

BIOVOICE: A MULTIPURPOSE TOOL FOR VOICE ANALYSIS

Maria Sole Morelli¹, Silvia Orlandi² Claudia Manfredi¹

¹ Department of Information Engineering (DINFO), Università degli Studi di Firenze, Firenze, Italy

² Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada

msole.morelli@gmail.com, sorlandi@hollandbloorview.ca, claudia.manfredi@unifi.it

Abstract: BioVoice is a user-friendly software for the acoustical analysis of the human voice. Here we introduce the last version of this software tool.

Results of the acoustical analysis of newborn cry signals are described and discussed. Furthermore, results of the acoustical analysis of 2 adults (1 male and 1 female) and 1 child emitting sustained vowels are shown.

Keywords: BioVoice, acoustic analysis, software tool, voice parameters, infant cry, fundamental frequency, formant frequencies.

I. INTRODUCTION

Speech and vocal productions are characterized by acoustic waves generated by physiological processes that involve the central nervous system, the respiratory system and the phonatory apparatus.

The acoustical analysis of the human voice is considered clinically relevant in assessing the health state of the phonatory apparatus and detecting possible neurological disorders [1, 2]. Main clinical applications concern screening, diagnostic support and evaluation of the effectiveness of treatments [3]. To date, few automated methods have been developed for voice analysis, the most used being PRAAT [4]. Another one is the LENA system, that investigates both healthy child recordings and child populations with language disorders [5].

BioVoice, a user-friendly software tool for voice analysis, was developed at the Biomedical Engineering Lab, Firenze University [6–9]. It allows recording the human voice from the newborn to the elder and performing both time and frequency analysis, estimating more than 20 acoustical parameters. Here, a short description of the software is proposed. Furthermore, some results on voice analysis of newborn cries, child sustained vowels and adult sustained vowels are shown.

II. METHODS

A. Biovoice: software description

In Fig. 1, the main interface of BioVoice is shown. The user has to follow few mandatory steps to perform voice analysis.

First, the user selects and uploads at least one audio file. Indeed, it is possible to upload at the same time several files of any time duration and concerning different age range, gender and kind of voice emission. The file selection is allowed from any folder on the computer or HD or USB key. Only wav. files can be analysed.

Before starting the analysis, the user has to specify the settings of the audio file(s). Specifically, age (newborn, child, adult), gender (male or female), and kind of vocal emission (voiced, singing, speech, cry) must be selected (Fig.2).

When the analysis starts, BioVoice first performs the selection of voiced/unvoiced (V/UV) audio segments [10]. Then, all the parameters of interests are extracted from each voiced segment. Specifically, in time domain, information about the number and length of voiced segments, length of silence segments and percentage of voiced segments are extracted and saved in an excel table. A picture shows the V/UV selection. In the frequency domain, fundamental frequency (F0), formant frequencies (F1, F2, F3), noise level (Normalized Noise Energy) and jitter are estimated. For F0 and for each formant, the mean, median, standard deviation, maximum and minimum values are calculated and saved in excel tables. Furthermore, differently from other automatic software tools, BioVoice computes the melodic shape of F0, identifying up to 12 melodic shapes (rising, falling, symmetric, plateau, low-up, up-low, double, frequency step, complex, unstructured, not a cry, other) [11]. It is also possible to perform a perceptual melodic analysis, looking at each melodic shape of F0 and classifying them manually. Some other options are available. Indeed, a selected audio file can be listened to, and new audio files can be saved using a connected external microphone. For newborn and child, it is also possible to perform the perceptual analysis of the melodic shape and compare it with the automatic results.

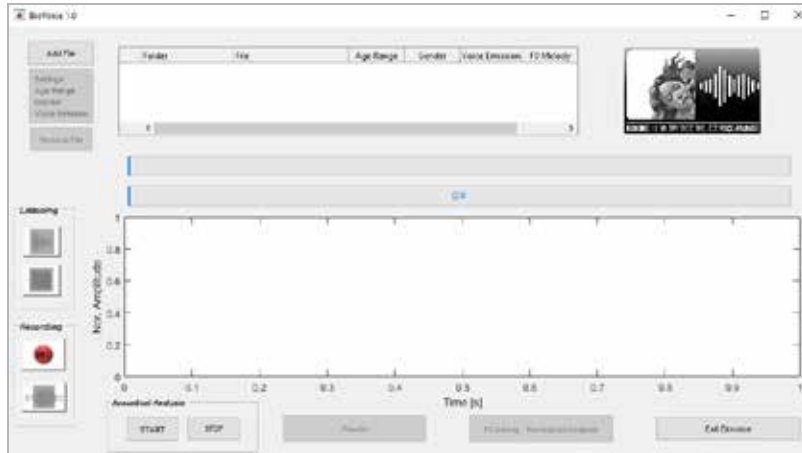


Fig.1 BioVoice user interface

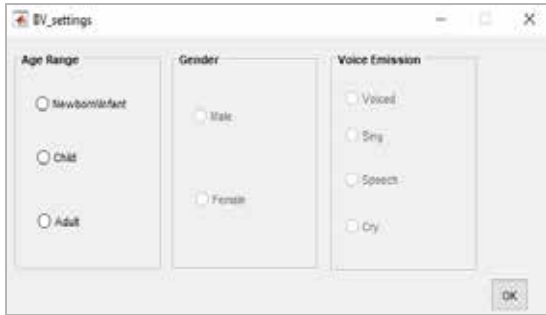


Fig.2. BioVoice settings

At the end of the analysis, BioVoice results and pictures are saved in a specific folder created in the same directory of the audio file. Tables (excel) contain F0 and formants values and statistical information about the parameters. Colour figures (.jpeg) include: V/UV selection, F0 shape and spectrogram with formants values superimposed, for each detected voiced frame.

B. Data analysis

To show some applications of BioVoice, we report the results of the analysis of four different recordings pertaining to different categories of human voice: a newborn cry (1st week of life) and sustained vowels of: a typically developing child (4 years old), an healthy adult male and an adult female (both 24 years old). Specifically, for the child and the adults only \a\, \i\ and \u\ vowels are considered here, as they allow building the vowel triangle, i.e. the plot of F2 vs F1 [12].

III. RESULTS

In Fig.3, an example of the results of newborn cry analysis is shown. Specifically, in this recording six voiced frames (cry units) were detected with BioVoice. For each cry unit, F0, along with its mean and std and

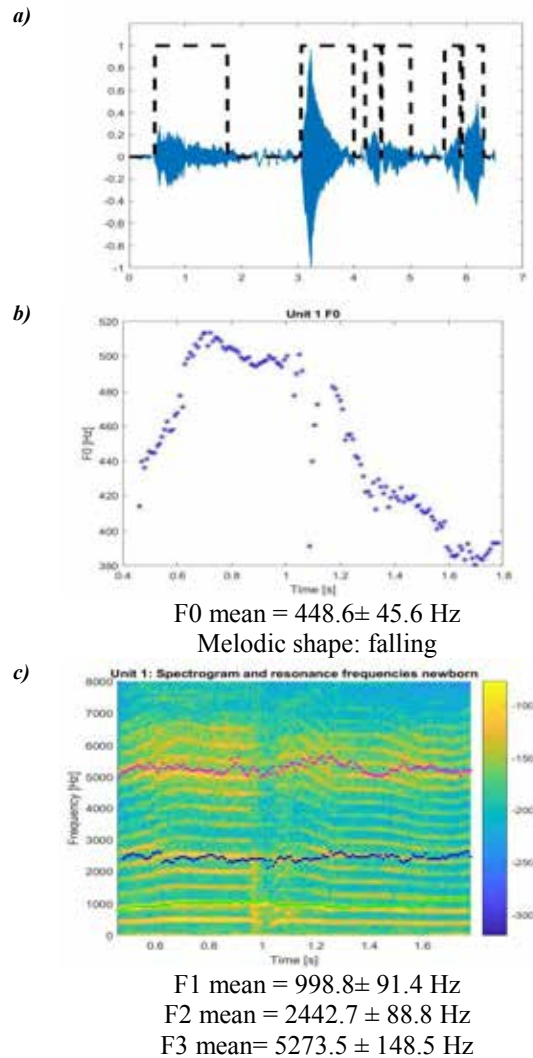


Fig.3. Newborn cry. a) V/UV selection; b) F0 shape; c) Spectrogram and formants. First cry unit.

its melodic shape are computed. Also, the spectrogram

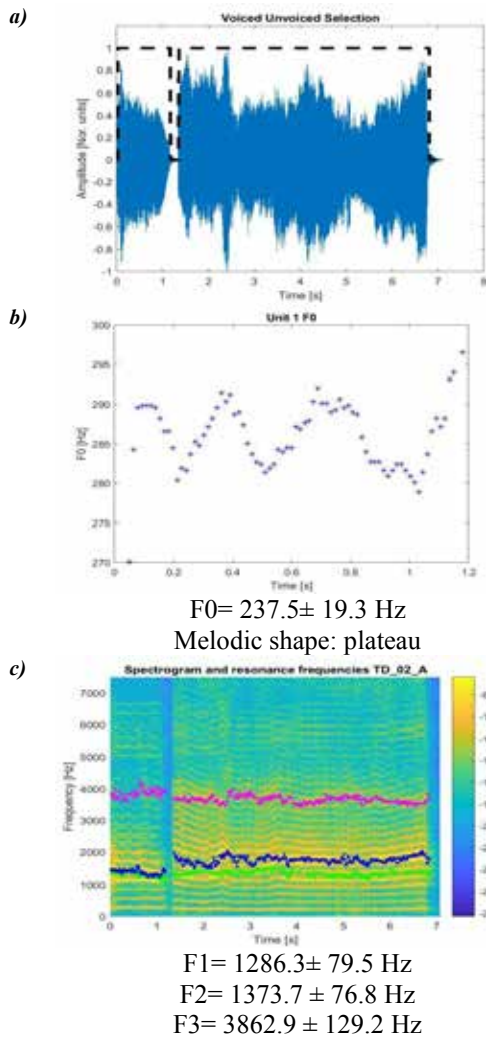


Fig.4. Child voice (sustained 'a')

a) V/UV selection; b) F0 shape; c) Spectrogram and formants for the first voiced unit.

with the first three formants superimposed is reported. In Fig.4, an example of sustained vowel for a child voice is reported ('a'), as well as its the melodic shape and spectrogram. Finally in Fig.5, an example concerning a male adult voice is shown ('a'). In Fig. 6, the vowel triangle of a child voice, of a male and a female adult voices are shown and compared to the normative values of male adult voice reported in [12].

IV. DISCUSSION

We presented the last version of BioVoice, a software developed at the Firenze University for the acoustical analysis of human voice. This updated version is improved as of the interface layout is concerned, but most of all concerning the methodologies for parameter estimation. In addition to F0 and formants values, it allows performing melodic

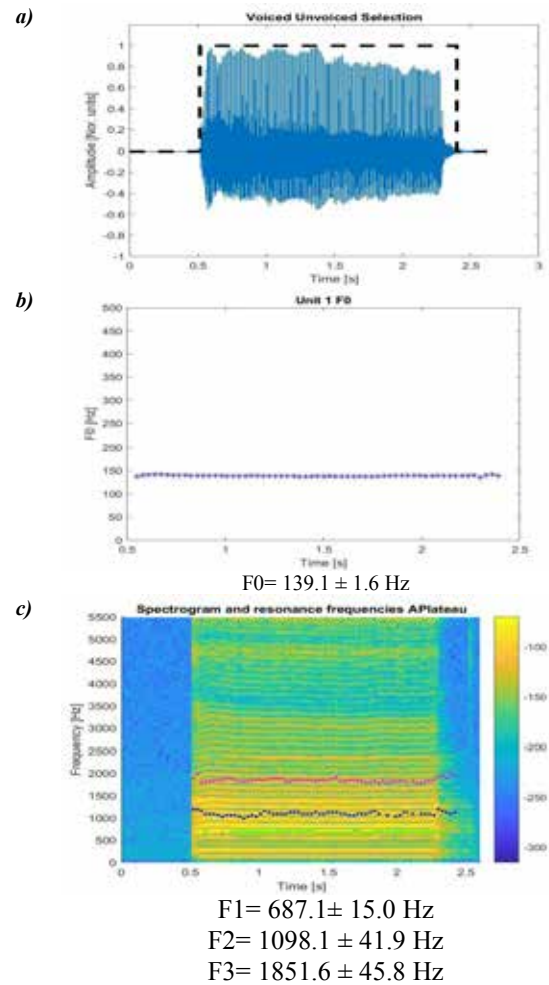


Fig.5. Adult male voice (sustained 'a')

a) V/UV selection; b) F0 shape; c) Spectrogram and formants.

shape analysis in newborn cry and child's voice, both automatic and perceptual.

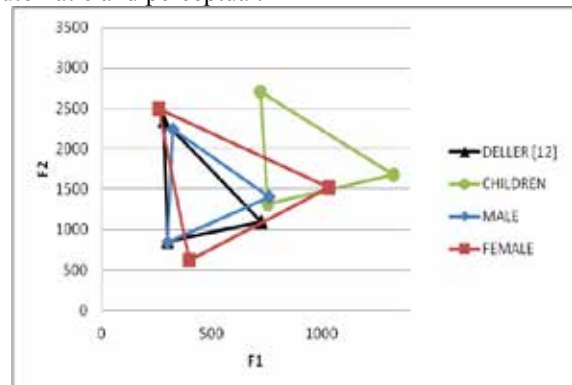


Fig. 6. Vowel triangle: Comparison of children and adults results to normative values in [12].

The reported results show some potential of this software. As for newborn cry analysis, BioVoice provides coherent values of F0 and formants in

different cry units. Concerning the melodic analysis, the software automatically detects the shape of F0, that can be compared to the perceptual one. As for child analysis, the comparison of its vowel triangle with normative values of adult male shows some differences. Indeed, as expected, children voice has higher values of F1 and F2 with respect to adults, the values of formants being related to the shape and size of the vocal tract: the larger the vocal tract cavities, the lower the resonance frequencies [12]. Concerning the adult's voice, the same consideration can be made. The vowel triangle of male voice is similar to the normative triangle reported in [12], while, as expected, for the female voice higher frequency values are shown, due to the smaller vocal tract cavities.

In the current version of BioVoice child and adult voice analysis is limited to voiced frames only. We are working towards including running speech analysis for both groups. Moreover, a revised version of the singing voice analysis is under development. In this case, two more formants, F4 and F5 must be estimated, as well as the so-called singer's formant and other acoustical parameters [13]. Finally, up to now the melodic shape is computed only for newborn cry and children voice, but it will be extended to the adult's voice too, as it might be related to the emotional state of the subject.

V. CONCLUSION

A new version of BioVoice is presented. This software could be very useful for estimating several acoustical parameters, from the newborn to the elder. The software is very intuitive and easy to use also by a less expert user. The executable version is freely available upon request.

REFERENCES

- [1] Niedzielska, G. (2001). Acoustic analysis in the diagnosis of voice disorders in children, *International Journal of Pediatric Otorhinolaryngology*, Vol. 57, No. 3, 189–193. doi:10.1016/S0165-5876(00)00411-0
- [2] Niebudek-Bogusz, E.; Fiszler, M.; Kotylo, P.; Sliwinska-Kowalska, M. (2006). Diagnostic value of voice acoustic analysis in assessment of occupational voice pathologies in teachers, *Logopedics Phoniatrics Vocology*, Vol. 31, No. 3, 100–106. doi:10.1080/14015430500295756
- [3] Laver, J.; Hiller, S.; Beck, J. M. (1992). Acoustic waveform perturbations and voice disorders, *Journal of Voice*, Vol. 6, No. 2, 115–126. doi:10.1016/S0892-1997(05)80125-0
- [4] Boersma, P. (2001). Praat, a system for doing phonetics by computer., *Glott International*, Vol. 5, No. 9/10
- [5] Ganek, H.; Eriks-Brophy, A. (2018). Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review, *Journal of Communication Disorders*, Vol. 72, 77–85. doi:10.1016/J.JCOMDIS.2017.12.005
- [6] Orlandi, S.; Bocchi, L.; Donzelli, G.; Manfredi, C. (2012). Central blood oxygen saturation vs crying in preterm newborns, *Biomedical Signal Processing and Control*, Vol. 7, No. 1, 88–92. doi:10.1016/J.BSPC.2011.07.003
- [7] Manfredi, C.; Bocchi, L.; Cantarella, G. (2009). A multipurpose user-friendly tool for voice analysis: Application to pathological adult voices, *Biomedical Signal Processing and Control*, Vol. 4, No. 3, 212–220. doi:10.1016/J.BSPC.2008.11.006
- [8] Rruqja, N.; Dejonckere, P. H.; Cantarella, G.; Schoentgen, J.; Orlandi, S.; Barbagallo, S. D.; Manfredi, C. (2014). Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia, *Biomedical Signal Processing and Control*, Vol. 13, 71–78. doi:10.1016/J.BSPC.2014.03.011
- [9] Manfredi, C.; Bocchi, L.; Orlandi, S.; Calisti, M.; Spaccaterra, L.; Donzelli, G. P. (2008). Non-invasive distress evaluation in preterm newborn infants., *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, Vol. 2008, 2908–2911. doi:10.1109/IEMBS.2008.4649811
- [10] Orlandi, S.; Dejonckere, P. H.; Schoentgen, J.; Lebacqz, J.; Rruqja, N.; Manfredi, C. (2013). Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring, *Biomedical Signal Processing and Control*, Vol. 8, No. 6, 799–810
- [11] Manfredi, C.; Pieraccini, G.; Orlandi, S.; Viellevoe, R.; Torres-García, A.; Reyes-García, C. (2019). Automated analysis of newborn cry: relationships between melodic shapes and native language, *Biomedical Signal Processing and Control*, Vol. (In print)
- [12] Deller, J. R.; Hansen, J. H. L.; Proakis, J. G. (2000). *Discrete-Time Processing of Speech Signals*
- [13] Manfredi, C.; Barbagallo, D.; Baracca, G.; Orlandi, S.; Bandini, A.; Dejonckere, P.H. (2015). Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers, *Journal of Voice*, Vol. 29, No. 4, 517.e1-517.e9