The melody of crying

György Várallyay Jr.*

Budapest University of Technology and Economics, Department of Control Engineering and Information Technology, Laboratory of Medical Informatics, Hungary

Received 26 April 2007; accepted 10 July 2007
Available online 27 August 2007

KEYWORDS
Acoustics; Fundamental frequency; Hearing impairment; Infant cry; Melody analysis

Summary As the speech of a normal hearing and a deaf person are different, author expects differences between the crying sound of normal hearing and hard-of-hearing infants as well. In this study the author determined by computerized algorithms the melody of 2762 crying sounds from 316 infants, and compared the results between infants with hearing disorders and normal hearing. The analysis of the crying sounds is aimed to work out a new, cheaper hearing screening method, which would give a new potential to the early detection of hearing disorders. All the applied steps were developed by automatic, computer-executed methods providing reproducible, objective results in contradistinction to some previous studies, which had applied manual methods and reached subjective results.

Several possible ways for digital signal processing of the infant cry are discussed. A novel melody shape classification system was created to obtain a more precise distribution of the melodies by their shapes. The system determined 77 different categories, where the first 20 categories covered the 95% of the melodies. The applied methods were created and tested in a huge number of melodies.

1. Introduction
1.1. Hearing screening today

Nowadays numerous Universal Newborn Hearing Screening (UNHS) programs are running in several countries around the world with an increasing success and efficiency [1]. By these programs the average date of detecting hearing disorders (and starting intervention) has decreased to the first few months of life (it was typically a few years). Hard-of-hearing infants, therefore, have better chance for an adequate mental, physical, physiological, social, and language development.

The great majority of the UNHS programs apply objective audiometry (as various OAE or BERA techniques), and use a multiple-stage strategy to increase the efficiency.

At the Newborn Hearing Screening (NHS) 2006 Conference (Cernobbio, Italy, June 1–3), in the keynote lecture from Bolajoko O. Olusanya and in several other reports as well the following recommendations and ideas were presented [2]:

* Correspondence address: Magyar tudósok krt. 2, BME IIT IB.311, H-1117, Budapest, Hungary. Tel.: +36 1 463 4027; fax: +36 1 463 2204.
E-mail address: varallyay@iit.bme.hu.

doi:10.1016/j.ijporl.2007.07.005
developing countries many a time cannot put up the money for starting a UNHS program (costs of equipments) or for carrying on the program (costs of explorations and servicing) and for more acceptable UNHS programs cheaper hearing screening methods should be developed.

Although the existing strategies of objective audiometry have a high reliability, the author’s research team finds it essential to try to work out a new, cheaper screening method, which would give a new potential to the early detection of hearing disorders. This research team is aimed to create a new hearing screening method which tests the sound of the infant cry.

Existing screening methods test some parts of the whole auditory system, e.g. OAE checks if the hearing system functions or not between the outer ear and the inner ear. During sound production the whole auditory system is involved to control the quality of the produced sound, in this way the acoustic analysis of the produced sound (as crying) might be used to check the function of the whole auditory system.

1.2. The information of sounds

After birth, the first obvious sign of life is crying. This is the first mean of communication; experienced mothers (as well as nurses) can translate the crying sound and understand the needs of their own babies. There are typical, common attributes of crying as well as individual ones: all of the crying sounds seem to be similar, but a mother can distinguish the crying sound of her own baby from the sounds from other infants.

During crying, the volume, the pitch and the tone color is changing. The changing of the pitch is called melody. Based on the experiments so far the author declares that the melody curve of crying, except sudden pitch shifts, is continuous.

In human life, the melody of speaking is essential to express the mood, intention, etc. During infancy, the most typical sound is crying. The melody of crying also gives information about the physical, physiological state of the infant: it is characteristic for typical states as pain, hunger, discomfort, boredom, etc. [2—12].

The state of health can modify the produced sound as well. People have a typical voice if they have a cold, if they suffer from pharyngo-laryngitis, or even if they have impaired hearing. A normal hearing person can control his produced sound, while a deaf person cannot. Thus it can be determined from the sound of speaking if the patient has problems with the hearing or not. By this conception, different produced sound is expected between normal hearing and hearing impaired infants, whose produced sound is crying.

In 1982 Hirschberg and Szende collected over 100 sound phenomena from infants suffering form several diseases [13]. They reported about extremely high-pitched melodies (around 1000—2000 Hz) of crying at hard-of-hearings contrary to the healthy infants’ mean frequency range around 400—500 Hz. In 1995 Cacace et al. tested several parameters of the crying of normal hearing pre-term and healthy normal-term infants [14]. They found connections between age, weight and some acoustic features of the crying. In a study of Schönweiler et al. the auditory feedback was tested [15]. With a digital audio tape (DAT) recorder a real-time amplified feedback of the crying was driven to a headphone on the infant. When this feedback was suddenly delayed, infants with hearing disorders kept crying while the healthy ones got confused. In 1999 Möller and Schönweiler analyzed infant cries for the early detection of hearing impairment [16]. They applied self-organizing topological feature maps to compare the Mel frequency cepstral coefficients between healthy and hearing impaired infants and obtained a classification accuracy of about 75%.

It can be declared that even now there is no method with high efficiency which can conclude on the state of hearing disorders from the crying sound. The author’s aim is to create algorithms which calculate parameters of crying and diagnose hard-of-hearing from them with a better efficiency than the above-mentioned methods had. After several studies dealing with other attributes of crying [17—21], the author turns to the analysis of the melody of crying with the same purpose.

2. Subjects and methods

2.1. Subjects

For this study data from 316 infants were collected. Most of the recordings were made in several hospitals in Hungary and furthers at home. There were 147 boys and 169 girls with a mean age at 189.82 days. As approximately 0.2—0.3% of the infants born with hearing loss [22], only 23 of the 316 infants had medium grade hearing loss or deafness (threshold of hearing under —40 dB). The grade of the hearing loss was diagnosed by objective audiometry (as OAE and BERA). The more important details about the infants involved are tabulated in the following table (Table 1).

It is worthy of note that the average age of the hearing impaired infants is higher than the normal
hearings, as the typical time of the detection of hearing loss during infancy is about 6–18 months after birth.

The typical duration of the recordings was 25–30 s in the hospitals and 60–80 s at home. The reason of crying was pain in most cases (evoked crying); there were only a few infants, at whom other reasons were found (spontaneous crying). In 2007 Branco et al. tested the pain cry of 111 newborns; their pain situation was the venepuncture [23]. They found that the smaller the newborn weight, the bigger the presence of higher fundamental frequency with tense strangled voice quality; the bigger the neonatal/infant pain scale (NIPS) score, the more frequent the cough.

For this study the cries were collected during examination of the eardrum: the doctor looked into the ears of the infant. The procedure was painful and the baby started to cry. The cries selected for analysis were, whenever possible, chosen from the start of the cry sequences. The infants were sitting on their mothers’ lip, the distance between the microphone and the mouth of the infant was 1–2 m.

All the sound recordings were made in quiet places in the hospitals or at homes, but not in special silence rooms. The first recordings were made in 2001, the recent ones in 2006. During these 5 years the used recording devices were: minidisk recorder (SONY MZ-R55), digital video camera (SONY DCR-TRV25), digital dictaphone (SONY ICD-P28) and PC sound card with several microphones (SONY ECM-M5907, AKG D555) attached. The melody is such a robust attribute of the crying, that is not impressionable by the type of the recording device.

The digitalization of the recorded crying sounds was applied at 44.1 kHz or 48 kHz (depending on the recording device), each sample was assigned to 16 bits. Finally, all the recorded sounds were saved onto PC as separate wave (.wav) files.

### 2.2. Database

In this study a MS Access 2000 database structure was applied, containing the following pieces of information: name of the infant (necessary for tracking); date of birth (necessary for tracking); identity number (ID), which refers to the name of the wave file of the recording; gender; date of the sound recording; recording device, type of the microphone applied; place of recording (necessary for tracking); length of the recording, further recording parameters; reason of crying; results of several hearing screening tests as OAE, BERA, etc.; comments.

### 2.3. Methods

The author used Matlab with self-developed functions for all the following steps. The first step of the melody analysis was the segmentation of each recording into crying segments. As every crying segment has its own melody, this step was very important, because the melodies were calculated from these resulted segments. Secondly, each segment was divided into short-time windows, which were the base for the fast Fourier transform (FFT) algorithm. In this step the spectrum was calculated by FFT for each window of each segment and the detection of the fundamental frequency ($F_0$) was performed in each spectrum. The detected consecutive $F_0$ values form the melody of the segment. At the third step of the analysis the obtained melodies were visualized and classified. As it will be shown the existing categories of melodies were not precise enough, in this way a novel method was developed to categorize the melodies of crying better. At the fourth step two groups (GH, group of hard-of-hearings and GC, control group) were defined and the melodies were compared between them.

#### 2.3.1. First step: segmentation of the crying

The whole recording contains not only crying parts, but other disturbing parts (as noises, pauses, coughs) as well. As only the crying sounds (often called segments) are important in this study, segmentation is needed to detect the crying sounds in the whole recording. On the top of Fig. 1 a short example is shown for a general crying signal. In this example there are two crying segments (i.e. useful parts, need to be recognized), which can be seen in the second line of the figure. Further parts of the whole recording are, e.g. coughs, pauses and respirations.

There are three main attributes of a crying segment which should be used at the automatic segmentation:
The amplitude (i.e. energy) of a crying segment is quite big. The duration of a crying segment is longer than a few tenths of seconds. The spectral structure is regular of a crying segment, because a harmonic signal contains only the fundamental frequency and its harmonics.

By applying these attributes as rules, an automatic crying segment detection algorithm was developed. From the 316 infants 2762 segments were detected, which means: one recording of an infant contained 8—9 segments on an average. As there were much less hearing impaired infants, 202 segments were found in their group, while 2560 segments were detected at the group of infants with normal hearing.

2.3.2. Second step: windowing, FFT and fundamental frequency detection

The fundamental frequency ($F_0$) is the smallest useful frequency component of the spectrum (see below in Fig. 2). The changing of the time-dependent fundamental frequency results the melody. To follow the fundamental frequency varying with time, the author divided the segment into short-time windows (see in Fig. 1C). During a window, the changing of the fundamental frequency is considered negligibly small. By this way the melody is determined from the detected fundamental frequency values. The smaller the window length is, the more calculation has to be performed but more points will form the melody and vice versa.

The spectrum of a window is calculated by the fast Fourier transform (FFT). As the infant cry is normally a harmonic signal, it has a regular structure: only the fundamental frequency and its harmonics are present (Fig. 2).

Fundamental frequency detection algorithms were applied to determine the $F_0$ value of each spectrum. The majority of the spectrums obtained were simple like Fig. 2, but there were many cases, when one or more of the following events happened:

- There was background noise in the recording, which resulted in extra peaks in the spectrum. In a better case, the background noise was deterministic and its effects could be filtered from the spectrum. In a worse case it was unknown at which frequencies to expect the noise components.
- During crying the infant had hoarse sound, which produced several extra peaks in the spectrum.

![Fig. 1](image1.png)

**Fig. 1** Explanation of why segmentation is needed. Only the crying segments have melody. The upper line (A) shows a whole crying signal: there are both more important parts (called crying segments) and less important ones. In the second line (B) only the detected crying segments are present. The third line (C) illustrates the windowing within each segment: they are divided continuously into short-time (around 50 ms) windows for the further steps of the analysis.

![Fig. 2](image2.png)

**Fig. 2** Spectrum of a window of a crying segment. As it is a harmonic signal, the structure of the spectrum is regular: only the fundamental frequency and its harmonics are present.
There were other special events during the sound production which could also affect the normal spectrum.

- The sound of the infant was clear, but the formant structure (i.e. the intensities of the harmonics) resulted in such a small \( F_0 \) component, that it looked like the smallest frequency component was the second harmonic.

In these cases either there were several misleading components in the spectrum, or the intensity of the fundamental frequency was too small to detect. The detection algorithm of the fundamental frequency had to be a precise and robust method. There were several methods available to detect the fundamental frequency, such as local maximum value detection between a predetermined frequency range [24]; the autocorrelation function [25]; cepstrum analysis [24,25]; the harmonic product spectrum (HPS) method [26]; and the smoothed spectrum method (SSM) [17,18].

For the detection of the fundamental frequency from infant cries HPS and SSM methods proved to be the best algorithms. HPS is a very robust algorithm which applies a multiplication of compressed spectrums, and results the product spectrum, where a significant peak at the fundamental frequency is presented. SSM is a highly precise algorithm to detect the most probable value of the fundamental frequency which is also based on the regular structure of the spectrum and applies signal processing and statistical processes as well. The precision of the result of SSM is much better than the resolution of the original spectrum. As it was mentioned above, there were some cases, when the spectrum contained extra, misleading peaks (which do not belong to the crying sound), that is why special fundamental detection methods were needed.

For this study the author used for windowing an approximately 40 ms long window size, resulting 56,660 windows from the 2762 crying segments. The spectrum was calculated and \( F_0 \) detection was performed for each window by the processes mentioned above.

The results of the \( F_0 \) detections show the typical interval of the fundamental frequency in case of crying. These results confirm several previous studies dealing with the fundamental frequency of the infant cries [19,27,28], moreover it gives a more particular view of the distribution of the fundamental frequency from 56,660 measurements. The distribution of the detected fundamental frequencies is shown in Fig. 3.

The fundamental frequency of the analyzed infant cries varies between 200 and 1000 Hz. The most typical values are between 300 and 600 Hz.
As frequency is not a linear quantity logarithmic scale is suggested. The five-line method is a special method to visualize the melody contour of the infant cry. It contains guidelines at five exact frequency values; the smallest line is at approx. 330 Hz, while the biggest one is at approx. 700 Hz. The method remains to the five lines on a music paper, but the frequency values are not exactly the same with the values presented on the music paper. FLM applies exact frequency limits in the visualization which provides a well-arranged view of the melody.

In Fig. 5 a short recording of crying is shown with examples of several ways to visualize the melody.

As it is shown in Fig. 5, the spectrogram gives more information than only the melody visualization: the whole harmonic structure is included. But this additional information affects a bad resolution of the fundamental frequency itself: neither the exact melody shape of the crying, nor the starting/ending frequencies of the melody can be seen clearly.

The simple 2D visualization at the third line of Fig. 5 is a more practical method to show the melody. But the “reading” of the figure is not perfect: e.g. it cannot be seen whether the current melody is low-pitched or high-pitched. The best visualization is provided by FLM, where the melodies are comparable. More examples for the obtained melodies are shown in the next figure (Fig. 6).

There are several kinds of melody shapes among the 2762 melodies. Schönweiler et al. classified the melody shapes of crying [15]. Their categories were: rising, falling, rising—falling, falling—rising, flat and glottal plosive. However, the elements of these categories are quite simple, they do not cover all of the melodies, e.g. in the previous figure (Fig. 6) ‘CR0265-09.wav’ could be treated as a rising—falling melody, but it has a more complex melody shape than simple rising—falling.

2.3.4. The idea of determining new categories
The author defined three fundamental units: falling (−1), flat (0) and rising (1). All the melodies can be treated as a sequence of some of these fundamental units. The idea of this method is to create categories according to the obtained sequences. For example a rising—falling crying can be put together from a rising unit (1) and a falling unit (−1), in this way the sequence code of this melody type is [1 to −1].

To automatize this algorithm the melodies were partitioned into various numbers of parts depending on the number of the local extremums; then all of these partitions were classified as a falling unit, a flat unit or a rising unit. By this way from the 2762 melodies the author determined 2762 sequences containing −1, 0 and 1.

Fig. 5 A short recording of crying and examples of several methods to visualize the melody contour. The figure on the top shows the time—amplitude of five crying segments. In the second line a spectrogram is shown: there are the varying fundamental frequency (i.e. the melody) and its harmonics. The last two figures present only the melody as a function of time, but second of them (FLM) contains additional information about the position of the melody: there are five guidelines on the last figure to make easier the reviewing of the figure.
2.3.5. Fourth step: comparison of melodies between hard-of-hearing and healthy infants
For the comparison two groups were defined:

- GH: the group of 175 crying segments from hearing impaired infants. The age of these infants was between 3 and 12 months.
- GC: control group, having the same number of elements and age distribution as it is in GH.

Only the pain evoked crying sounds were collected into both GH and GC. Three attributes of the melody were compared between GH and GC: the duration of the melody, the frequency values of the melody and the shape of the melody (using the new categories).

3. Results

3.1. Results of the new classification system
The new classification system partitions the melody into fundamental units, and the information about the shape of the melody is carried by the order of these units; this method enables to make a precise categorizing of the all the melody shapes. After assorting the obtained sequence codes the author found the following results. A total of 77 different sequence codes were found from the easier ones (having only one unit) to the more complex ones (having 6–8 units). Out of the 77 categories there were 20 which include the 95% of the 2762 melodies. The distribution of these "top 20" categories and the schemes of their shapes are shown in Table 2.

The most typical category of this classification system was the [1 to –1] category with 933 cases. It can be clearly seen that in the top 20 all the simplest shapes (including one or two units) are present; while the most complex shapes have no more than four units.

Typically the starting partition of the melodies of crying is a rising unit or a flat unit, while the typical ending partition is a falling unit, or a flat unit. It might have happened because before and after sound production the tightness of the vocal cords was in static conditions.

Seventy-five percent of the crying samples could be categorized into the categories from Schöneweiler et al., the remaining 25% had more complex melody shapes. It can be declared that there are more than 5 important categories of the melody shapes of the infant cry.

3.2. Comparing the melodies between healthy and hard-of-hearing infants
There are three main attributes a melody can be described by: the duration, the fundamental frequency and the melody shape. The durations were determined after the automatic segmentation of the recordings (first step). The mean value of the fundamental frequencies the author calculated after windowing, FFT and fundamental frequency detection (second step). Finally, by the new classification system the sequence codes of the melodies were obtained (third step).

The mean duration of the melodies was \[1.0211 \pm 0.5734\] s and the median was 0.9288 s. The mean duration was somewhat smaller in the hearing impaired group (GH), \[1.0460 \pm 0.5813\] s, than in the control group (GC), \[0.9961 \pm 0.5659\] s;
the difference was not statistically significant (d.f. = 349, \( F = 0.66, p = 0.4165 \)). The results are shown in Fig. 7.

The mean value of the fundamental frequency of the melodies was 417.13 ± 72.13 Hz and the median was 413.21 Hz. The mean duration was higher in the hearing impaired group (GH), 425.51 ± 78.10 Hz, than in the control group (GC), 408.74 ± 64.77 Hz, the difference was almost statistically significant (d.f. = 349, \( F = 4.78, p = 0.0294 \)). The obtained distributions are shown in Fig. 7.

The most typical melody type was the rising—falling, signed \([-1 1]\), at both of the groups (Fig. 7). Approximately 30% of the 175 melodies could be ranked into this category in GH and in GC as well. While it was shown before the second most common

---

**Table 2** Illustrating the 20 most common melody shapes

<table>
<thead>
<tr>
<th>Code</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1 -1]</td>
<td>933</td>
</tr>
<tr>
<td>[1]</td>
<td>458</td>
</tr>
<tr>
<td>[-1]</td>
<td>383</td>
</tr>
<tr>
<td>[0]</td>
<td>176</td>
</tr>
<tr>
<td>[0 -1]</td>
<td>136</td>
</tr>
<tr>
<td>[1 0]</td>
<td>87</td>
</tr>
<tr>
<td>[-1 0 -1]</td>
<td>66</td>
</tr>
<tr>
<td>[-1 0 -1]</td>
<td>57</td>
</tr>
<tr>
<td>[-1 -1 1 -1]</td>
<td>53</td>
</tr>
<tr>
<td>[-1 1]</td>
<td>47</td>
</tr>
<tr>
<td>[1 -1 0 -1]</td>
<td>45</td>
</tr>
<tr>
<td>[1 0 -1]</td>
<td>41</td>
</tr>
<tr>
<td>[-1 1]</td>
<td>41</td>
</tr>
<tr>
<td>[1 0 0 -1]</td>
<td>40</td>
</tr>
<tr>
<td>[-1 0 -1]</td>
<td>26</td>
</tr>
<tr>
<td>[0 1 -1]</td>
<td>23</td>
</tr>
<tr>
<td>[0]</td>
<td>13</td>
</tr>
<tr>
<td>[1 0 1 -1]</td>
<td>9</td>
</tr>
<tr>
<td>[1 0]</td>
<td>9</td>
</tr>
<tr>
<td>[1 -1 1 0]</td>
<td>9</td>
</tr>
</tbody>
</table>

**Fig. 7** Histograms of the distributions of the tested parameters between the two groups (GH: light grey, GC: dark grey).

No significant differences were found between the two groups by duration. The distribution of the hearing impaired group has a lower center frequency than the other group has. The classification of the melody shapes produced quite a same distribution in the two groups.
The melody of crying was the rising melody. On the contrary both of the groups had a smaller ratio materially in this category. At both GH and GC there was generally the same distribution of the shape classes, no significant differences were found between the group of hearing impaired infants and the group of healthy infants.

4. Conclusions

The infant cry is an important acoustic signal, which cannot only be treated as a mean of communication, or a sign of mood, but it can be used to diagnose several diseases as well. A few decades before this idea was accomplished by some medical experts manually. Nowadays plenty of the classical analogue diagnostic methods are reproduced into digital ones to help and improve the diagnostics with their objective decisions and high efficiencies.

This study was aimed to compare the melody of crying between hard-of-hearing infants and healthy ones. As the acoustic feedback is damaged or missing in case of hearing disorders, the author expect significant differences between the sounds of crying of the two groups.

In this study 2762 melodies of crying were analyzed. From these melodies 56,660 crying windows were created, and the distribution of the fundamental frequencies of the infant cry was shown. The author suggested the harmonic product spectrum method and the smoothed spectrum method as the best fundamental detection algorithms in case of crying sound.

The author suggested a novel way to visualize the melody contour of crying. The five-line method applies five guidelines to place reference frequency values on the image.

As the obtained melody shapes were quite various, a new melody shape classification system was created. The system is based on the segmentation of the melody into elementary units (rising, flat or falling), and the categories are defined by the order of these units. From the classification of the melodies the author showed that 30% of the melody shapes were simple rising—falling; the more elementary units are present in a melody, the less number of cases that kind of melody occurs. This classification system can be used as a more precise system than previous studies had and by its objective, automatized decisions the results of previous studies dealing with the melody shapes of crying can be checked and refined [21].

Three main parameters can describe a melody: the duration, the fundamental frequency and the melody shape. The author determined a hearing impaired and a control group (with the same number of melodies and same age distribution) and compared these three parameters between them. No significant differences were found between the two groups by these tests.

In this study the common distribution of these parameters was not tested. The author is to continue the investigations in this topic by checking new attributes of crying and extending the scope of the tested diseases.

Acknowledgements

This research has been supported by Hungarian National Scientific Research Foundation (OTKA T0422990, T029830) and Medical Research Council (ETT 089/2003).

This study is a part of a biomedical project dealing with early diagnostics, started in 2001 in Hungary. Besides the author further members of the team are Professor Zoltán Benyó and Professor András Illényi from the Budapest University of Technology and Economics; Zsolt Farkas and Gábor Katona chief doctors from the Heim Pál Hospital for Sick Children, Budapest. In this study the author discussed his own research in this project.

Thanks for all the hospitals involved in the data collection: Heim Pál Hospital for Sick Children (Budapest), Szent István Hospital (Budapest), Schöpf-Merei Hospital (Budapest), and Borsod County Hospital (Miskolc). The author thanks Professor Jenő Hirschberg, Zita Makói, and Rainer Schönweiler for their consultations.

References


Available online at www.sciencedirect.com

ScienceDirect