Newborn pain cry analysis based on pitch frequency tracking

Delphine Bard[a], Per Runefors[b] and Einar Arnbjörnsson[c]

[a]Lund University, Division of Engineering Acoustics, John Ericsson väg 1, 221 00 Lund, Sweden
[b]Department of Paediatrics, Växjö Hospital, 351 85 Växjö, Sweden
[c]Department of Paediatric Surgery, University Hospital Lund, 221 00 Lund, Sweden
delphine.bard@acoustics.lth.se

The aim of the newborn pain cry analysis is to test the hypothesis that cry can be used as a tool to detect signs of nocioceptive pain. Previous studies applying signal processing techniques to analyze the sound of these cries have been done. The subject of this paper is to adapt and improve the original method with the help of new signal processing methods. The pitch frequency is extracted from the waveform of the recorded babies' cries using time domain methods. The fluctuations of this parameter are analyzed in terms of jitter. In particular, a sliding buffer approach is presented, as well as an improvement of the Average Mean Difference Function (AMDF). Comparison between original and news results has been done.

1   Introduction

In the early 90's, Lüdge et. Al. [1,2] have published an analysis method of Central Nervous System (CNS) disturbances of newborn children, based on computer analysis of recorded baby cries. The method uses signal intervals of 8ms, on which an autocorrelation function is performed. From the maximum peak of this function, the fundamental period and frequency are determined. Thereby, we obtain a curve of the fundamental frequency as a function of the time. The fine variability of this curve is extracted and used to determine the fundamental frequency jitter. A normalized jitter index is then calculated to differentiate ill from healthy babies. An update of this methodology is considered in order to get finer results and with respect to the computation power that is available commonly nowadays.

2   Audio material

A large number of baby cries has been recorded on a tape cassette some years ago by P. Runefors and E. Arnbjrönsson [5] and we could get a copy of it. The records to be found on this tape have been digitized using the sound card of a computer, at a sampling frequency of 48 kHz and with a resolution of 16 bits. The signal is stored on the computer as a lossless WAV PCM file. In order to simulate the behavior of original material and software, a copy of this record has also been realized by down sampling the signal at a frequency of 8 kHz and the resolution has been artificially reduced to 12 bits.

3   Initial Method

This section describes the methods used by Lüdge et. Al. [1,2] and the results we can get from those methods with some newborn child recordings. The recordings have been realized using an audio tape recorder connected to a microphone. A magnetic audio tape in cassette recorder/player ordinarily has typically a clean frequency response ranging from 30-40Hz to about 12-15kHz.

4   Sampling

The analog signal is converted in a digital signal prior to the analysis. The sampling has been achieved at a frequency of 8kHz and with a resolution of 12 bit. Due to hardware limitation at this time, and especially the amount of RAM memory (512kB), only 2 seconds of recording could be analyzed at a time. This signal of 2s recording is then broken down into 250 sequential sections of 8 ms each.

Each section therefore contains 64 digitized measurement points. One sampling period of a baby cry is represented on the figure 1, with a sampling frequency of 8 kHz and a resolution of 12 bits. We can recognize that this signal is periodic, as one pattern is repeated almost identically several times. In addition, we can also notice the existence of higher order sub harmonics that are present in this sound, although those latter are hard to recognize.
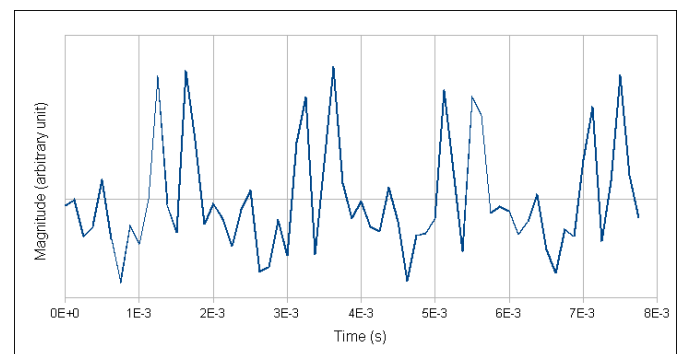

Fig.1: Waveform of the content of one buffer (8ms), when using a sampling frequency of 8 kHz.

5   Pitch Detection

For each signal section stored in the memory of the computer, the *Average Magnitude Difference Function* (AMDF) is calculated. This means that the resolution of the pitch frequency curve that we obtain is 8ms. For a recorded signal of 2s we therefore get 250 pitch frequency values. The AMD function f_{AMD} is defined as follows [1]:

$$f_{AMD}(p) = \frac{1}{M}\sum_{v=q}^{q+M-1}|x(v) - x(v+p)| \qquad (1)$$

where p is the time shift, q is the starting point, x is the digitalized signal section and M is the number of values in this section. If the signal x is periodic with a period of p_0, then we would have $x(p_0) = x(v+p_0)$. Therefore, and as for the autocorrelation function, the AMDF function presents a minimum at p_0 if the signal x is periodic with a period of p_0. However, the AMDF function requires less calculation time and power, since it does not need any multiplications. Also, the AMDF function presents the advantage of narrower peaks, making it easier to determine the period.

The figure 2 represents the result of the AMDF function as a function of the time. Note that the time scales extends only up to the half of the buffer length, since we do a summation of two time-shifted terms. By definition, the function returns 0 at the origin. The fundamental period is defined as the position of the absolute minimum of the AMDF function. The fundamental frequency is therefore simply the inverse of this minimum. We look for the minimum of the function over the considered time interval. In this precise case, the minimum takes place for t=2 ms, which corresponds to a pitch frequency of 500 Hz. We can now recognize 4 local minima taking place between the origin and the absolute minimum. They correspond to 5th order harmonics already observed on the figure 1.
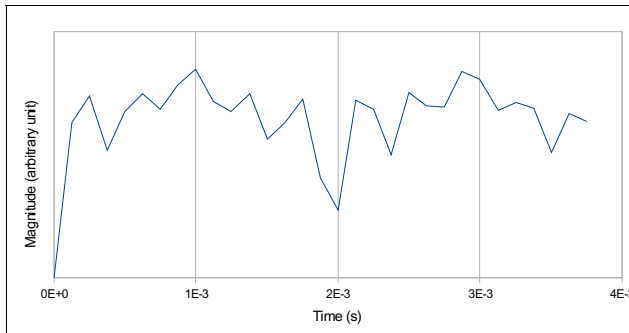

Fig.2**:** AMDF function calculated for the sound sample represented on the figure 1.

6   New Method

This second part describes some improvements to the initial methods and how the modifications can impact the result. The raw recording is the same as the material used for the early analysis techniques, namely a magnetic audio cassette with its 3.8 mm tape.

6.1   Sampling

In this section, we use the same infant cry record but with different settings. As stated in 2, the content is the same as in 3, except a higher resolution of 16 bits and a sampling frequency of 48 kHz. The figure 3 represents the same buffer as the figure 1, but with a sampling frequency 6 times higher, and an improved resolution. The waveform is much better defined, since we have 6 times more samples for the same buffer length (8 ms).
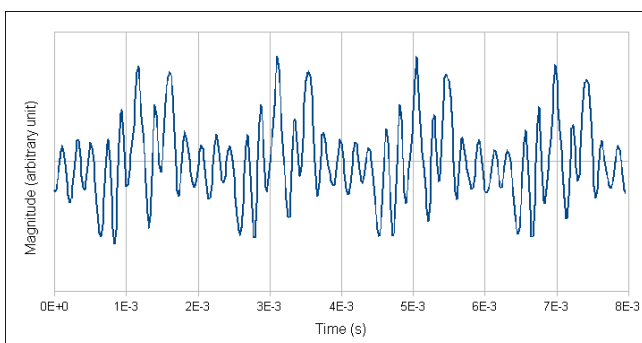

Fig.3: Waveform of the content of one buffer (8ms), when using a sampling frequency of 48 kHz.

6.2   Pitch detection

The main benefit of using a higher sampling frequency is that we can get a much better defined result of the AMDF function. The figure 4 represents the result of the AMDF function over the length of a buffer, corresponding to the waveform of the figure 3.
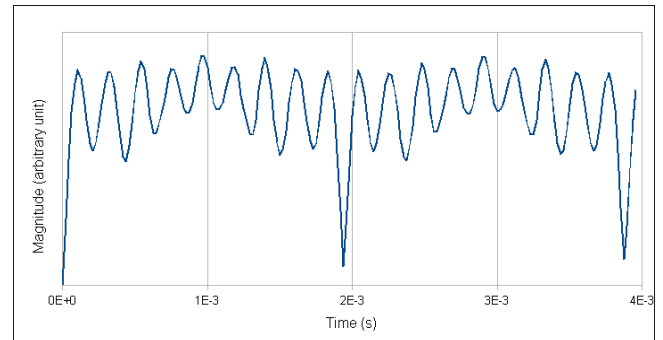

Fig.4: AMDF function calculated for the sound sample represented on figure3.

We can now identify the absolute minimum with much more precision. In this case, the position of the minimum is at 1.94 ms, which corresponds to a pitch frequency of 515 Hz. Considering the local minima, we can also clearly identify that the signal contains the 9th harmonic of this pitch frequency, thanks to the positions of the local minima.

We also try to get a better precision of the peak position by using linear interpolation with the help of neighboring sample points. Finding a minimum of the function is equivalent to finding the position at which its derivate is equal to 0. The derivate of the AMDF function can be defined by:

$$f'_{AMD}(p) = \frac{df_{AMD}(p)}{dp} = \frac{f_{AMD}(p+1) - f_{AMD}(p-1)}{(p+1) - (p-1)}$$
$$= \frac{f_{AMD}(p+1) - f_{AMD}(p-1)}{2}. \quad (2)$$

The conditions for having a local minimum at a given position p_m of this curve is that $f'_{AMD}(p) \leq 0, f'_{AMD}(p+1) > 00$ and $p \leq p_m < p+1$. If those conditions are fulfilled, the position of the curve minimum can be interpolated from the values of the function at the points p and $p+1$.

$$p_m = p + \frac{f_{AMD}(p+1) - f_{AMD}(p-1)}{(f_{AMD}(p+2) - f_{AMD}(p)) - (f_{AMD}(p+1) - f_{AMD}(p-1))}$$

(3)

6.3   Sliding buffer

We have seen in 3.1 that the initial method was limited to a relatively low number of pitch frequency values per second. This limitation is inherent to the relatively poor calculation power of the computers and acquisition systems at the date where the reference articles were written, as compared to the means available nowadays. No doubt that those latter will seem to be ridiculously low as well in ten or even five years. One way to circumvent this limitation is to use sliding buffers. When we do a classical

sampling, we fill in a buffer; we process it and release it for the next series of acquisition. The principle of the sliding buffers is to fill in and to store in memory several buffers, instead of one. The number of buffers can be two or three, or even more, depending how we want to use them. In our case, we will use three distinct buffers. Let the buffers be defined by their length M. After having filled in and stored three buffers, we calculate the first pitch frequency value by proceeding to the summation and using the first two buffers. We need two buffers for this because of the definition of the AMDF itself (1), which takes into account a total of 2M points. For the second pitch value, we shift the summation indexes by one single sample. This basically means that we will ignore the first sample of the first buffer, but in order to keep the total number of members in the summation equal to M, we have to take into account the first sample of the third buffer. For each of the following points, we precede the same manner, by shifting the indexes by one sample at a time. For the calculation of the Mth pitch value, we will use 2M samples, starting from the last sample of the first buffer to the last sample of the third buffer. After that, the first sample will be of no use anymore, and we can read in a new set of M points and put those samples in the first buffer. We will then calculate the next M pitch values the same manner as described before, but instead of using buffer 1, then 2, then 3, we use buffer 2, then 3, then 1. This makes sense, since buffer 1 will contain the most recently introduced samples. Logically, the next set of M points will use buffer 3, then 1, then 2, since we will have filled buffer 2 at last. After those consecutive cycles, we naturally start over with the first scheme. To get a graphical representation (5), we can consider that the three buffers are chained as a ring. Therefore, buffer 1 logically takes his place after buffer 3.
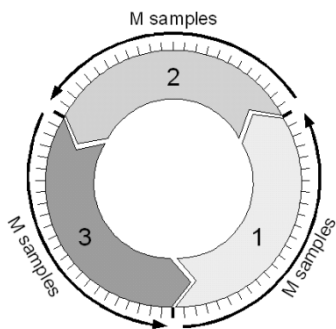

Fig.5: Three buffers connected together to form a ring.

$$AMDF(p) = \frac{1}{M} \left[\sum_{v=q}^{q+M-1} |x_{b_1}(v) - x(v+p)| \right. $$
$$\left. + \sum_{v=q}^{q+M+1} |x(v) - x(v+p)| \right]$$

(4)

where b_1, b_2 and b_3 indicate which buffer is used. By using the sliding buffer approach, we obtain one pitch frequency value for each signal sample. This way, we get rid of the limitation of the low number of pitch frequency values and have a much finer time granularity.

6.4 Weighted sliding buffer method

The method can be further improved by multiplying the time-domain signal for each buffer by a weighting function. This function shows a maximum in the middle of its range and decreases at both ends of the range. This allows for a reduction of the border effects when the time-domain signal is cut into smaller parts. Among the popular weighting functions, one can mention the Hamming, Hanning or Blackman functions, or simply a triangular function. The Hanning window has been chosen for this study.

7 Cry analyses

The figure 7 shows a typical sonogram for a series of subsequent baby cries and figure 8 represents the corresponding AMDF result. The sonogram allows us to clearly distinguish the variations of the cry intensity while the AMDF plot shows the variations in terms of pitch frequency. The pitch frequency is not calculated when the sound level is too low, because its value would not make any sense. Therefore, we can isolate three cries from that graphic, that can be analyzed in terms of length and minimum, maximum and average frequency.
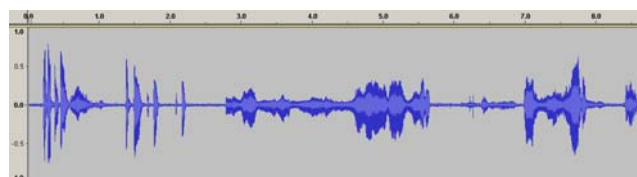

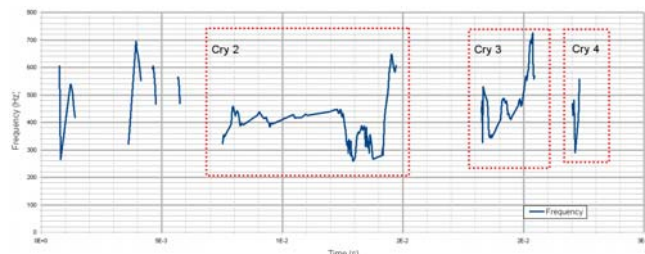Fig.6: Sonogram for a short sequence of cries.

Fig.7: AMDF plot for the same sequence of cries.

We observed that the new method gives slightly different average frequency values, especially when the pitch frequency varies a lot during one same cry. This can be explained by the fact that the new method gives more accurate results, and presents a finer pitch frequency granularity.

8 Conclusion

This study presents a solution to improve an existing baby cry analysis method. Not only modern equipment allows having faster and more accurate results by using the same original recording, but also it is possible to extend the scope of the analysis by using adequate signal processing methods.

References

[1] W. Lüdge, P. Gips, " Microcomputer-aided studies of cry jitter uttered by newborn children based upon high-resolution analysis of fundamental frequencies ", *Computer Methods and Programs in Biomedicine*, 28, p. 151-156, Elsevier, 1989.

[2] W. Lüdge, H. Rothänger, "Early diagnosis of CNS disturbance from computer analysis of infant cries: A new method of fundamental frequency jitter computation with high resolution in frequency and time", *Early Child Development and Care*, 65, p. 83-90, Gordon and Breach Science Publishers S.A., 1990.

[3] H. Rothgänger, K. Michelsson, W. Lüdge, E.-L. Grauel, "Comparing examination of cry and speech signals with the method of high-resolution fundamental frequency analysis", *Proceedings of the Third Congress of the International Clinical Phonetics and Linguistics Association*, 9-11 August 1993, Helsinki, Publications of the Department of Phonetics, University of Helsinki 39, p. 159-166.

[4] K. Michelsson, O. Michelsson, "Phonation in the newborn, infant cry", Children's Hospital, University of Helsinki (Finland), *International Journal of Pediatric Otorhinolaryngology*, 49 Suppl. 1, p. 297-301, 1999.

[5] P. Runefors, E. Arnbjörnsson, "A sound spectrogram analysis of children's crying after painful stimuli during the first year of life", Department of Paediatrics, Växjö Hospital and Department of Paediatric Surgery, University Hospital Lund, (Sweden), *Folia Phoniatrica et Logopaedica* 2005, 57, pp. 90-95.