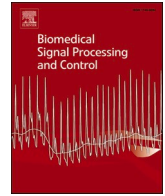




Contents lists available at ScienceDirect

# Biomedical Signal Processing and Control

journal homepage: [www.elsevier.com/locate/bspc](http://www.elsevier.com/locate/bspc)

## BioVoice: A multipurpose tool for voice analysis

Maria Sole Morelli<sup>a,b,\*</sup>, Silvia Orlandi<sup>c,1</sup>, Claudia Manfredi<sup>d</sup>

<sup>a</sup> Sant'Anna School of Advanced Study, Pisa, Italy

<sup>b</sup> Fondazione Toscana Gabriele Monasterio, Pisa, Italy

<sup>c</sup> Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada

<sup>d</sup> Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy

### ARTICLE INFO

#### Keywords:

Acoustical analysis  
Software tool  
Adult voice  
Child voice  
Newborn cry  
Singing voice

### ABSTRACT

**Objective:** This paper introduces BioVoice, a user-friendly software tool for the acoustical analysis of the human voice. It estimates more than 20 acoustical parameters with advanced and robust analysis techniques specifically developed for different vocal emissions, from the newborn to the adult and the singer.

**Methods:** BioVoice performs both time and frequency analyses, detecting the number, length, and percentage of voiced and unvoiced segments and computing fundamental frequency (F0), formant frequencies (F1-F3), noise level, and jitter. The software tool computes the melodic shape of F0 out of 12 basic shapes and allows performing perceptual analysis in newborn and child voice. In the singing voice case, formants up to F5 are computed as well as the quality ratio and parameters concerning vibrato and its regularity. Colour figures and Excel tables show F0, the spectrogram with formants, voiced segments, and quality ratio.

**Results:** Examples of voice analysis in adults, children, newborns, and singers are presented. They show the specific capabilities and the high performance of BioVoice also as compared to another existing software tool.

**Significance:** BioVoice is a free user-friendly software tool for voice analysis that implements new estimation techniques. Basic parameters are computed as well as new ones specifically developed for newborn cry and singing voice analysis, not available with current software tools.

**Conclusions:** BioVoice is capable to deal with low to high pitched voices implementing dedicated tools. Thanks to its simple and intuitive interface, colour figures and Excel tables, it is a valuable tool suitable also for the inexperienced user.

### 1. Introduction

The analysis of the human voice has been and is still challenging, due to its high variability across gender, age, pathological or emotional state of human beings. Applications mainly concern clinical aspects such as hoarseness and in general the quality of the voice as related to pathologies of the vocal folds. Also, and somewhat more recently, there is a growing interest in the relationship between voice and neurological conditions. This correlation is indeed relevant, as voice production is strictly dependent on the correct functioning not only of the vocal folds but also of several areas of the central nervous system.

Thanks to the sophisticated human auditory apparatus, perceptual analysis is the first and basic approach for the assessment of voice quality, the most used scale being the GIRBAS scale [1]. However, the need for an objective evaluation of voice quality soon proved

fundamental, in order to minimize the errors due to perceptual and, consequently, individual analysis.

Therefore, along with the increased reliability and processing speed of computers, in the last fifty years several voice analysis tools and software were developed and some of them are available on the market. Perhaps the first and most used one was the Multi Dimensional Voice Program (MDVP) [2], developed and commercialized in the nineties by the Computerized Speech Lab (Kay Elemetrics Corporation, Lincoln Park, NJ, USA), later Pentax Medical, that requires the use of Computerized Speech Lab (CSL). Though widely used basically for the analysis of dysphonia, MDVP has some limitations concerning the license cost and, mainly in its earliest version, its poor usability by the non-expert user as well as analysis techniques often not adequate for the complexity of the signal and the wide range of applications. Other widely used tools are DrSpeech [3] and Wevosys [4] useful in addition to

\* Corresponding author at: Sant'Anna School of Advanced Study, Pisa, Italy.

E-mail address: [msmorelli@monasterio.it](mailto:msmorelli@monasterio.it) (M.S. Morelli).

<sup>1</sup> Authors contributed equally.

<https://doi.org/10.1016/j.bspc.2020.102302>

Received 26 May 2020; Received in revised form 17 October 2020; Accepted 24 October 2020

Available online 21 November 2020

1746-8094/© 2020 Elsevier Ltd. All rights reserved.

traditional speech therapy methods and for voice range profile.

Several open-source tools were also developed, such as VoceVista [5], Wavesurfer [6], Speech Analyzer [7], WASP [8], and VAT [9], this one developed in the MATLAB programming environment. They perform fundamental frequency, spectrographic and spectral analysis, some being particularly suited as a feedback tool to practice singing and to document vocal development over the course of a voice education or therapy. The most widely used by top-level researchers is Praat [10] though its usage, being not very intuitive, may be challenging for those with little computing expertise.

Those mentioned here are just a few out of the many existing tools, both for purchase and free, quite often devoted to a specific application, some of them being constantly updated and some others not.

BioVoice was developed as a software tool for voice analysis in its various facets and is based on new algorithms that make it particularly suited also for high-pitched and quasi-stationary voices, such as the newborn cry, the children vocalizations, and the singing voice. It is the result of several years of research and development carried on with the MATLAB® toolbox (MathWorks Inc, Natick, MA). This has prevented its spread, hitherto requiring the installation of MATLAB whose license of use is quite expensive and thus prohibitive for some interested users. Only recently an executable version that does not require a MATLAB license was set up, and it is this version that is described here along with a summary of its main features.

BioVoice allows recording the human voice, uploading several .wav files at the same time of any time duration and concerning different age range, gender, and kind of voice emission. Each recording can be listened to and its time wave plotted through the BioVoice interface.

BioVoice performs both time and frequency analysis. In the time domain, the number, length and percentage of voiced and unvoiced segments (V/UV) are detected and saved in an Excel table. In the frequency domain, fundamental frequency (F0), formant frequencies (F1-F3), noise level (Normalized Noise Energy - NNE), and jitter are estimated. For F0 and for each formant, the mean, median, standard deviation, maximum, and minimum values are calculated.

Furthermore, differently from other automatic software tools, BioVoice implements specific tools both for the newborn cry and the singing voice. Specifically, for newborn cry and child voice, it computes the melodic shape of F0, automatically identifying up to 12 melodic shapes. It also allows performing the perceptual melodic analysis: the user looks at the shape of each F0 voiced frame and classifies it manually. Objective and perceptual results are then compared and the percentage of the match is displayed. In the singing voice case F0 is displayed in cents (100 cents = 1 semitone =  $1/12$  octave). Two more formants are computed (F4 and F5) as well as the quality ratio (QR) [11] defined as the ratio of the Power Spectral Density (PSD) corresponding to the cluster of the first two formants and that corresponding to the third, fourth, and fifth formants. Parameters concerning vibrato rate and extent and its regularity (jitter and shimmer) are also provided.

BioVoice results and pictures are saved in a dedicated folder automatically created in the same directory as the one containing the analyzed audio file. Data and parameters, as well as statistical information, are saved in Excel tables. For each voiced frame detected with the implemented V/UV selection procedure, color figures (.jpg) are also created showing the F0 time evolution and the spectrogram with formants values superimposed. More pictures concern the singing voice analysis. Specifically, F0 in cents and the PSD (normalized with respect to its maximum value) along with the QR are shown.

This work extends the one presented at the 11th MAVEBA Workshop [12]. Here, BioVoice is described in detail, starting from the installation procedure and going through its settings and usage, highlighting its peculiar features, its versatility and ease of use. The technical details of its numerous functions, specifically created for the calculation of parameters in different cases, were already widely described in previous papers [13–19]. The performance of BioVoice has also been successfully tested on synthetic signals of newborn cry [20,21] and adults' voice

[19]. In this paper, additional applications of BioVoice on real cases are shown, to illustrate its most peculiar features that make it unique.

## 2. Materials and methods

The following sections describe for the first time a detailed tutorial of BioVoice, from the installation procedure to the analysis provided by the software tool. In the first part, the requirements that a computer must meet to install BioVoice are listed. Then a complete description of the software is presented and the information for its correct usage is detailed. In the last section, examples of analysis, concerning newborns, adults, and singers, are described, highlighting BioVoice's capabilities.

### 2.1. Getting started

BioVoice is a software platform designed in MATLAB® 9.3 (R2017b). In order to use BioVoice, the computer must be equipped with at least an Intel® Core™ i3 processor running a 64-bit version of Windows OS (from Windows 7 on). It has not been tested on Mac OS or Linux. An executable beta version is available. Users without MATLAB, must install MATLAB®Runtime R2017b, which can be freely downloaded from: <https://www.mathworks.com/products/compiler/matlab-runtime.html>. BioVoice is delivered in a .zip file named BioVoice-master.zip. It is freely downloadable following the link <https://github.com/ClaudiaManfredi/BioVoice>. The website also contains a Readme.txt file that describes the main features of BioVoice and a BioVoice\_Installing\_Guide.pdf file explaining how to install it. The download may take several minutes, depending on the computer in use. Quite a long installation time is indeed one of the drawbacks of BioVoice.

### 2.2. Execute BioVoice

BioVoice can be launched from the folder where it was installed or through a link on the desktop. A picture showing the logo is displayed (Fig. 1). After 1'-2' the main BioVoice interface appears along with a warning message (Fig. 2). This is just a reminder for the user. Indeed, if Excel files or figures created by a previous analysis session are not closed, BioVoice could stop running. The BioVoice user interface is shown in Fig. 3. The BioVoice interface includes only immediate functionalities that can be called through a few buttons located on the left and lower side of the window. No technical specification is asked to the



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

Fig. 1. BioVoice Logo.

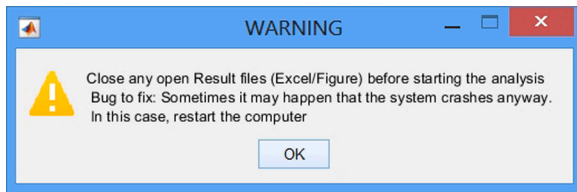


Fig. 2. Warning: close figures and Excel files from previous analysis.

user. Buttons allow four possible choices:

1. Recording a new voiced file
2. Uploading one or more voiced files
3. Listening to a file among the uploaded ones
4. Analyzing the uploaded files

To avoid possible discrepancies due to differences in audio compression, BioVoice analyzes only .wav files that are automatically resampled at 44.1 kHz.

### 2.2.1. Recording a new voiced file

To record a new voiced file a microphone must be connected to the computer. By selecting the red button labelled as "REC" in the Recording box (bottom left side of the interface in Fig. 3), recording starts and the button "STOP REC" just below "REC" becomes active. To perform a recording any microphone is allowed, but the choice could affect the results. Obviously, the recording must be made carefully, sufficiently close to the microphone, by avoiding any background noise that could overlap and corrupt the recorded signal. To stop recording, the "STOP REC" button must be selected. A window opens through which the user can select the folder in which the recorded file must be saved.

### 2.2.2. Upload one or more voiced files

By clicking the "Add File" button on the top left side of the main BioVoice interface, a window opens where the folder of interest can be searched for. Files can be selected from different folders. Holding down the "ctrl" button several files can be selected from the same folder. Clicking on the "Open" button makes the uploaded files visible in the white box in the upper part of the interface, arranged in the same order

as they are uploaded. An example is shown in Fig. 4. If the user forgets to upload one or more files, they can be added by repeating the same procedure. Instead, if one or more files were unintentionally uploaded, they can be removed simply by selecting them in the list and clicking on the "Remove File" button on the left. If the user tries to upload an already uploaded file, a message is shown (Fig. 5) and the user is asked to select a different file. Notice that, to avoid mistakes, the "Remove File" button is not active during the upload step. By selecting one of the uploaded files, the signal waveform appears in the large white box in the middle of the interface. It displays the normalized amplitude of the signal as a function of time. This gives a first view about the signal that will be analyzed, providing visual information such as its time duration, its time evolution, and amplitude peaks. An example is shown in Fig. 6.

### 2.2.3. Listening to a file among the uploaded ones

Listening can be done for a single file at a time. By selecting the file in the list and clicking on the "PLAY" button in the Listening box (mid left side of the interface in Fig. 3), the voice signal is played through the speakers of the computer or through headphones or loudspeakers connected to it. To stop listening before the end of the audio file, click the "STOP" button in the same box. When PLAY is clicked again, the file is reproduced from the beginning.

### 2.2.4. Analysis of the uploaded files

The analysis is performed for all uploaded files. In order to obtain correct results, BioVoice implements different analysis procedures according to the gender and age of the subject. Therefore it requires some simple preliminary specifications regarding voice files. This avoids the user having to specify any technical detail or setting that could be difficult to choose by a non-expert user. These specifications must be entered before starting the analysis, by selecting a file and then clicking on the "Settings: Age Range, Gender, Voice Emission" button (upper left side of the interface). As shown in Fig. 7, this window is divided into three boxes, each one concerning a specification of the subject that emitted the vocal sound:

- Age Range
- Gender
- Voice emission

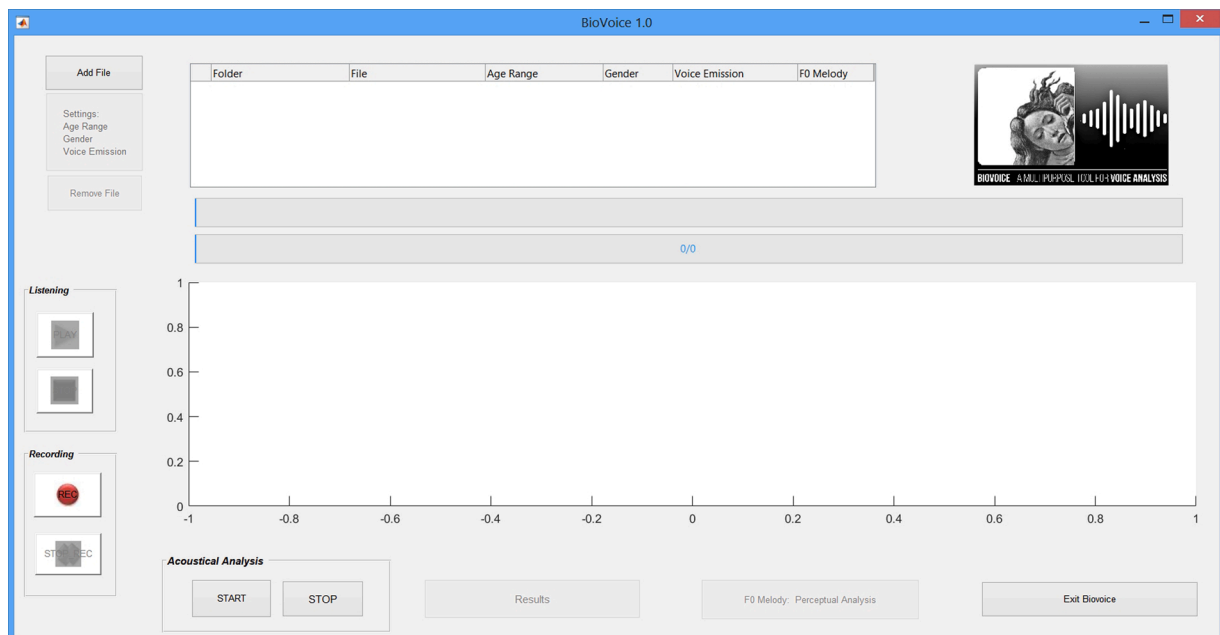


Fig. 3. BioVoice user-friendly interface. Buttons on the left and lower side allow several choices.

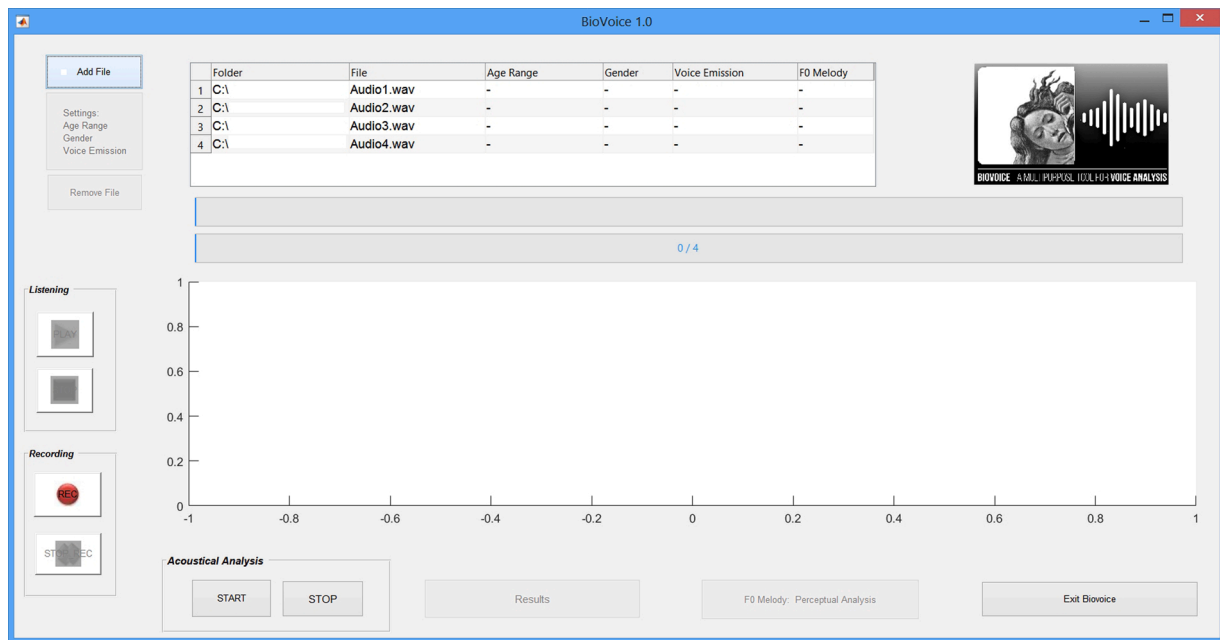


Fig. 4. Uploaded files are displayed in the BioVoice interface (top panel).

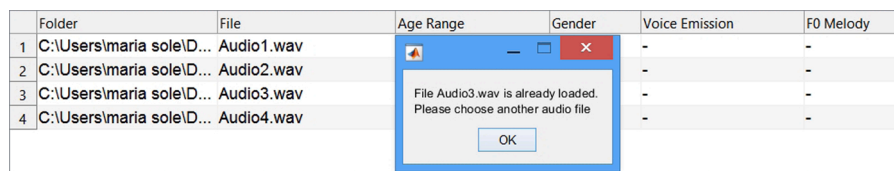


Fig. 5. BioVoice message to avoid uploading duplicates.

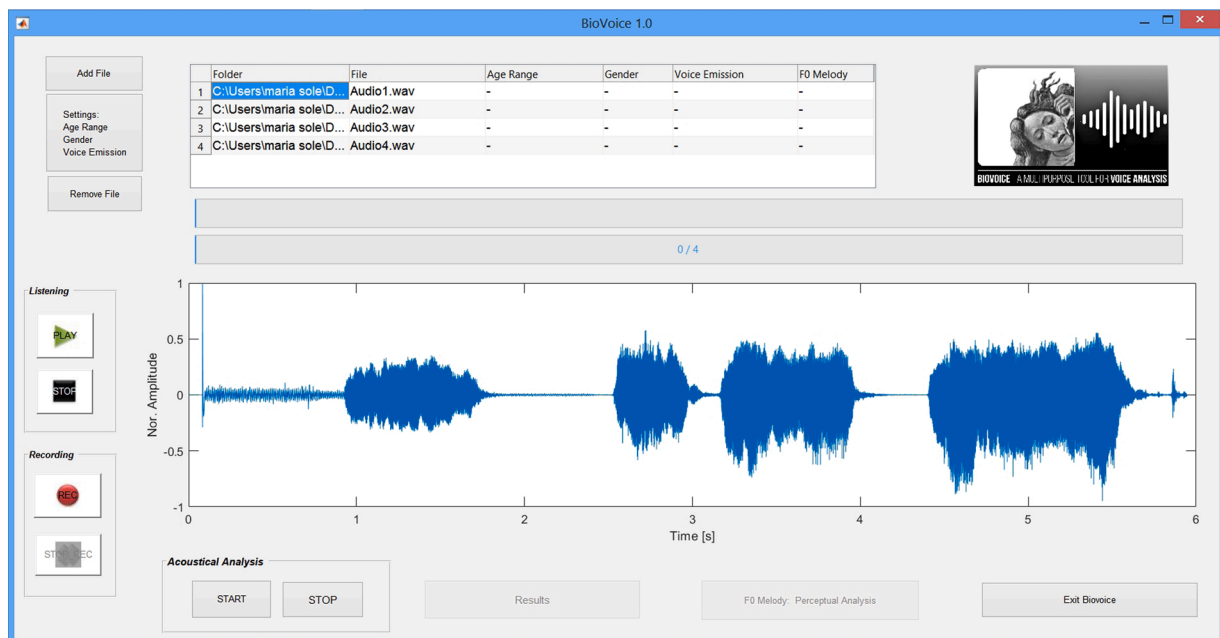


Fig. 6. The normalized waveform of the selected Audio1.wav file is displayed in the central panel of BioVoice interface.

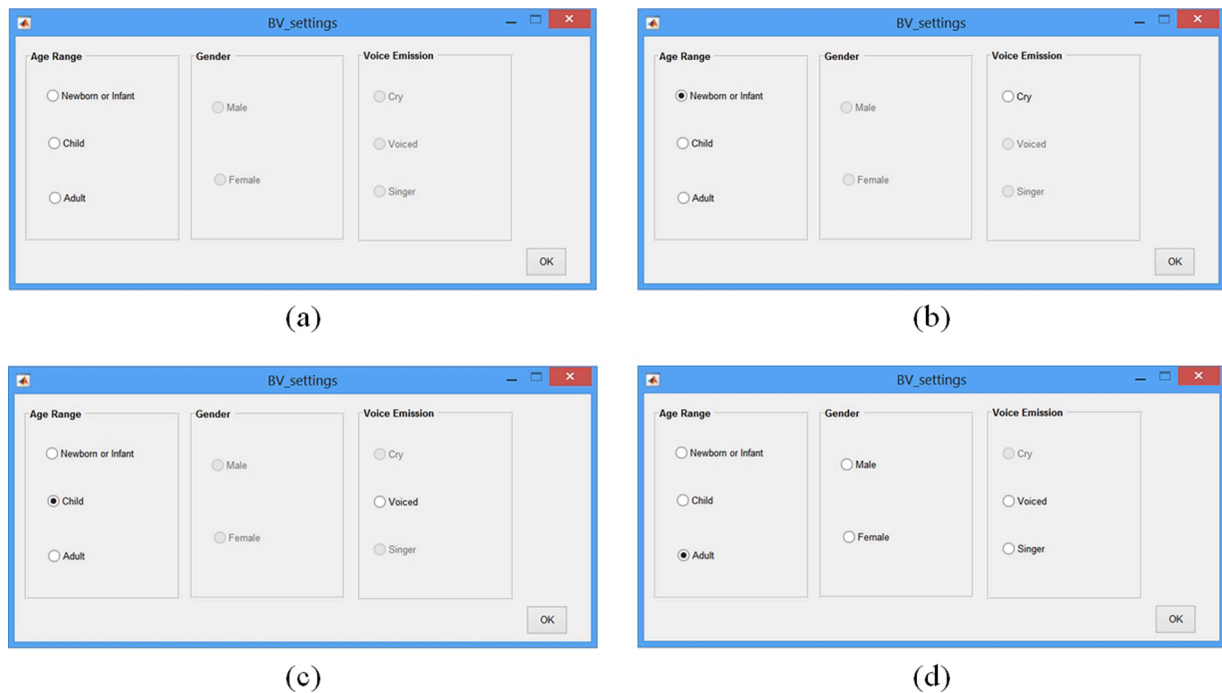


Fig. 7. Setting Interface. (a) Settings for age, gender, and type of voiced emission. (b) Setting for Newborn/Infant choice. (c) Setting for Child choice. (d) Setting for Adult choice.

First, it is necessary to select the age range among the three available options:

- Newborn / Infant
- Child
- Adult

For adults only, the gender of the individual's voice must be specified: Male or Female. In this case, the selectable vocal emissions are: Voiced or Singer. To avoid mistakes, notice that for each choice only the buttons of allowed sub-choices are active. That is, for example, if Newborn / Infant is selected, gender choice is not active and the type of vocal emission can only be Cry. The same for Child: gender is not active and the vocal emission is just Voiced. Once the required choices are made, the user clicks on OK and repeats the procedure for the next file, until the specifications for all the files are entered. If there are two or more files with the same specifications, they can be made in a single step simply holding down the "ctrl" key on the computer keyboard, selecting all the files having the same specifications, and then clicking on "Settings: . . .".

Once the specifications have been entered, voice analysis is launched by clicking the "START" button in the Acoustical Analysis box placed in the lower part of the interface (see Fig. 3):

Notice that, before launching the analysis, all Excel or .jpeg files must be closed. Otherwise, a warning is displayed (Fig. 8) and the elaboration is stopped until the files are closed. If BioVoice starts to analyze a file for which some settings were only partially fixed, or not fixed at all, the analysis stops and an error window appears warning the user about the

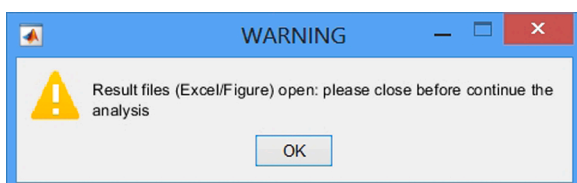


Fig. 8. Warning message: Excel files and figures must be closed.

missing information. The messages are shown in Fig. 9. Pressing the OK button, the user enters the settings for that file. The analysis restarts from the file where it was left off. The uploaded files are analyzed one at a time, in the same order in which they were uploaded. As soon as the analysis of a file begins, a new folder with the same name of the file, followed by "\_BIOVOICE", is created inside the same folder where the file is located. At the end of the analysis of the file, all the results obtained are saved in this new folder. BioVoice automatically proceeds with the analysis of the next file until all the uploaded files have been analyzed. To allow the user checking how the analysis proceeds, two bars are shown in the middle of the interface, respectively showing the number of voiced units analyzed and the file number in the list. An example is shown in Fig. 10. Voiced units are detected at the beginning of the elaboration according to a procedure described in detail in [18]. Notice that, to avoid unintentional pressure of other buttons during the file processing, only the "STOP" button is active. A figure displaying the detected voiced parts in the whole recording is saved. Examples are reported in the Results Section. If an already analyzed file is uploaded, BioVoice warns the user as shown in Fig. 11. If the user wants to proceed anyway with the analysis, current results will overwrite previous ones.

If more than 20 voiced frames are detected in a single file, BioVoice displays a message to inform the user that the figure showing the Voiced/UnVoiced (V/UV) selection will not be displayed nor saved. The reason being that the V/UV frames would be too compressed in the figure and thus unreadable. An empty figure is displayed instead, as shown in Fig. 12. This warning does not affect the processing of the file that goes on anyway, even if the warning message is not closed. This could be useful for example if the user is not watching the monitor during the elaborations.

The analysis can be stopped by selecting the "STOP" button in the Acoustical Analysis box. In this case, BioVoice completes the analysis of the current file but does not continue with that of the next file. A window opens that specifies which was the last analyzed file before the stop (Fig. 13). If the analysis is not stopped, BioVoice analyzes all the files in the stack and, once finished, the window shown in Fig. 14 is displayed. BioVoice informs the user that the analysis has been completed. There are three possible choices:



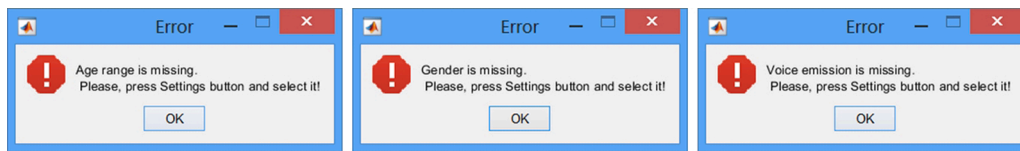


Fig. 9. Error messages for missing settings: (a) Missing Age range; (b) Missing Gender; (c) Missing Voice emission.

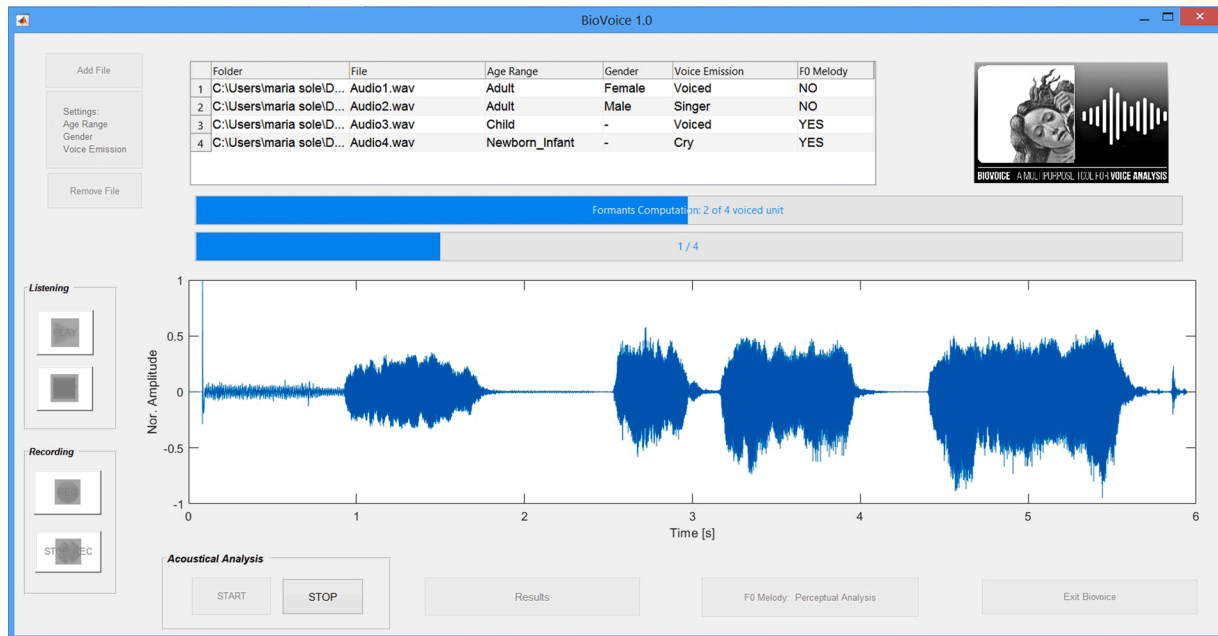


Fig. 10. BioVoice displays two progress bars showing the file and the vowel unit currently being processed.

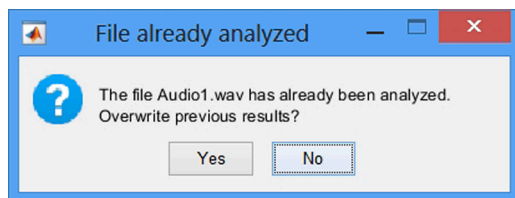


Fig. 11. BioVoice warning: the file has already been analyzed.

1. Select one of the analyzed files and click on the "Results" button to view the results obtained for that file; this opens the folder that contains .jpeg figures and Excel tables.
2. For files of newborns or children only, the user can select the "F0 Melody: Perceptual analysis" button to perform a manual melodic analysis in addition to the automatic one performed by BioVoice, as explained below.
3. Close BioVoice by selecting "Exit". The results are saved in their respective folders and can be viewed later.

### 2.3. Perceptual melodic analysis

At the end of the analysis of a newborn or child type file, selecting the button "F0 Melody: Perceptual analysis" in the interface, the window shown in Fig. 15(a) is displayed.

The button on the top: "HOW TO PERFORM PERCEPTUAL ANALYSIS? Click here" opens a short explanatory box, indicating how to proceed to perform the perceptual analysis (Fig. 15(b)). For correct use, the user must be already sufficiently informed and experienced about the infant cry analysis [20–26].

With BioVoice, for each voiced/cry unit (CU), the users may select the one that in their opinion is the most appropriate to describe the shape of F0 among twelve possible choices. The melodic shapes are: Falling (F), Rising (R), Symmetrical (S), Plateau (P), Low-Up (LU), Up-Low (UL), Frequency Step (FS), Double (D), Complex (C), Undefined (U), Not-a-Cry (NC) and Other (O). These shapes are a subset of those listed in [23,24]. To go to the next unit select "Next", to return to the previous one select "Prev". Once all the units have been analyzed, the window shown in Fig. 15(c) opens.

Clicking on "Results" opens the folder that, among other results, includes an Excel sheet with the comparison, CU by CU, between the automatic analysis carried out by BioVoice and the perceptual one performed manually by the operator. The sheet also shows the percentage of matching between the two. An example is shown in Fig. 15(d) where three Cry Units (voiced frames) are compared. A 100% match is obtained in this case.

To exit without completing the analysis, press "Interrupt Analysis". This may be necessary due to lack of time of the user or other reasons, however, the analysis can be restarted anytime from the point where it was stopped. Partial results are anyway saved.

### 2.4. Excel files and figures

The folder of each analyzed file contains several Excel files and pictures.

#### 2.4.1. Excel files

For any audio file named "audioname.wav", the following tables are created: "audioname\_F0.xls", "audioname\_F1.xls", "audioname\_F2.xls", "audioname\_F3.xls", "audioname\_Info.xls", "audioname\_PSD\_Units.xls", "audioname\_Voiced\_Unvoiced\_selection.xls", "audioname\_Voiced\_units.

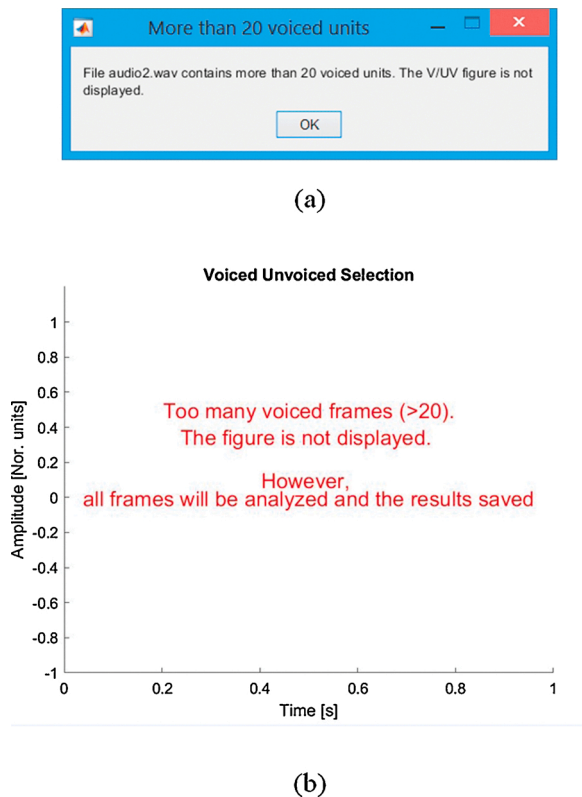


Fig. 12. More than 20 voiced units detected: (a) Warning message; (b) the V/UV selection plot is not displayed.

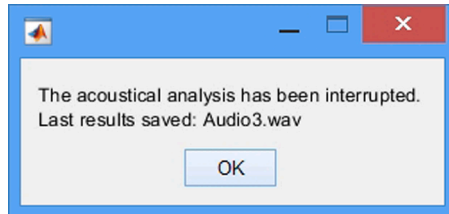


Fig. 13. BioVoice message when the STOP button is pressed.

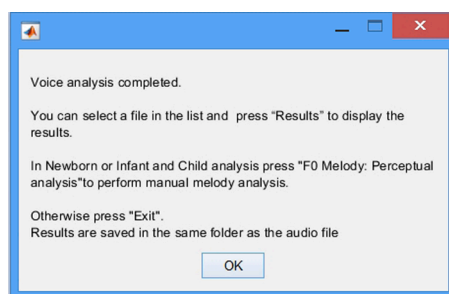


Fig. 14. A message is shown when all the .wav files are elaborated. Information is given about next steps.

xls" and "audioname\_Whole\_Recording\_Report.xls". For Newborn and Child one more excel file is created named "audioname\_Auto\_matic\_Melody\_Analysis.xls". Moreover, if the perceptual analysis is performed, the Excel file "audioname\_Automatic\_vs\_Perceptual\_Melody\_Analysis.xls" is created. This file contains a comparison, unit by unit, between the automatic analysis carried out by BioVoice and the perceptual one manually performed by the user, and the percentage of

matching between the two. Finally, in the case of a Singer voice, "audioname\_F4.xls" and "audioname\_F5.xls" are created and additional parameters (cents, quality ratio (QR), mean and standard deviation of vibrato rate, vibrato extent, vibrato jitter, and vibrato shimmer) are added in "Voiced\_Unit.xls" and "Whole\_Recording\_Report.xls". In Table 1, all the parameters saved in the Excel files are reported. Examples of the Excel files extracted from the analysis of a newborn cry are reported in supplementary materials. In the folder(s) containing the analyzed files, other Excel files are also created providing a report of all the results obtained for each category of adults, children, newborns, and singers, respectively. The Tables are named Report\_BIOVOICE\_Adult.xls, Report\_BIOVOICE\_Child.xls, Report\_BIOVOICE\_Newborn\_Infant.xls, and Report\_BIOVOICE\_Adult\_Singer.xls. Report tables are saved in the same folder as the analyzed audio files. Therefore, if some files for which a single report is required are in different folders, several tables will be created. It is therefore advisable that such files are collected in the same folder.

The implemented estimation methods for F0, formants, noise, and jitter are described in detail in [13–19].

For newborns and children, the estimation of the melodic shapes is described in [20,21,26]. For the singing voice, BioVoice implements the parameters described in [11].

#### 2.4.2. Figures

All figures are provided in .jpg format, to be accessible with any OS.

1. "audioname\_VUV.jpg" Picture in which the voiced units detected in the signal are highlighted by a dotted line. It returns an empty figure in the case of >20 units for which this figure would be illegible.
2. "audioname\_spectrogram.jpg" For each voiced unit a spectrogram is provided in colored scale with formant values superimposed.
3. "audioname\_F0.jpg" For each voiced unit a plot is provided showing the time evolution of F0. For clarity, the frequency scale is set according to the min-max F0 values.
4. In the singing voice case, for each voiced unit two more pictures are created: "audioname\_QR.jpg" displays the PSD with the low (red) and high frequency (green) parts, the two frequency thresholds and the QR value, while "audioname\_cents.jpg" shows the F0 plot in the cents scale along with the values of all the singing voice parameters (mean and std): Vrate, Vextent, C, Vocal intonation, Vibrato jitter and Vibrato shimmer.

The eventually large number of tables and figures concerning a single .wav file is one of the reasons for possible quite long computing time with BioVoice. The other reason being the time required for performing computations with some of its functions. Long computing time is indeed the main drawback of BioVoice.

#### 2.5. Real data analysis

To highlight the features of BioVoice, some examples are presented concerning the five categories: adult male, adult female, child, newborn, and singer. Apart from the singer, these examples concern voices recorded at home with cheap microphones or integrated into portable devices (i.e. smartphones). The singer's voice was recorded with a professional microphone Shure SM58 and a Tascam US-144 audio board. Results related to synthetic signals of both adults and newborns were already reported elsewhere. Specifically, the good performance of BioVoice in estimating F0, formants, jitter, and noise for synthetic adult voices was highlighted in [14–17], while newborn cry analysis of synthetic and real cries was described in detail in [20,21,27,28]. It is worth mentioning that recently BioVoice was successfully applied to newborn cry for differentiating the melodic shape of cry units of Italian, French, and Arabic newborns according to their mother tongue [26].

For adults, two groups were considered. Recordings from 8 male voices (age:  $37.7 \pm 9.3$  years old) and 8 female voices (age:  $37.2 \pm 11.8$

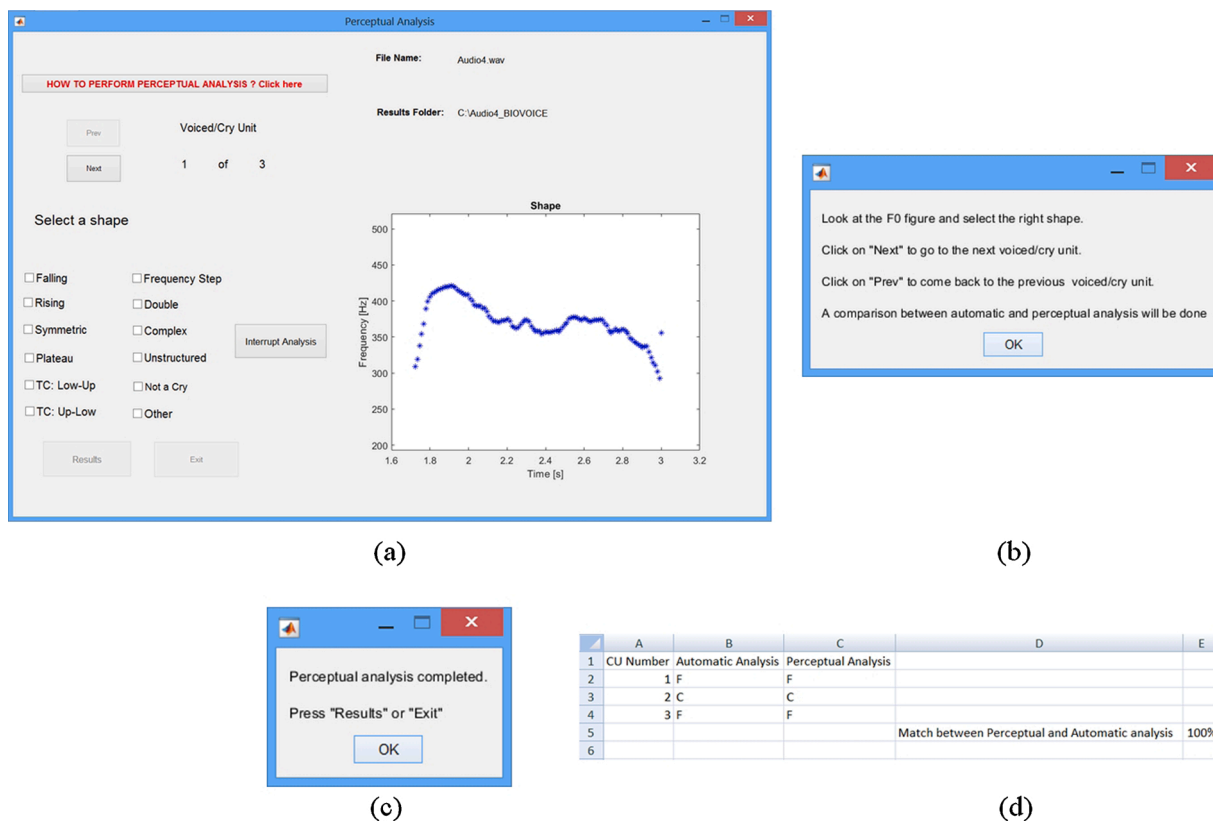


Fig. 15. Perceptual analysis (newborn cry or child’s voice): (a) Main interface; (b) explanatory box, (c) end of the analysis, (d) Comparison between automatic and perceptual analysis.

years old) were collected. All subjects were native Italian speakers and with healthy voices. Two children, a male and a female (though gender is not relevant in this case), both aged 4 years and native Italian speakers were recorded too. For adults and children, the vowel triangle is obtained from the estimated formant values F1 and F2. Results are compared to the vowel triangle reported in [29] and to the one obtained with Praat. The Mann-Whitney test was applied to compare results obtained with BioVoice and Praat. The same statistical test has been applied to compare the reference values reported in [29] to BioVoice and Praat, respectively. Moreover, the vowel space area (VSA) and the formant centralization ratio (FCR) [30] are computed. VSA represents the two-dimensional area bounded by lines obtained connecting first and second formant frequency coordinates (F1/F2) of vowels [31]. FCR is an acoustic metric that weighs formants with higher sensitivity to vowel centralization against formants that are expected to be lower [30]. VSA and FCR are acoustic metrics correlated to speech intelligibility. In general, clear and healthy voices present larger VSA values [32] and FCR values close to 0.9 [30]. FCR is expressed as  $(F2u + F2a + F1i + F1u)/(F2i + F1a)$  and VSA as  $ABS((F1i*(F2a - F2u) + F1a*(F2u - F2i) + F1u*(F2i - F2a))/2)$ , where F1a, F1i, F1u, F2a, F2i, and F2u represent the first and second formants of the vowels /a/, /i/ and /u/ [30].

Obviously, significant variability exists across speakers and languages. Features of American English are reported in [29], wherein vowels /a/ (/A/), /i/ (/I/) and /u/ (/U/) are represented and their pronunciation is reported. Although the Italian language slightly differs from both lowercase and uppercase vowels of American English, we consider such values as guidelines in this paper, as well as the same notation as in [29]. Therefore, the reference triangle is given by the averages of such values that are reported in Table 2 (upper case in brackets), where the correspondence between the vowels and the symbols of the International Phonetic Alphabet (IPA) is also shown. This

allows disregarding any specific characteristics of the Italian language that could distort the results based on the different ranges for F0 and formants in BioVoice and Praat. Table 2 also shows VSA and FCR for both lower case and upper case vowels.

Parameter settings used in BioVoice and Praat are reported in Table 3. Computations in Praat were performed by setting its default values as reported in the software tutorial [33]. The Praat tutorial does not indicate a specific value of the F0 and formant upper limit for children’s voices, therefore we used the same upper limit as in BioVoice. As suggested in the Praat tutorial [33], the number of formants to be estimated was set to 5 and the first three were selected. Instead, BioVoice automatically sets ranges, upper limits and the number of formants to three for adults, children and newborns, and to 5 for singers.

The example related to the newborn cry concerns a healthy at term subject. The recordings were performed before feeding, therefore they reasonably concern hunger cry. The first recording lasting about 10 s was made 30 days after birth. A second recording was made 15 days later. Both parents were native Italian speakers. Both automatic and perceptual analyses were performed. The perceptual analysis was made by an expert in the field of newborn cry analysis. Lastly, the example of singing voice concerns a professional male opera singer, aged 40 years, recorded while singing in operatic style a musical phrase from Gershwin’s opera Porgy and Bess: “Summertime, and the livin’ is easy.” For vibrato analysis, the sustained /aɪm/ from /sʌmətʌɪm/ was extracted. More details are reported in [11].

### 3. Results

#### 3.1. Adult results

The results of the acoustical analysis for adult males and females are reported in Table 4.



**Table 1**

List of all parameters and excel tables saved by BioVoice. All Excel files are saved adding the audio recording filename as a prefix.

	PSD_Unit_#.xls	Automatic_Melody_Analysis.xls	Automatic_vs_Perceptual_Melody_Analysis.xls	Info.xls	F0.xls	F1.xls	F2.xls	F3.xls	F4.xls	F5.xls	Voiced_Unit_#.xls	Voiced_Unvoiced_Selection.xls	Whole_recording_report.xls
Newborn	x	x	x	x	x	x	x	x			x	x	x
Child	x	x	x	x	x	x	x	x			x	x	x
Adult	x			x	x	x	x	x			x	x	x
Singer	x			x	x	x	x	x	x	x	x	x	x
<b>PARAMETERS</b>													
Age range (Newborn, Child, Adult)				x									
Automatic melody results		x											
Cent											x		x
F0					x								
<i>Frequency values [Hz]</i>					x								
<i>Time instants [s]</i>					x								
<i>Mean [Hz]</i>		x	x								x		x
<i>Median [Hz]</i>		x									x		x
<i>Std [Hz]</i>		x									x		x
<i>Min [Hz]</i>		x									x		x
<i>T (min)</i>											x		x
<i>Max[Hz]</i>		x									x		x
<i>T(max)</i>											x		x
<i>Skewness</i>		x											
<i>Kurtosis</i>		x											
<i>Percentiles (up to 100)</i>		x											
<i>Frequency values [Hz]</i>						x							
<i>Time instants [s]</i>						x							
<i>Mean [Hz]</i>											x		x
F1											x		x
<i>Median [Hz]</i>											x		x
<i>Std [Hz]</i>											x		x
<i>Min [Hz]</i>											x		x
<i>Max [Hz]</i>											x		x
<i>Frequency values [Hz]</i>							x						
<i>Time instants [s]</i>							x						
<i>Mean [Hz]</i>											x		x
F2											x		x
<i>Median [Hz]</i>											x		x
<i>Std [Hz]</i>											x		x
<i>Min [Hz]</i>											x		x
<i>Max [Hz]</i>											x		x
<i>Frequency values [Hz]</i>								x					
<i>Time instants [s]</i>								x					
<i>Mean [Hz]</i>											x		x
F3											x		x
<i>Median [Hz]</i>											x		x
<i>Std [Hz]</i>											x		x
<i>Min [Hz]</i>											x		x
<i>Max [Hz]</i>											x		x
<i>Frequency values [Hz]</i>									x				
<i>Time instants [s]</i>									x				
<i>Mean [Hz]</i>											x		x
F4											x		x
<i>Median [Hz]</i>											x		x
<i>Std [Hz]</i>											x		x
<i>Min [Hz]</i>											x		x
<i>Max [Hz]</i>											x		x
<i>Frequency values [Hz]</i>										x			
<i>Time instants [s]</i>										x			
<i>Mean [Hz]</i>											x		x
F5											x		x
<i>Median [Hz]</i>											x		x
<i>Std [Hz]</i>											x		x

(continued on next page)

Table 1 (continued)

	PSD_Unit_#. xls	Automatic_Melody _Analysis.xls	Automatic_vs_Perceptual _Melody_Analysis.xls	Info. xls	F0. xls	F1. xls	F2. xls	F3. xls	F4. xls	F5. xls	Voiced_Unit_#. xls	Voiced_Unvoiced _Selection.xls	Whole_recording _report.xls
Min [Hz]											x		x
Max [Hz]											x		x
Gender (Male, Female)				x									
Jitter											x		x
Kind of vocal emission (Cry, Voiced, Singer)				x									
Match between perceptual and automatic analysis			x										
NNE [dB]											x		x
Number													x
Pauses													x
Duration - mean [s]													x
Duration_std [s]													x
Duration_min [s]													x
Duration_max [s]													x
Perceptual analysis			x										
PSD		x											
Frequency [Hz]		x											
Normalized PSD [dB]		x											
QR											x		x
Total signal duration [s]													x
Vext											x		x
Mean											x		x
Std											x		x
Vibrato											x		x
Jitter											x		x
Shimmer											x		x
% Voiced													x
Duration [s]													x
Number													x
Voiced parts													x
Duration_mean [s]													x
Duration_std [s]													x
Duration_min [s]													x
Duration_max [s]													x
Start [s]											x	x	
Stop [s]											x	x	
Voiced frames											x	x	
Duration [s]												x	
Pause until next Episode [s]													x
Vrate											x		x
Mean											x		x
Std											x		x
Unit number/CU number		x									x		
			x										

Abbreviations: Fundamental frequency (F0), Formants (F1, F2, F3, F4, F5), Normalized Noise Energy (NNE), Power Spectral Density (PSD), Quality Ratio (QR).

**Table 2**

First two formant reference values for both lower case and upper case (in brackets) vowels [29].

American English	IPA Symbols	F1 [Hz]	F2 [Hz]	VSA	FCR
/a/ (/A/)	ɑ (ʌ)	730 (640)	1090 (1190)		
/i/ (/I/)	ɪ (ɪ)	270 (390)	2290 (1990)	20,9700 (2,09,150)	0.84 (1.16)
/u/ (/U/)	ʊ (ʊ)	300 (440)	870 (1020)		

VSA and FCR are shown in the last two columns.

**Table 3**

F0 and formant settings used in BioVoice and in Praat.

	BioVoice		Praat	
	F0 range [Hz]	Formant upper limit [kHz]	F0 range [Hz]	Formant upper limit [kHz]
Adult male	50–250	5.5	75–300	5.0
Adult female	100–350	5.5	100–500	5.5
Child	110–600	7.0	110–600	7.0
Newborn	150–900	10.0	—	—
Male singer	50–500	7.0	—	—
Female singer	100–1200	7.0	—	—

In Fig. 16, the vowel triangles for adult male and female are shown. Furthermore, VSA and FCR are computed, as well as the comparisons between the two software tools and the reference values. Specifically, for male subjects  $VSA = 168082.93 \pm 68305.74$  ( $p = 0.74$ ) and  $FCR = 1.05 \pm 0.14$  ( $p = 0.55$ ) with BioVoice, while with Praat  $VSA = 147433.41 \pm 79922.75$  ( $p = 0.31$ ) and  $FCR = 1.20 \pm 0.13$  ( $p = 0.14$ ). For females:  $VSA = 287010.54 \pm 120663.75$  ( $p = 0.08$ ) and  $FCR = 0.99 \pm 0.12$  ( $p = 0.84$ ) with BioVoice, while with Praat:  $VSA = 247355.95 \pm 131045.22$  ( $p = 0.46$ ) and  $FCR = 1.14 \pm 0.18$  ( $p = 0.11$ ).

**3.2. Children results**

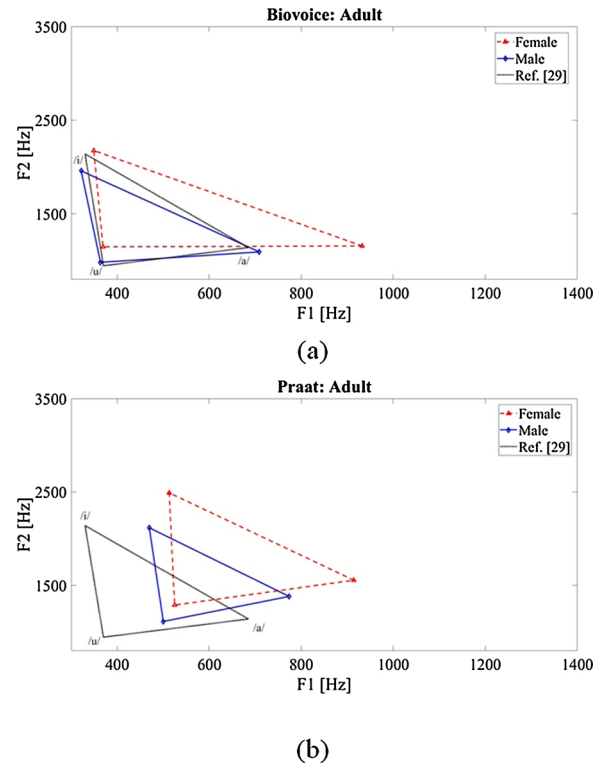
The children results are shown in Fig. 17 and Table 5. With Praat, the first three formants out of five were selected. The F0 range and formants upper limit were selected in order to optimize the results. They will be described in detail in the next section. Statistical analysis was not

**Table 4**

Mean and standard deviation (SD) of formant frequencies for adult males and females estimated using BioVoice and Praat. Statistical comparisons between BioVoice and Praat are reported as well as the comparisons between the two software tools and the reference values (ref. values) [29].

		Formant Frequencies [Hz]		Statistical Analysis (p-value)		
		BioVoice (mean ± SD)	Praat (mean ± SD)	BioVoice vs. Praat	BioVoice vs. ref. values [27]	Praat vs. ref. values [27]
<b>MALE</b>						
/a/	F1	708.57 ± 86.24	773.49 ± 60.06	0.04*	0.46	0.01*
	F2	1093.23 ± 82.44	1381.79 ± 104.93	0.01*	0.15	0.01*
/i/	F1	321.53 ± 26.71	469.59 ± 88.49	0.01*	0.38	0.01*
	F2	1958.81 ± 279.17	2116.92 ± 102.04	0.20	0.11	0.55
/u/	F1	363.55 ± 24.86	500.10 ± 102.06	0.01*	0.84	0.01*
	F2	981.74 ± 86.21	1113.30 ± 153.74	0.02*	0.25	0.04*
<b>FEMALE</b>						
/a/	F1	932.27 ± 91.31	914.12 ± 53.82	0.31	0.01*	0.01*
	F2	1156.39 ± 61.32	1554.96 ± 137.69	0.01*	0.64	0.01*
/i/	F1	349.01 ± 41.68	512.93 ± 105.05	0.01*	0.20	0.01*
	F2	2175.96 ± 410.32	2490.76 ± 162.37	0.11	1.00	0.01*
/u/	F1	369.05 ± 50.25	524.62 ± 129.14	0.01*	1.00	0.01*
	F2	1147.39 ± 51.75	1288.50 ± 289.43	0.38	0.01*	0.01*

\* p-value <0.05.



**Fig. 16.** Comparison between BioVoice (a) and Praat (b): vowel triangle of adult males (solid line) and females (dashed line). Reference values are reported in Table 2 (dotted line).

performed in this case due to the limited number of participants included in this category.

**3.3. Newborn results**

In the first recording, eight voiced parts (Cry Units, CU) were detected. CUs obtained with BioVoice are shown in Fig. 18. The Excel table with the comparison between automatic and perceptual melodic analyses is displayed in Table 6. The third CU is classified as Complex (C) by the automatic method, but perceptually as a Falling (F) shape. Therefore the match is 88%. The two shapes are shown in Fig. 19. In the analysis of the first audio recording, most of the CUs were classified as Falling cries.

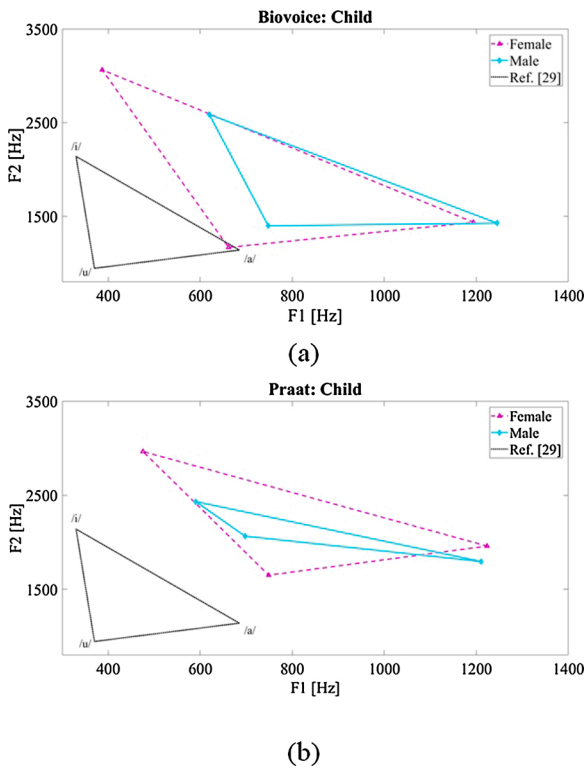


Fig. 17. Comparison between BioVoice (a) and Praat (b): vowel triangle of a male (solid line) and a female child (dashed line). Reference values are reported in Table 2 (dotted line).

Table 5  
Formant frequencies for a male child and a female child.

		Formant frequencies [Hz]	
		BioVoice (mean)	Praat (mean)
<b>MALE</b>			
/a/	F1	1193.70	1210.66
	F2	1435.74	1795.08
/i/	F1	386.84	590.3
	F2	3063.89	2432.03
/u/	F1	661.42	697.44
	F2	1168.26	2064.74
<b>FEMALE</b>			
/a/	F1	1245.13	1223.37
	F2	1427.32	1959.45
/i/	F1	619.50	475.4
	F2	2587.46	2964.95
/u/	F1	747.86	748.66
	F2	1398.67	1649.67

The second recording was made 15 days later and lasts about 10 s. In this case, three CUs were detected, as shown in Fig. 20, and classified as reported in Table 7. The match is now 100%.

As an example, the second CU is displayed in Fig. 21. It is correctly classified as C both automatically and perceptually. Though only three CUs were detected, the majority of C shapes might indicate that the newborn cry evolves with age towards more complex sounds such as babbling and pre-speech, in agreement with [22].

3.4. Singer results

Vibrato is clearly visible in Fig. 22 and its main parameters are shown: vibrato rate, vibrato extent, vocal intonation, vibrato jitter, vibrato shimmer, and cents (reported in Fig. 22(a) respectively as VRate, VExtent, Vocal Int., C, Vibrato Jitt. and Vibrato shimm). They are also

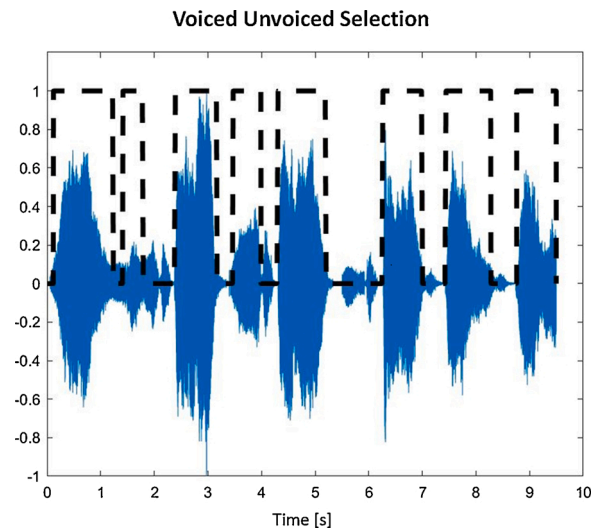


Fig. 18. Voiced/UnVoiced (V/UV) selection obtained with BioVoice. Eight cry units (CUs) are found. They are indicated with a black dotted line.

Table 6  
Melodic shapes: comparison between automatic and perceptual analyses.

CU Number	Automatic Analysis	Perceptual Analysis	F0 Mean
1	F	F	367.07
2	S	S	263.31
3	C	F	466.51
4	S	S	318.23
5	F	F	447.79
6	F	F	436.99
7	F	F	380.48
8	F	F	384.05

collected in Excel tables in the folder created by BioVoice for this subject. Fig. 22(b) shows the quality ratio QR, computed as the ratio between the cumulative energy of F1 and F2 divided by that of F3-F5. A smoothed Normalized Power Spectral Density (PSD) is shown, to enhance the energy peaks that correspond to the first 5 formants. The two PSD areas are well distinguishable ( $F1-F2 < Fref\_inf$ ,  $Fref\_inf < F3-F5 < Fref\_sup$ ) as well as the two dynamic thresholds as obtained with BioVoice [11]. The spectrogram in Fig. 22(c) shows the time evolution of F1-F5. The fundamental frequency (first harmonic, around 350 Hz) and its multiples are clearly visible, as well as their oscillating shape related to vibrato.

4. Discussion

In this paper, the BioVoice software tool for voice analysis is presented, its innovative features and ease of use are described in detail. Examples of real data analysis are presented to highlight the software performance and its features also as compared to Praat, the currently most used software tool for voice analysis. The comparison between BioVoice and Praat was performed only for adults and children because, on one hand, newborns and infants cannot be asked to produce a specific vowel emission and thus there are no reference vowels for comparison, and on the other hand because Praat does not estimate the cry melody or the parameters of the singing voice. However, comparisons of synthetic newborn cries have been made and published in [20,21,23]. Other software tools perform newborn cry analysis, but they are neither available online nor freely downloadable [34–36].

In the vowel triangle for adult’s voice analysis, higher values for F1 and F2 are obtained with respect to the reference values with Praat, both for males and females. With BioVoice, only the female group exhibits very high F1 values for the vowel /a/. This is however quite consistent

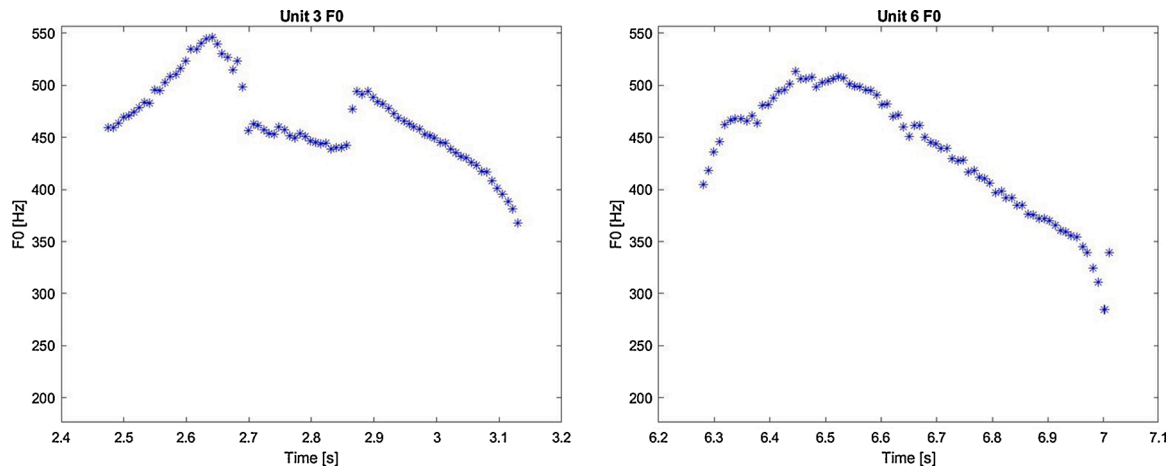


Fig. 19. F0 plot of the 2nd and the 6th CU computed with BioVoice. (a) The 3rd CU is classified as Complex (C) by the automatic method, but perceptually as Falling (F). (b) Falling shape is clearly visible for the 6th CU.

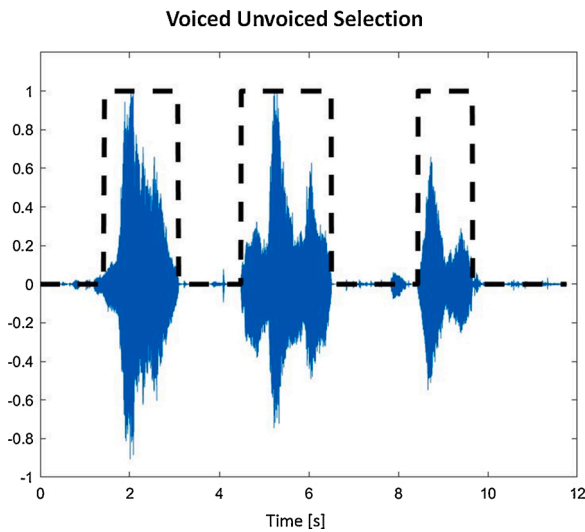


Fig. 20. Voiced/UnVoiced (V/UV) selection obtained with BioVoice. Three cry units (CUs) are found. They are indicated with a dotted line.

Table 7  
Melodic shapes: comparison between automatic and perceptual analyses.

CU Number	Automatic Analysis	Perceptual Analysis	F0 Mean
1	F	F	369.38
2	C	C	362.43
3	C	C	314.11

with the differences in the vocal tract between males and females.

The comparative analysis between Praat and BioVoice (Table 4) showed statistically significant differences in the formant frequency estimations of all three vowels. Instead, there are only a few statistically significant differences when BioVoice is compared to the reference values.

The acoustic metrics given by VSA and FCR shows the reliability of formant frequencies algorithms implemented in BioVoice. VSA and FCR values obtained through BioVoice are slightly better aligned to the reference values [29] than those obtained with Praat, except for the VSA of female voices. For most parameters, the statistical comparison between BioVoice and the reference values in [29] shows a high correlation (p-value up to 0.84), while slightly low values were found with Praat (p-value < 0.46).

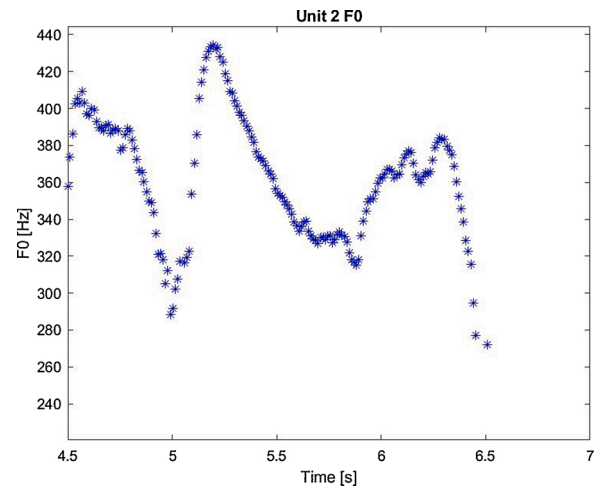


Fig. 21. F0 plot of the 2nd CU computed with BioVoice. The Complex shape is clearly visible.

As the adult participants included in this work had healthy voices, it worth noticing that the FCR values obtained with BioVoice are closer to 0.9 as should be in control subjects with clear speech [30].

In previous papers [19,37,38], the analysis of sustained vowels emitted by healthy adult subjects gave almost identical results for F0 both with BioVoice and Praat, and quite similar results are found for the first two formants. On the other hand, as regards the third and subsequent formants, the calculation is certainly more complex due to the lower energy and higher bandwidth of these formants [29] and results often differ. Praat gives the users the possibility to work around this issue, by empirically increasing the number of formant frequencies to be estimated and then choosing the first three, and/or changing the frequency range settings. This requires some skill of the user, who must decide which is the best action to take looking at the spectrogram. As already mentioned, BioVoice does not require any actions from the users, thus providing standardized results among each voice category. Indeed the estimation of formants with numerical methods is still challenging and further improvements have to be developed and implemented. On this topic, some papers on simulated data show the robustness of the methods implemented in BioVoice [14–17,19].

Generally, the vowel triangles of adults obtained with the values estimated by BioVoice are similar to the reference triangle, while those of children show higher formant frequencies. This difference is consistent with the higher frequencies in children’s voices, as the vocal tract is



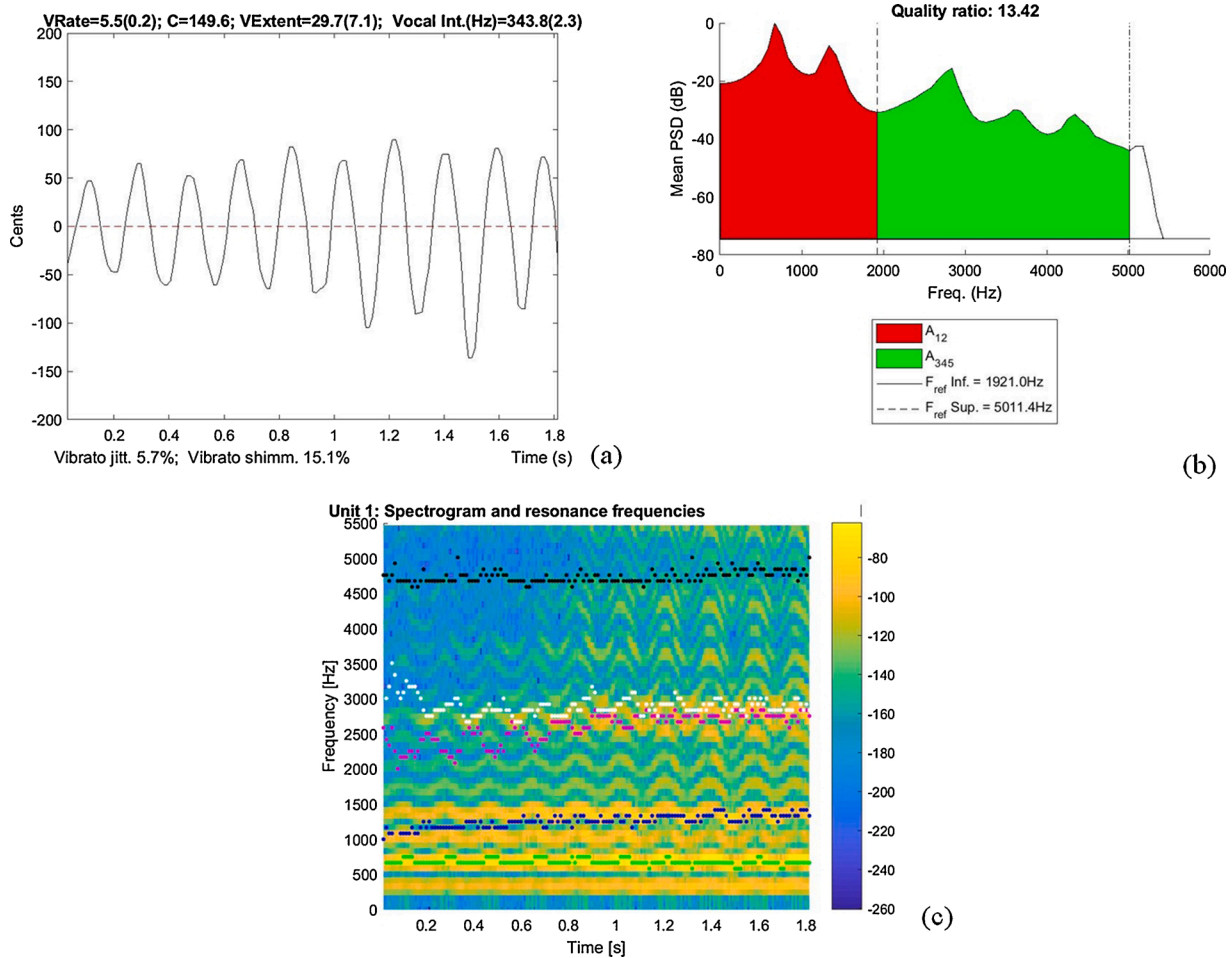


Fig. 22. Singing voice analysis. Vowel /a/ with vibrato. Upper left: F0 shape in cents; upper right: Power Spectral Density and QR; lower: spectrogram with F1-F5 superimposed.

shorter than in adults.

An example is shown in Fig. 17 and in Table 5 concerning children’s vocalizations. As can be seen, the estimates provided by BioVoice and Praat differ significantly from each other, especially as regards F2, which is generally higher with Praat. The reference values reported in Table 3 were used for BioVoice, while for Praat several attempts were made: in fact, varying both the minimum of F0 ( $75\text{ Hz} < F_{0\text{min}} < 110\text{ Hz}$ ), the number of formants  $F_{\text{nmax}}$  ( $5 < F_{\text{nmax}} < 8$ ) and their range  $F_{\text{range}}$  ( $5.5\text{ kHz} < F_{\text{range}} < 8\text{ kHz}$ ) results were found remarkably dissimilar to each other. The result shown in Fig. 17 was obtained by setting  $F_{0\text{min}} = 110\text{ Hz}$ ,  $F_{\text{nmax}} = 5$  and  $F_{\text{range}} = 7\text{ kHz}$ .

The main innovative part of BioVoice presented in this paper consists of the melodic analysis for newborns and children and the singing voice analysis. As an example, Fig. 15 concerning the newborn cry shows the difference between automatic and perceptual analyses in the third melodic shape. Difficult to say which of the two is the correct one. The classification of the cry units (CUs) is indeed a quite hard task, both numerically and perceptually, and there are few clinically assessed relationships between melodic shapes and neurological conditions in the newborn [23,27,34,35]. Recently, the study of melodic shapes in the neonatal cry has proven to be an effective tool in characterizing linguistic differences between infants whose mothers have different mother tongues [25,26,39,40]. This aspect is relevant as it allows verifying the newborns’ ability to listening to external sounds in the last weeks of gestation, and their learning ability before birth. Therefore, this option offered by BioVoice might be an added value and the comparison between automatic and perceptual analysis could help with

solving doubtful situations.

The singing voice, as high pitched as that of the newborn, requires specific quality parameters. In this case, the greatest difficulty lies in generalizing the calculation of the QR, which fully depends on the vocal characteristics of the subject. To this aim, variable thresholds are implemented to separate F1 and F2 from F3-F5, according to the specific vocal emission [11]. Indeed, the problem of a reliable estimation of formants is particularly critical here, where up to the fifth formant is required.

It is obvious that the difficulties in both newborn cry and singing voice are considerable, as these voices have a very high pitch and peculiar characteristics. Moreover, the newborn cry is an extremely complex signal, because it is always characterized by high irregularity and the recordings are made without the possibility of asking the subject to emit a specific vocalization.

Finally, we remark that the methods implemented for adult and singing voices can also be applied in the case of actor/actress voice. Recently, BioVoice was applied with the aim of evaluating possible improvements of the performance in the use of semi-occluded vocal tract exercises [41,42]. Moreover, the acoustical analysis of the singing voice could usefully combine imaging techniques to deepen the study of phonatory mechanisms in singing practice [43].

4.1. Study limitations

With BioVoice the required computational time for elaborations is highly dependent on the duration of the audio file, but even more on the

number of detected voiced parts: indeed, the many acoustical parameters are computed for each voiced unit, as well as figures and tables. This inevitably slows down the completion of the processing. For example, 38 s were needed to carry out the analysis of the newborn recording described in Figs. 20 and 21 and Table 7 on a 64-bit Windows 10 laptop equipped with Intel Core i7-3632QM CPU @ 2.20 GHz and 8GB of RAM.

In general, the computing time is affected by two main factors: the number of time windows on which F0 and formants are calculated, whose length is inversely proportional to the variable frequency of F0 [13,21] and therefore higher in the case of infants, children and singers; and the further processing required for the automatic analysis of the melody.

In the current version of BioVoice, child and adult voice analysis is limited to voiced frames only or vocalic words [37]. We are working towards including running speech analysis for both groups. Lastly, up to date, the melodic shape is computed only for newborn cry and children's voice, but it will be extended to the adult's voice too, as it might be related to the emotional state of the subject.

Reducing computational time, code optimization, and the implementation of additional voice features will be included in future BioVoice releases.

Finally, in this work, we did not address the possible interaction between F0 and the first formant in high pitched voices, that might occur in adult voice and more often in newborns and children's voices. As a matter of fact, in BioVoice the algorithms for the computation of F0 and formants are implemented in such a way as to avoid possible interferences between them, but this topic will require further and in-depth studies.

## 5. Conclusions

This study presents and explains a user-friendly software tool for the acoustical analysis of the human voice named BioVoice. Biovoice estimates more than 20 acoustical parameters with advanced and robust analysis techniques for different vocal emissions, from the newborn to the adult and the singer.

The software is the result of several years of research and some of its features and results were partially presented in previous papers. Here, we described in detail and in a unified way all the capabilities of BioVoice in its recently available online version. With respect to previous versions, new relevant features of BioVoice are also introduced concerning the analysis of the newborn cry and of the singing voice.

Basic parameters are computed as well as new ones, specifically developed for newborn cry and singing voice analysis not provided by other software tools of the same kind. For newborn cry and child voice, BioVoice computes the melodic shape of F0 out of 12 basic shapes and allows performing perceptual analysis. In the singing voice case, formants up to F5 are computed as well as the quality ratio QR and parameters concerning vibrato and its regularity.

Great effort and attention have been paid to the creation of a simple and intuitive interface, which completely frees the user from setting technical parameters often difficult to understand, that might lead to incorrect use of the software and inaccurate results. BioVoice is thus a valuable tool including new relevant features as compared to other existing software tools and suitable also for the inexperienced user.

## CRedit authorship contribution statement

**Maria Sole Morelli:** Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Silvia Orlandi:** Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Claudia Manfredi:** Conceptualization, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration.

## Declaration of Competing Interest

The authors report no declarations of interest.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.bspc.2020.102302>.

## References

- [1] S. Fex, Perceptual evaluation, *J. Voice* 6 (1992) 155–158, [https://doi.org/10.1016/S0892-1997\(05\)80130-4](https://doi.org/10.1016/S0892-1997(05)80130-4).
- [2] Kay Elemetrics Corporation, Operations Manual: Multi-Dimensional Voice Program (MDVP) Model 4305, Kay Elemetrics Corporation, Lincoln Park, NJ, 1993.
- [3] Dr.Speech, 2011 <http://www.drspeech.com/>. (Accessed 2 October 2020).
- [4] WEVOSYS Home English, 2018 <https://www.wevosys.com/>. (Accessed 2 October 2020).
- [5] User Manual – Sygyt Software, 2020 <https://www.sygyt.com/en/documentation/>. (Accessed 2 October 2020).
- [6] TMH KTH: WaveSurfer, 2020 <http://www.speech.kth.se/wavesurfer/>. (Accessed 2 October 2020).
- [7] Speech Analyzer - SIL Language Technology, 2018 <https://software.sil.org/speech-analyzer/>. (Accessed 2 October 2020).
- [8] Windows Tool for Speech Analysis, 2020 <https://www.phon.ucl.ac.uk/resource/sfs/wasp/>. (Accessed 2 October 2020).
- [9] Vocal Analysis Tool | SPLab, 2016 <http://splab.cz/en/download/software/software-pro-analyzu-fonace>. (Accessed 2 October 2020).
- [10] P. Boersma, Praat, a system for doing phonetics by computer, *Glott Int.* 5 (2001) 341–347.
- [11] C. Manfredi, D. Barbagallo, G. Baracca, S. Orlandi, A. Bandini, P.H. Dejonckere, Automatic assessment of acoustic parameters of the singing voice: application to professional western operatic and jazz singers, *J. Voice* 29 (2015) 517.e1–517.e9, <https://doi.org/10.1016/j.jvoice.2014.09.014>.
- [12] M.S. Morelli, S. Orlandi, C. Manfredi, BioVoice: a multipurpose tool for voice analysis. *Model. Anal. Vocal Emiss. Biomed. Appl.*, 2019.
- [13] C. Manfredi, L. Bocchi, G. Cantarella, A multipurpose user-friendly tool for voice analysis: application to pathological adult voices, *Biomed. Signal Process. Control* 4 (2009) 212–220, <https://doi.org/10.1016/J.BSPC.2008.11.006>.
- [14] P. Dejonckere, J. Schoentgen, A. Giordano, S. Fraj, L. Bocchi, C. Manfredi, Validity of jitter measures in non-quasi-periodic voices. Part I: perceptual and computer performances in cycle pattern recognition, *Logop. Phoniatri. Vocol.* 36 (2011) 70–77, <https://doi.org/10.3109/14015439.2011.578078>.
- [15] C. Manfredi, A. Giordano, J. Schoentgen, S. Fraj, L. Bocchi, P. Dejonckere, Validity of jitter measures in non-quasi-periodic voices. Part II: the effect of noise, *Logop. Phoniatri. Vocol.* 36 (2011) 78–89, <https://doi.org/10.3109/14015439.2011.578077>.
- [16] P.H. Dejonckere, A. Giordano, J. Schoentgen, S. Fraj, L. Bocchi, C. Manfredi, To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition, *Biomed. Signal Process. Control* 7 (2012) 37–42, <https://doi.org/10.1016/j.bspc.2011.05.002>.
- [17] C. Manfredi, A. Giordano, J. Schoentgen, S. Fraj, L. Bocchi, P.H. Dejonckere, Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools, *Biomed. Signal Process. Control* 7 (2012) 409–416, <https://doi.org/10.1016/j.bspc.2011.06.004>.
- [18] S. Orlandi, P.H. Dejonckere, J. Schoentgen, J. Lebacqz, N. Rruqja, C. Manfredi, Effective pre-processing of long term noisy audio recordings: an aid to clinical monitoring, *Biomed. Signal Process. Control* 8 (2013) 799–810.
- [19] N. Rruqja, P.H. Dejonckere, G. Cantarella, J. Schoentgen, S. Orlandi, S. D. Barbagallo, C. Manfredi, Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia, *Biomed. Signal Process. Control* 13 (2014) 71–78, <https://doi.org/10.1016/J.BSPC.2014.03.011>.
- [20] C. Manfredi, A. Bandini, D. Melino, R. Viellevoe, M. Kalenga, S. Orlandi, Automated detection and classification of basic shapes of newborn cry melody, *Biomed. Signal Process. Control* 45 (2018) 174–181, <https://doi.org/10.1016/J.BSPC.2018.05.033>.
- [21] S. Orlandi, A. Bandini, F.F. Fiaschi, C. Manfredi, Testing software tools for newborn cry analysis using synthetic signals, *Biomed. Signal Process. Control* 37 (2017) 16–22, <https://doi.org/10.1016/j.bspc.2016.12.012>.
- [22] K. Wermke, W. Mende, C. Manfredi, P. Bruscoloni, Developmental aspects of infant's cry melody and formants, *Med. Eng. Phys.* 24 (2002) 501–514.
- [23] G. Várallyay Jr., Infant cry analyzer system for hearing disorder detection, *Periodica Polytechnica, TU Timișoara, Trans. Autom. Control Comput. Sci.* 49 (2004) 57–60.
- [24] G. Várallyay, The melody of crying, *Int. J. Pediatr. Otorhinolaryngol.* 71 (2007) 1699–1708, <https://doi.org/10.1016/J.IJPO.2007.07.005>.
- [25] B. Mampe, A.D. Friederici, A. Christophe, K. Wermke, Newborns' cry melody is shaped by their native language, *Curr. Biol.* 19 (2009) 1994–1997.
- [26] C. Manfredi, R. Viellevoe, S. Orlandi, A. Torres-García, G. Pieraccini, C.A. Reyes-García, Automated analysis of newborn cry: relationships between melodic shapes and native language, *Biomed. Signal Process. Control* 53 (2019) 101516, <https://doi.org/10.1016/j.bspc.2019.101561>.

- [27] S. Orlandi, C.A. Reyes García, A. Bandini, G. Donzelli, C. Manfredi, Application of pattern recognition techniques to the classification of full-term and preterm infant cry, *J. Voice* 30 (2016) 656–663, <https://doi.org/10.1016/j.jvoice.2015.08.007>.
- [28] S. Orlandi, L. Bocchi, G. Donzelli, C. Manfredi, Central blood oxygen saturation vs crying in preterm newborns, *Biomed. Signal Process. Control* 7 (2012) 88–92, <https://doi.org/10.1016/J.BSPC.2011.07.003>.
- [29] J.R. Deller, J.H.L. Hansen, J.G. Proakis, *Discrete-Time Processing of Speech Signals*, 2000.
- [30] S. Sapir, L.O. Ramig, J.L. Spielman, C. Fox, Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech, *J. Speech Lang. Hear. Res.* 53 (2010) 114–125, [https://doi.org/10.1044/1092-4388\(2009/08-0184\)](https://doi.org/10.1044/1092-4388(2009/08-0184)).
- [31] S. Sandoval, V. Berisha, R.L. Utianski, J.M. Liss, A. Spanias, Automatic assessment of vowel space area, *J. Acoust. Soc. Am.* 134 (2013) EL477–EL483, <https://doi.org/10.1121/1.4826150>.
- [32] A.R. Bradlow, T. Bent, The clear speech effect for non-native listeners, *J. Acoust. Soc. Am.* 112 (2002) 272–284, <https://doi.org/10.1121/1.1487837>.
- [33] P. Boesma, *Acoustic analysis*, in: R.J. Podesva, D. Sharma (Eds.), *Res. Methods Linguist.*, Cambridge University Press, 2014.
- [34] M.A. Ruiz Díaz, C.A. Reyes García, L.C. Altamirano Robles, J.E. Xalteno Altamirano, A. Verduzco Mendoza, Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis, *Biomed. Signal Process. Control* (2012) 43–49, <https://doi.org/10.1016/j.bspc.2011.06.011>. Elsevier.
- [35] H. Ganek, A. Eriks-Brophy, Language ENvironment Analysis (LENA) system investigation of day long recordings in children: a literature review, *J. Commun. Disord.* 72 (2018) 77–85, <https://doi.org/10.1016/J.JCOMDIS.2017.12.005>.
- [36] B. Reggiani, S.J. Sheinkopf, H.F. Silverman, X. Li, B.M. Lester, A flexible analysis tool for the quantitative acoustic assessment of infant cry, *J. Speech Lang. Hear. Res.* 56 (2013) 1416–1428, [https://doi.org/10.1044/1092-4388\(2013/11-0298\)](https://doi.org/10.1044/1092-4388(2013/11-0298)).
- [37] A. Bandini, F. Giovannelli, S. Orlandi, S.D. Barbagallo, M. Cincotta, P. Vanni, R. Chiaramonti, A. Borgheresi, G. Zaccara, C. Manfredi, Automatic identification of dysprosody in idiopathic Parkinson's disease, *Biomed. Signal Process. Control* 17 (2015) 47–54, <https://doi.org/10.1016/j.bspc.2014.07.006>.
- [38] E. Biondi, A. Bandini, L. Lombardo, S. Orlandi, G. Siciliani, C. Manfredi, Phonetic analysis during treatment with rapid maxillary expander, *Orthod. Craniofacial Res.* 20 (2017) 21–29, <https://doi.org/10.1111/ocr.12136>.
- [39] A. Prochnow, S. Erlandsson, V. Hesse, K. Wermke, Does a 'musical' mother tongue influence cry melodies? A comparative study of Swedish and German newborns, *Music. Sci.* 23 (2019) 143–156, <https://doi.org/10.1177/1029864917733035>.
- [40] K. Wermke, Y. Ruan, Y. Feng, D. Dobnig, S. Stephan, P. Wermke, L. Ma, H. Chang, Y. Liu, V. Hesse, H. Shu, Fundamental frequency variation in crying of mandarin and German neonates, *J. Voice* 31 (2017) 255.e25–255.e30, <https://doi.org/10.1016/j.jvoice.2016.06.009>.
- [41] V. Di Natale, G. Cantarella, C. Manfredi, A. Ciabatta, C. Bacherini, P.H. Dejonckere, Short term effect of "Semioccluded Vocal Tract Exercises" on healthy actors' voices, in: 11th Int. Work. Model. Anal. Vocal Emiss. Biomed. Appl. MAVIBA 2019, Firenze University Press, 2019, pp. 167–170, <https://doi.org/10.36253/978-88-6453-961-4>.
- [42] V. Di Natale, G. Cantarella, C. Manfredi, A. Ciabatta, C. Bacherini, P. H. DeJonckere, Semioccluded vocal tract exercises improve self-perceived voice quality in healthy actors, *J. Voice* (2020), <https://doi.org/10.1016/j.jvoice.2020.07.024>.
- [43] P.H. Dejonckere, J. Lebacqz, L. Bocchi, S. Orlandi, C. Manfredi, Automated tracking of quantitative parameters from single line scanning of vocal folds: a case study of the "messa di voce" exercise, *Logop. Phoniater. Vocol.* 40 (2015) 44–54, <https://doi.org/10.3109/14015439.2013.861014>.