ARE PRAAT’S DEFAULT SETTINGS OPTIMAL FOR INFANT CRY ANALYSIS?

Giulio Gabrieli
Psychology Program, School of Social Sciences
Nanyang Technological University, Singapore
GIULIO001@e.ntu.edu.sg

Andrea Bizzego
Department of Psychology and Cognitive Science
University of Trento, Italy
andrea.bizzego@unitn.it

Wan Qing Leck
Psychology Program, School of Social Sciences
Nanyang Technological University, Singapore
LECK0006@e.ntu.edu.sg

Gianluca Esposito
Psychology Program, School of Social Sciences
Nanyang Technological University, Singapore
Department of Psychology and Cognitive Science
University of Trento, Italy
qianluca.esposito@ntu.edu.sg
qianluca.esposito@unitn.it

ABSTRACT

In recent years, the number of studies investigating possible non-invasive health screening techniques for infants has increased exponentially. Amongst those, one of the most prominent is health screening based on the acoustic investigation of infant cry. Clinicians involved in the field moved from visual inspection of the audible spectrum to automatized analysis of cry samples using computer software. A software that has been more widely adopted in recent years is Praat, a free software designed for speech analysis. Unfortunately, the software’s default settings are not suitable for investigation of cry samples, yet rarely used settings are reported in final manuscripts. In this article, we tested 4 different computer generated signals, with frequency features comparable to cry frequencies, and 3 real cry samples using both Praat’s standards and tuned settings. Our results highlight the importance of properly tuning software’s parameters when expanding their field of usage, and provide a starting point for the development of optimal Praat algorithm’s parameters selection for cry analysis.

1. INTRODUCTION

Screening of infants’ health statuses can lead to early recognition of developmental pathologies, this allows clinicians to define an intervention program, which can lead to enhanced outcomes when adopted in earlier stages of life. Among infants’ health screening methods, non-invasive techniques received the highest level of attention within the community of pediatricians and researchers. Starting from the second half of the Twentieth Century, researchers investigated several possible ways to identify different pathologies and developmental issues through non-invasive methods.

For example, pulse oximetry, a non invasive technique that measures the amount of oxygenated and deoxygenated hemoglobin in blood by mean of infrared light, has been widely tested for early screening of congenital heart defects in asymptomatic newborn babies [1,2,3,4,5]. Recently, in a review by Thangaratinam et al. [5], authors compared the overall sensitivity of this method and false-positive ratio against other screening techniques, including prenatal ultrasounds and routine physical exams [6,7]. [what are the results?] One of the techniques in which researchers’ interest increased exponentially during the last sixty years is the empirical analysis of infant cry [8,9]. Acoustical properties of infant cry have been associated with different developmental pathologies, including Autism Spectrum Disorders (ASD), Sudden Infant Death Syndrome (SIDS), hearing impairments and unilateral cleft lip and palate (UCLP) [10,11,12].

1.1. Properties of Infant cry

Cry sound utterances are produced by the larynx during the expiratory phase of respiration. Pressure differences of air streams flowing through the larynx cause vocal folds to open and close rapidly, from about 250 to about 550 times per second in healthy infants [13,14,15,16,17,18,19]. This ratio of vibration is defined as the fundamental frequency (F0). Position of the vocal folds is modulated by central nervous system (CNS), and therefore activity of the vocal folds can be used to estimate an infant’s developmental status. Moreover, the lower vocal tract produces different sound characteristics, including the loudness of the expiratory phase. The upper vocal folds concord in the production of higher frequencies, resonants of the fundamental frequency [20,21]. During the first two years of life, an infant’s body evolves. The vocal tract shapes during this period, and therefore acoustical properties of cry vocalizations changes accordingly [24,25,26].

Research studies conducted on infants suffering from pathological conditions highlighted a positive shift in the spectrum of cry frequency properties, as compared to those of healthy infants. For example, investigation of infants at high risk of developing ASD disorders showed that the fundamental frequency of their cry vocalization can be higher than 700 Hz [27]. Analogous, F0 collected from vocalization of infants suffering from colic were significantly higher than those collected from healthy infants [28].

1.2. Cry analysis

In a typical cry experiment, audio recordings are collected by inducing infants to cry using a specific paradigm, or trigger (e.g. pain caused heel prick test [29]). Collected samples are then preprocessed to increase the signal to noise ratio.
During the 1960s, when systematic analysis of infant cry began, researchers relied on visual inspection of spectrograms. With the advent of more powerful computing devices, techniques and algorithms employed in cry analysis became more sophisticated, producing more accurate and useful results.

Because of the similarities between infant vocalization and adult voice, cry researchers adopted software designed for speech analysis. One of the software most widely used within the field is Praat, a free software developed by Paul Boersma and David Weenink, specifically designed for acoustic analysis of adult voice. In the last 18 years, Praat has been used in 41.3% (N=36) of the articles published within the field during this period (N=87), detailed information about the software in use is provided. Despite being a robust tool for speech analysis, Praat’s default parameters are not suitable for accurate analysis of cry samples. In this article we discuss the role of Praat in cry analysis, highlighting the reasons for which standard settings are not suitable and provide suggestions on how to apply it successfully on cry samples.

2. PRAAT

Praat features a graphical user interface that fits the needs of different researchers, from phoneticians to musicians and biologists involved in the acoustic analysis of animal vocalizations. Written in C and C++, Praat provides tools for analysis of signals’ pitch ($F_0$) and formants in audio signals. Not only that, Praat comes with a picture tool which produces high-quality graphics ready to be used in manuscripts and dissertations.

The software uses a general purpose scripting language that can be used to automatize the analysis of multiple files, allowing for fast processing of large amount of auditory samples. Praat implements an auto-correlation algorithm for pitch analysis. According to Boersma, the applied algorithms is not only more accurate than other frequency-based pitch detection procedures, but is also less dependent on the length of selected window and more resistant to rapid shifts and external noise.

2.1. Praat settings

In this work, settings have been verified on Praat version 6.0.43 (8 September 2018), running on a Linux machine (Linux Mint 19 Tara x86_64, Kernel: 4.15.0-42-generic).

2.1.1. Pitch

Default pitch settings point the algorithm to search for $F_0$ in the frequency range that goes from 75Hz to 500Hz. As introduced above, healthy infant cry’s fundamental frequency usually lays between 250 and 550 Hz, with the latter higher in sick infants. With those settings, there are at least two possible situations in which Praat cannot identify the real fundamental frequency value:

- $F_0$ is above the upper cutoff: In this situation, Praat will identify a wrong value (lower) for the fundamental, or provide no pitch information within a window.
- $F_0$ lays between the cutoff values but a strong noise with a frequency between 75 and 250Hz is present: In the situation where a strong periodical noise is recorded within the signal, such as the presence of a split-system air-conditioner within the recording environment, it is possible that the software identifies this lower frequency as the real fundamental, especially when this noise is about half of the real fundamental frequency.

2.1.2. Formants

Standard formant settings are used to obtain up to 5 formants with a frequency lower than 5500Hz. The GUI returns $n - 1$ formants’ frequency values, where $n$ is the number of formants indicated within the settings.

3. ANALYSIS OF COMPUTER GENERATED SIGNALS

To better illustrate pitch and formant extraction errors, we tested Praat with standard and cry suitable settings on a set of computer generated signals with a specific $F_0$, to which white noise was added. Formants (N=5), with a frequency of about $F_0*(n+1)$ and decreasing amplitudes have been added to the generated signals. For half of the files, noise at a specific frequency band, close to $F_0/2$, was added. Four different signals of 5s length have been generated. Audio files and the source code written in Python which were used to generate those signals are available online. Used frequency values for $F_0$, formants and, where added, $F_0/2$, are reported in Table 1 (Real). To verify the validity of generated files, a visual inspection of the spectrum was conducted. Frequency peaks are shown in parenthesis in Table 1.

Using Praat we extracted value of Pitch and Formants at $t = 2.5s$, using both Praat’s standard (Praat S) and cry-optimized (Praat O) settings:

- Pitch:
  - Pitch range (Hz) = 250.0 - 800.0 Hz
- Formants:
  - Maximum formants (Hz) = 4500.0 Hz

Fundamental frequencies and formant have also been verified by visual inspection of the spectrogram using Audacity version 2.2.1 and the following settings:

- Algorithm: Spectrum
- Window size: 1024
- Function: Hanning window
- Axis: Logarithmic frequency

Pitch and formants frequency obtained using the two set of settings, and their Mean Absolute Percentage Error (MAPE) are reported in Table 1.

4. DISCUSSION

As described above and demonstrated by analysis on simple computer generated samples, Praat’s default settings are not suitable for the analysis of infant cry. In example A.wav, $F_0$ is located between the pitch cutoff values and no periodic noise was added. We can observe, that parameter optimization led to a general improvement of formant estimation, with the MAPE drastically reduced. Similarly, in example B.wav, where $F_0$ was still between pitch cutoff values and periodic noise was above the lower cut-off with standard settings but not parameters optimized, the latter configuration
granted a better recognition of the fundamental as well as of the formant.
In example C.wav, F0 was higher than the upper cut-off for the pitch of Praat’s standard settings. Here, pitch recognition identified the wrong peak as the signal pitch. This situation did not occur when parameters were optimized and the higher cut-off was increased. This is especially important when working with pathological infants or where the risk of developmental pathology is high, and therefore acoustic properties of cry are expected to differ from those of healthy infants.

Finally, as shown with file D.wav, when the presence of periodic noise was at about half of the fundamental frequency (with a high fundamental frequency), it led the software to a recognition error even with optimized parameters. This did not happen when the spectrum was visually inspected, since it was clear that the amplitude of F0/2 was lower than the amplitude of the peak of F0, as visible in Figure 1.

Parameter tuning sharpens extracted features, but because of the properties of cry, researchers still have to pay special attention to obtained values, as well as to the quality of collected data.

Generally, we can expect Praat with standard settings to perform poorly when employed in infant cry studies, because of the complexity of the signal itself and of the presence of external noise. In the next section, we will show the performances of Praat on real cry sample, using both the standard and optimized settings.

5. ANALYSIS OF REAL CRY SIGNALS

In order to provide a demonstration of Praat’s performance on real cry samples, we analyzed infant utterances from a public dataset [36]. More specifically, we assessed the first three utterances from the file “BabyCrying.wav”, therefore named here as “Utterance1”, “Utterance2” and “Utterance3”. F0 and formants have been first obtained by visual inspection with Audacity, using the same configurations used to obtain the spectrum of computer generated signals. Because of the properties of cry, reported value are the mean values of a whole utterance. Frequencies’ peaks are reported in Table 2. Then, each utterance have been analyzed in Praat, using both the default settings (Praat S.) and our suggested settings (Praat O.). For each pair of file and settings, we estimated the Mean Absolute Percentage Error using as actual value the peak obtained manually in Audacity by visual inspection of the spectrum. Pitch and formants frequency values and MEAP per file and settings are reported in Table 2.

6. DISCUSSION

As shows in Table 2, the difference in the estimated MEAP of investigates samples follows what have been shown for computer generated signals in Table 1. Similarly to the previous examples, the higher the formant number, the higher the difference between the peak detected in Praat or by visual inspection.

With an average reduction in the estimated MEAP of 18.4%, a fast optimization of pitch and formant detection parameters demonstrated to be helpful in increasing the accuracy of estimated features. As demonstrated by our examples, differences in the used settings can result in a large variance in estimated frequency values. Because of that, we expect researchers involved in cry studies to tune the software properly and to report used settings in final manuscripts. Unfortunately, this is not the case: only in 12 out of 36 studies in which the researchers used Praat, details about the used settings were provided [8, 9].

7. CONCLUSIONS

In this work, we demonstrated the different level of performance that Praat, an open source software designed for speech analysis, can achieve when used with infant cry samples when the parameters are or aren’t tuned. In the first part of this work, we generated different acoustic signals with features similar to those of real cry samples. Generated files have been analyzed first by visual inspection, then using Praat standard settings and finally by fine tuning the algorithms’ parameters. The performance of the software has been evaluated using the Mean Absolute Percentage Error (MAPE). In the second part of this work, we applied the same procedure to a set of real cry utterances. Our results show that Praat standard settings are not suitable for the analysis of cry signal, and therefore the software should not be employed in cry studies without tuning. Researchers have to carefully examine collected data, to ensure that no external sources of periodic noises are recorded within the signals. Furthermore, because of the high inter-individual variability of cry properties, it may be advisable to tune pitch and formant extraction settings according to the investigated participants and their health statuses. We advise researchers of the field to test Praat’s parameters with more complex and extreme cry sounds so as to identify the extent to which the software can be correctly integrated in cry studies.

8. REFERENCES


Table 1: Real and Praat’s estimated values for generated acoustic signal properties. For Praat estimation, standard (Praat S.) and optimized (Praat O.) settings were used and values were computed at t=2.5s. In parenthesis are values obtained by visual inspection of the spectrum generated with Audacity version 2.2.1. For each pair of file and settings, the Mean Absolute Percentage Error (MAPE) has been calculated.

<table>
<thead>
<tr>
<th></th>
<th>A.wav</th>
<th>B.wav</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F₀ 440.0 (444) 448.2 448.0</td>
<td>F₀ 440.0 (444) 223.0 447.3</td>
</tr>
<tr>
<td></td>
<td>F₁ 867.0 (882) 881.6 667.7</td>
<td>F₁ 873.0 (882) 864.7 777.7</td>
</tr>
<tr>
<td></td>
<td>F₂ 1339.0 (1384) 1704.9 1367.4</td>
<td>F₂ 1338.0 (1384) 1688.4 1470.5</td>
</tr>
<tr>
<td></td>
<td>F₃ 1752.0 (1768) 2655.9 1945.4</td>
<td>F₃ 1755.0 (1768) 2573.1 2078.3</td>
</tr>
<tr>
<td></td>
<td>F₄ 2196.0 (2208) 4444.7 2719.0</td>
<td>F₄ 2190.0 (2208) 3916.1 2753.1</td>
</tr>
<tr>
<td></td>
<td>F₅ 2641.0 (2756) 2621.0 213.0 (216)</td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>36.9% 12.3%</td>
<td>40.4% 13.3%</td>
</tr>
<tr>
<td></td>
<td>C.wav</td>
<td>D.wav</td>
</tr>
<tr>
<td></td>
<td>F₀ 575.0 (581) 293.4 588.7</td>
<td>F₀ 575.0 (581) 293.6 293.5</td>
</tr>
<tr>
<td></td>
<td>F₁ 1166.0 (1131) 932.4 712.1</td>
<td>F₁ 1169.0 (1190) 1030.9 688.4</td>
</tr>
<tr>
<td></td>
<td>F₂ 1714.0 (1769) 1176.2 1411.0</td>
<td>F₂ 1707.0 (1769) 1898.7 1397.6</td>
</tr>
<tr>
<td></td>
<td>F₃ 2305.0 (2321) 2617.7 2110.4</td>
<td>F₃ 2292.0 (2321) 2765.6 2105.0</td>
</tr>
<tr>
<td></td>
<td>F₄ 2889.0 (2942) 3578.0 2880.4</td>
<td>F₄ 2884.0 (2942) 3695.7 2890.0</td>
</tr>
<tr>
<td></td>
<td>F₅ 3455.0 (3676) 3433.0 (3676)</td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>27.6% 13.5%</td>
<td>24.1% 23.3%</td>
</tr>
</tbody>
</table>


### Table 2: Real and Praat’s estimated values for generated acoustic signal properties.

For Praat estimation, standard (Praat S.) and optimized (Praat O.) settings were used and values where computed at t=2.5s. In parenthesis are values obtained by visual inspection of the spectrum generated with Audacity version 2.2.1. For each pair of file and settings, the Estimated Mean Absolute Percentage Error (Estim. MAPE) has been estimated using as Actual value the peak highlighted by Audacity trough visual inspection.

<table>
<thead>
<tr>
<th></th>
<th>Utterance1.wav</th>
<th>Utterance2.wav</th>
<th>Utterance3.wav</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_0$</td>
<td>531</td>
<td>380.5</td>
<td>477.3</td>
</tr>
<tr>
<td>$F_1$</td>
<td>904</td>
<td>1019.5</td>
<td>927.1</td>
</tr>
<tr>
<td>$F_2$</td>
<td>1343</td>
<td>2078.0</td>
<td>1637.8</td>
</tr>
<tr>
<td>$F_3$</td>
<td>1797</td>
<td>2832.4</td>
<td>2691.8</td>
</tr>
<tr>
<td>$F_4$</td>
<td>2271</td>
<td>3685.5</td>
<td>3288.1</td>
</tr>
</tbody>
</table>

Estim. MAPE: 43.2% 25.8% 30.0% 20.4% 55.1% 27.0% 30.0% 20.4% 55.1% 27.0%