vast and partly unexplored field of voice production. As usual, the 10th Workshop MAVEBA was organized into Session, devoted to topics that have been debated over the years, but over the years enriched by new scientific findings and technological innovations: models of vocal folds dynamics, voice quality assessment and monitoring, voice and neurocognition, voice and speech in neurology.

This Special Issue points out the most advanced research in the field of modeling and analysis of the human phonatory apparatus.

The purpose of this short Editorial is to facilitate the reader, summarizing and highlighting the peculiar characteristics of each paper. The interdisciplinary approach and international participation that has always inspired MAVEBA will clearly appear from the subjects and the authors' affiliation.

Models and analysis concern objective measures of voice quality, that is the measure, assessment and classification of voice irregularities, also thanks to developments in smartphone technology and software for the possibility of remote patient-clinician interaction.

Our voice is in fact something whose use is so obvious that we never think about it unless "something wrong" occurs. Indeed, the quality of our voice is of fundamental importance in our daily life and relationships: "hoarseness" might heavily compromise our working day: be it introducing a meeting, giving a talk, teaching, defending a lawsuit in a courtroom, singing, acting, playing wind instruments, and so on.

Specifically, in this Special Issue this subject is addressed by four papers.

The paper from *S. Pravin Kumar, Jan G. Švec,* titled *"Kinematic model for simulating mucosal wave phenomena on vocal folds"* exploits the physical phenomena related to mucosal wave propagation, presenting a kinematic model that simulates mucosal waves on human vocal folds. The vocal fold geometry is based on a parametrically adjustable M5 model. The vibration characteristics including the mucosal wave movements are visualized using a synthetic kymogram. The model can be applied for educational and research purposes.

The second paper, from *J. Lebacq and P.H. Dejonckere,* titled *"The dynamics of vocal onset"* concerns a large number of vocal onsets of different types in various conditions of modal healthy phonation. Vocal onset is the process occurring between the first detectable oscillatory glottal movement and the steady state vibration of the vocal folds. High speed imaging, photo-, electro-, flow- and ultrasonoglottography and sound analysis have been used in combinations to allow detailed qualitative insight into the phenomenon. The authors show that hard, soft and breathy onsets can be clearly identified.

Another paper, from *A. Bouvet, A. Amelot, X. Pelorson, S. Maeda and A. Van Hirtum* is titled: *"External lighting and sensing photoglottography: Characterization and MSePGG algorithm"*. External photoglottography (ePGG) relies on external glottal transillumination and sensing, being therefore suitable for non-invasive and continuous observation. Nevertheless, a formalized relationship between ePGG signal and glottal area is lacking. The paper proposes a Multi-Signal-ePGG (MSePGG) algorithm approach based on characterization of ePGG measurements under controlled conditions using mechanical glottal replicas. Application to a human subject is illustrated and discussed.

The fourth paper of this group is from *P. Aichinger, F. Pernkopf, J. Schoentgen*. The title is: *"Detection of extra pulses in synthesized glottal area waveforms of dysphonic voices"*. The study proposes and tests a procedure that automatically detects extra pulses, which

https://doi.org/10.1016/j.bspc.2019.101670
1746-8094/

may occur in voiced source signals of pathological voices in addition to cyclic pulses. To this aim, glottal area waveforms are synthesized. The model allows for extra pulse trains and additive noise variations. A mean accuracy between 82.9% and 88.3% is achieved, that decreases with increasing energy level of additive noise.

Voice quality assessment is closely related to a deep understanding of the dynamics of the vocal folds, one of the most complex and varied mechanisms of the human body, not yet fully exploited: our vocal folds are the source of the infinite range of sounds that make each of us unique and unrepeatable.

There are five papers in this Special Issue concerning this topic. The first two make a two-part series devoted to review the current state of the art of automatic voice condition analysis systems. In both papers the authors are *J.A. Gómez-García, L. Moro-Velázquez and J.I. Godino-Llorente*. The first one is titled *"On the design of automatic voice condition analysis systems. Part I: Review of concepts and an insight to the state of the art"*. The goal of this paper is to provide the scientific community with introductory concepts, a categorisation of different aspects of voice pathology and a systematic literature review that describes the methodologies and methods mostly used in voice condition analysis systems. A discussion about some variability factors that affect the performance of these systems is also presented. The title of the second paper is: *"On the design of automatic voice condition analysis systems. Part II: Review of speaker recognition techniques and study on the effects of different variability factors"*. The aim of this study is to examine several variability factors affecting the robustness of systems that automatically detect the presence of voice pathologies by means of audio recordings. Experiments are performed to test the influence of the speech task, extralinguistic aspects, such as gender, the acoustic features and the classifiers. Experiments are carried out using state-of-the-art classification methodologies often employed in speaker recognition. The suitability of these techniques in the clinical setting is discussed.

The other three papers in this set concern the evaluation of Parkinson's disease (PD) by means of speech processing and speech articulation kinematics. The authors of first one, titled *"A forced gaussians based methodology for the differential evaluation of Parkinson's Disease by means of speech processing"* are: *L. Moro-Velazquez, J.A. Gomez-Garcia, J.I. Godino-Llorente, J. Villalba, J. Rusz, S. Shattuck-Hufnagel and N. Dehak*. The aim of the paper is to give a description of speech articulation dynamics as a probability density function of the kinematic features derived from the evolution of formants in the time domain. The statistical distribution of the dynamic behaviour of articulation features can be used to estimate differences between speech features from subjects with Parkinson's dysarthria and normative subjects, improving the differentiation capability of Vowel Space Area and the Formant Centralization Ratio. The second paper is titled: *"Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics"*, by: *P. Gómez, J. Mekyska, A. Gómez, D. Palacios, V. Rodellar and A. Álvarez*. This work proposes new approaches for the automatic detection of PD from speech by using new classification schemes to compare independently the different phonetic units of patients and controls during a discourse. Results suggest that PD affects the articulatory sequence as a whole, influencing more phonetic units requiring a higher narrowing of the vocal tract. Additionally, text dependent utterances are the recommended speech task for the detection of PD. The accuracy of the systems for the automatic detection of PD using speech is mostly below 95%. Finally, the last paper titled *"A Transversal Study of Fundamental Frequency Contours in Parkinsonian voices"* is from *P. Rodríguez-Pérez, R. Fraile, M. García-Escrig, N. Sáenz-Lechón, J.M. Gutiérrez-Arriola and V. Osma-Ruiz*. Thirty patients suffering from Parkinson's disease (PD) and 32 healthy speakers were recorded while reading a text without voiceless phonemes. Results based on fundamental frequency indicate that the influence of PD on some aspects of intonation can be masked by the effects of aging, especially for male voices. Correlations between relative fundamental frequency range and PD stage reach moderate-to-high values in the case of women. For both male and female populations, significant correlations were found between parameters obtained from the modulation spectrum of fundamental frequency and the PD stage. Nevertheless, such measures are likely to be of limited value in the early diagnosis of PD due to inter-speaker variability.

During the 10th MAVEBA Workshop methods were proposed for the study of relationships between voice and **emotional or pathological states** in adults and the elderly. Indeed, not only physical diseases but also stress, depression, lack of confidence, fear, joy, happiness, psychological and psychiatric conditions and in general any emotional state influences our voice.

A paper of this Special issue is devoted to this subject: *"Analysis of speech features and personality traits"*, by *A. Guidi, C. Gentili, E.P. Scilingo and N. Vanello*. In this study, a correlation analysis between some speech-related features and the personality traits is performed. An experimental protocol, consisting of two structured speech tasks, was administered to eighteen healthy subjects. Significant correlations among personality traits and speech features were observed. The adopted speech task was found to influence the results. The impact of personality traits and speech production on the characterization of mental disorders and emotional/mood state of the speaker are discussed.

Another subject exploited in the MAVEBA workshop was the **relationship between speech and neurocognition**. Specifically a paper titled: *"Transfer learning in imagined speech EEG-based BCIs*" was presented by *J.S. García-Salinas, L. Villaseñor-Pineda, C.A. Reyes-García and A. Torres-García*. Brain–Computer Interfaces (BCI) based on electroencephalograms (EEG) are systems aiming at providing a communication channel to any person with a computer. These devices allow sending messages or controlling devices using the brain signals. Imagined speech is one of the most recent neuroparadigms. The proposed method extracts characteristic units of the EEGs associated with the words of an initial vocabulary. Subsequently, a new imagined word is represented with these code words, and then a classification algorithm is applied. The method was tested on a 27 subjects' dataset. Accuracy ranges between 65.65% and 68.9%.

Another paper deals with **new-born cry and language development** at birth. The title is: *"Automated analysis of newborn cry: relationships between melodic shapes and native language"*, by *C. Manfredi, R. Viellevoye, S. Orlandi, A. Torres-García, G. Pieraccini and C.A. Reyes-García*. Since the last trimester of pregnancy human fetuses exhibit a specific sensitivity to prosodic features such as melody, intensity, and rhythm that are essential for an infant to learn and develop the native language. First results are presented concerning the automated mother language identification of a set of about 7,500 cry units coming from French, Arabic and Italian mother-tongue healthy full term newborns. Acoustical parameters and 12 different melodic shapes are computed with the BioVoice software tool and their classification is performed with Random Forest and 4 neuro-fuzzy classifiers. Results show up to 95% differences among the three languages.

This Special Issue highlights that, since the first MAVEBA, the progress of joint and interdisciplinary research has deepened and widened the fields of investigation. However, nor the curiosity nor the enthusiasm of the old and new participants, was reduced.

As evidenced by the list of contribution, the scientific community that met in Firenze comes from all over the world, confirming that the study of the human voice, our main means of communication, has no geographical boundaries.

As for past editions, the Workshop was cheered by the generous hospitality and participation of artists and experts of the voice:

the Luigi Cherubini Conservatory, actor Ugo Donato, flute player Gianluigi Nuccini and the Accademia Bartolomeo Cristofori, with its collection of instruments of historical interest.

I wish to thank participants, chairpersons, round tables moderators, young collaborators and the website administrator with whom I share this experience since several years for their invaluable and qualified help.

To conclude this brief presentation, I wish everyone a good reading of this Special Issue and I look forward to meeting anyone interested at the next MAVEBA that will take place December 17–19, 2019 still in Firenze. We will celebrate twenty years of advances in the study of the human voice!

Details at http://maveba.dinfo.unifi.it/.

Claudia Manfredi
Department of Information Engineering, Universita' degli Studi di Firenze, Via di Santa Marta 3, 50139 Firenze, Italy
E-mail address: claudia.manfredi@unifi.it

Available online 4 October 2019