Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis

Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis

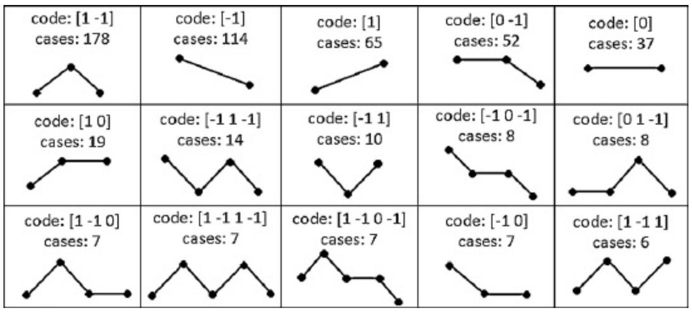

Fig. 5. Main categories found with the FLM.

and the categories are defined by the order of these values. All the melodies can be described as a combination of these elemental forms. For instance a melody of rising–falling type, is a combination of the rising form (1) and falling (−1), in this way, this type of melodies where classified as belonging to the category (1, − 1). At the end, 39 categories were established, 20 of them include 93% of the 580 analyzed melodies. The distribution of these categories and the schemas of their forms are shown in Fig. 5.

3. Proposed method

3.1. Automatic cry units detection

The cry units automatic detection is a relevant stage in the qualitative analysis, due to that qualitative features identification is performed over the cry units detected. To illustrate the way infant cry units are used, in [15] the cry units average duration is obtained, as well as the melody shape and the fundamental frequency average of the cries. In all cry recordings there are undesirable sounds, like environmental sounds and inspiratory cries, which do not provide any useful information to the analysis but they are usually present between cry units.

Inspiratory cries are produced when the infant inhales air while crying, which creates an audible sound. For qualitative features analysis these inspirations are not needed, reason why they have to be removed. There are also a variety surrounding environments and noisy devices around the crying infant during the recording process producing unwanted noises, which are noticeable also between cry units. These are sound to be removed too, as well as silence segments, to create the cry waves containing only crying sounds. The intensity and the type of the infant cry are other points to be taken in count. In the recordings we can find high-pitched or low-pitched, nasal, veiled, reedy, woody, etc. type of cries. And there are variations of the intensity due to the reduction or increase in the intensity of the cry that the infant can make during the same recording.

The process where the sound cries are detected and separated from undesirable sounds is called cry units detection, and previously was performed manually as described in [2,16]. Some software oriented to process speech has been applied to the cry units detection with no good results, because in several ways infant cry is different from speech.

Our Automatic Infant Cry Detection system was implemented in MATLAB. The first step was to identify the significant cry segments of each recording, eliminate unwanted noises and silence segments. To accomplish this task, we applied a threshold in each recording, this threshold was applied based on results of our experiments and it was proposed in [17] too, this threshold is repr
following equation:

$$\text{Energy threshold} = \frac{E_n(R)}{4}$$

where E_n is the short-time energy function and R i
recording.

This threshold is applied to each window of the an
ing, in this work, the width of the window was
the energy of the analyzed window is greater than
proposed threshold, this window is considered a cr

In cry units of very short duration is not possi
qualitative features, like melodic shape or others of
physicians. For this reason, the next step was to elim
of duration of less than 200 ms. With this second
able to eliminate inspiratory sounds which have
than 200 ms.

The steps to follow in the proposed method are s
Fig. 6(a) shows a cry signal, in Fig. 6(b) there is a cry
through the energy threshold proposed, and in Fig. 6(
cry units wave after eliminating the segments of le
is depicted.

After detecting the cry units and separating t
recordings, several other relevant attributes are obt
them are start and finish of each cry unit, duratio
number of cry units in the sample, which are obtaine
These attributes are useful for the expert physician
components in subsequent analysis.

3.2. Extraction of the fundamental frequency

Once we have detected the cry units, we extract
tal frequency values of each cry unit, for this we
algorithm proposed by Boersma in [20], the parame
the following:

- *Time step*: 0.05 s (frame duration).
- *Pitch floor*: 75 (standard value), candidates below
  are not recruited.
- *Pitch ceiling*: 1000, candidates above this frequen

The algorithm performs an acoustic periodicity d
basis of an accurate autocorrelation method. This m
accurate, noise-resistant, and robust, than methods
strum or combs, or the original correlation metho
why other methods were developed, was the failu
that to estimate a signal's short-term autocorrelati
a windowed signal the autocorrelation function of

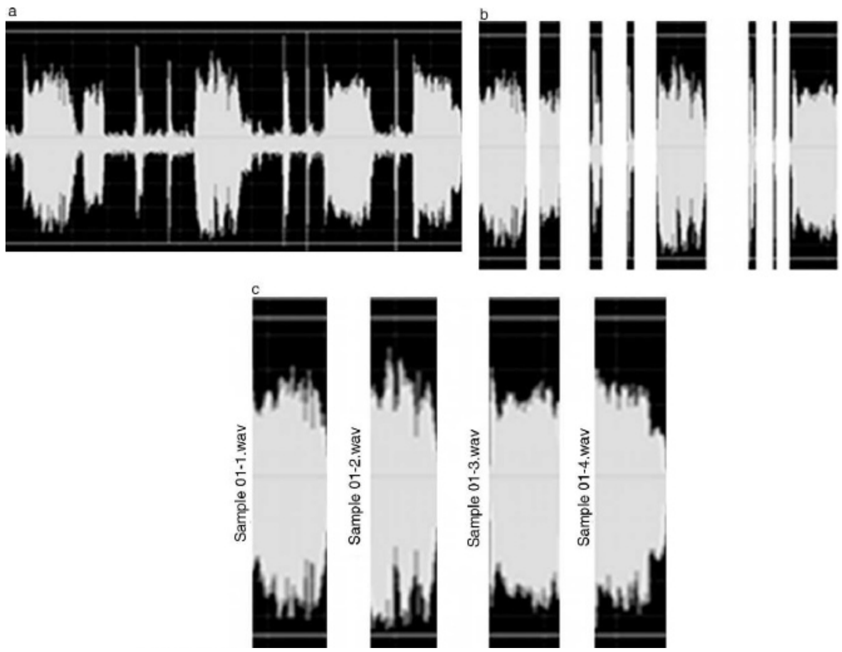

Fig. 6. Main steps in the automatic cry units detection: (a) recorded signal; (b) cry units resulting from the application of the short-time energy func
cry units obtained eliminating sound segments of less than 200ms.

signalshould be divided by the autocorrelation function of the window:

$$r_x(\tau) = \frac{r_{xw}(\tau)}{r_w(\tau)} \quad (4)$$

3.3. Identifying the melodic shape

Our proposed melodic shape identification method was implemented following the one presented in [12]. It is called the *dodecagram* method and it was implemented in MATLAB.

The fundamental frequency of each cry unit is fixed in the dodecagram. The value of the point P_i, is the value of the fundamental frequency of the first window, it can be noticed in Fig. 7 that in the second window the signal passes to the $P_i + 40$ row, in the third window the signal passes to the $P_i + 120$ row, in the next window the signal keeps in the $P_i + 120$ row, finally in the last windows the signal passes to the $P_i + 40$, $P_i - 40$ and $P_i - 120$ rows.

The next step is to encode the unit cry by the application of the following rules:

- 1 if the fundamental frequency passes to an upper row.
- 0 if the fundamental frequency stays in the same row.
- -1 if the fundamental frequency passes to a lower row.

In top of Fig. 7 the code corresponding to the cry unit in the dodecagram is shown. In the code, a number 1 corresponds to a rising of the fundamental frequency, the 0 corresponds to no meaningful changes and -1 corresponds to a falling in the fundamental frequency. With the obtained code the melodic
qualitative features can be obtained by following

- If in four consecutive windows the fundamental
are 0 a concentration of noise is confirmed.
- If all the digits of the code are 0s then the melc
- If there is no noise concentration and the mel
flat, then the 0s are eliminated and only the 1s
- The vectors are reduced, as shown in Fig. 8. S
digits are reduced to only one digit.

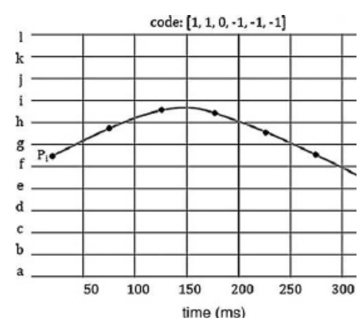

Fig. 7. Determination of a melodic shape by the use of the dc

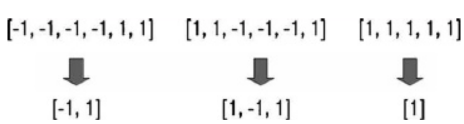

Fig. 8. Vector reduction.

- Finally the melodic shape that corresponds to the reduced vector is assigned in the following way:
  - ◊ 1: rising melody.
  - ◊ -1: falling melody.
  - ◊ 1, -1: rising–falling melody.
  - ◊ -1, 1: falling–rising melody.
  - ◊ The vectors with size higher than 2, are considered without melodic shape.
- In order to identify the shifts, from the spectrogram, we measure the differences of the fundamental frequencies along the signal. If the difference passes 100 Hz and is less than 600 Hz, within less than 10 ms, it is considered a shift (there can be more than one in a cry unit).

the Puebla City. From them 123 correspond to norm
belong to the pathological class. The same pediatrici
of each recorded sample, do the class labeling. The
recorded with ICD-67 Sony digital recorders, and tl
8000Hz [21]. The crying corpus from Cuba, compose
from the same number of babies, was collected at
Hospital of Santiago de Cuba with the help of the Spe
Group from the Universidad de Oriente (UO). For
they used a cassette recorder AKAI PM-R55 and
by the acquisition system PCVOX A/D [22]. The c
recorded under different conditions and different st

With our method we detected 182 unit cries from
12s long taken from the Mexican set. And we detect
from the Cuban set, with 13 recordings 12s long. Tl
compared versus the results obtained by manual
manual detection was performed by expert physicia
and 2 show in detail the obtained results in the re
ments.

4.2. Identification of qualitatives features

- In the same way, in order to identify the glides, and also from the spectrogram, we measure the differences of the fundamental frequencies along the signal. If the difference is equal or passes the 600 Hz within less than 10 ms, it is considered a glide (there can be more than one in a cry unit).

4. Experiments and results

4.1. Detection of cry units

The proposed cry units detection method was tested over two different sets of crying samples. The first set was taken from the database called Baby Chillanto property of the Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE, México), the other is a set of Cuban infant cries. The mexican infant cry corpus available is a set of 195 samples directly recorded from 112 babies by pediatricians from the Instituto Nacional de Rehabilitación-INR (Mexican Rehabilitation Institute) in Mexico City and the Instituto Nacional del Seguro Social-IMSS (Mexican Institute for Social Security) in

Our cry units identification method was tested
set, in which the labels were attached by expert p
the Instituto Nacional de Rehabilitación-INR (Mexic
tion Institute). These samples were taken from the
base.

From the results in Table 3 it can be noticed th
provides competitive results. Overall, we obtained 9
racy in qualitative features identification against th
made by human expert. For comparison purpose, t
labeled cry units was evaluated using the original
FLM proposed in [12].

In Table 4 we can see that the success percenta
50.27%. Our interpretation of those results is that t
of the low outcome is due to the fact that fundame
values from some cry units detected in our sample
below the values proposed for FLM, that is why mar
coded with 0s only.

Our method was able to correctly identify 238 c
tures, the results were compared versus the manual

Sample	Manual detection	Automatic detection	Accuracy	False positive
026.wav	10	10	100%	0
028.wav	8	8	100%	0
067.wav	12	12	100%	0
079.wav	10	10	100%	0
083.wav	9	9	100%	0
084.wav	5	5	100%	0
087.wav	10	9	80%	0
088.wav	10	10	100%	0
090.wav	6	6	100%	0
091.wav	14	14	100%	0
094.wav	5	5	100%	0
096.wav	12	12	100%	0
097.wav	13	13	100%	0
098.wav	5	5	100%	0
099.wav	13	13	100%	0
100.wav	13	13	100%	0
101.wav	7	7	100%	0
103.wav	8	8	100%	0
105.wav	10	9	80%	0
113.wav	7	9	71%	2
Total	187	187	96.55%	2

Table 2
Results of the manual and automatic infant cry units detection experiments for the Cuban set.

Sample	Manual detection	Automatic detection	Accuracy	False positive
C020812.wav	6	6	100%	0
C060812.wav	5	5	100%	0
C070812.wav	4	4	100%	0
C150812.wav	3	3	100%	0
C170812.wav	5	5	100%	0
C200812.wav	4	4	100%	0
C210812.wav	3	3	100%	0
C240812.wav	7	9	71%	2
C250812.wav	7	9	71%	2
C280812.wav	6	6	100%	0
C290812.wav	3	3	100%	0
C300812.wav	6	6	100%	0
C310812.wav	2	2	100%	0
Total	61	65	95.53%	4

Table 3
Results of the manual and automatic qualitative features identification using our proposed method on the Mexican set.

Qualitative feature	Manual identification	Automatic identification	Success percentage
Rising melody	36	35	97.22
Falling melody	36	32	88.89
Rising–falling melody	58	50	86.21
Falling–rising melody	18	16	88.89
Flat melody	34	29	85.29
Without melody shape	33	29	87.88
Shift	6	6	100.00
Glide	2	2	100.00
Noise concentration	15	12	80.00
Total			90.49

Table 4
Results of the manual and automatic qualitative features identification using the FLM method.

Qualitative feature	Manual identification	Automatic identification	Success percentage	Units out of the proposed scale
Rising melody	36	21	58.33	6
Falling melody	36	18	50.00	13
Rising–falling melody	58	34	58.62	10
Falling–rising melody	18	6	33.33	4
Flat melody	34	19	55.88	
Without melody shape	33	15	45.45	5
			50.27	

We like also to enlarge our databases, possibly w
tional infant cry corpus, and make both qualitativ
comparisons.

References

[1] J. Hirschberg, Dysphonia in infants, Int. J. Pediatr. Otc
(1999) S293–S296.
[2] K. Michelsson, O. Michelsson, Phonation in the newborn,
atr. Otorhinolaryngol. 49 (1) (1999) S297–S301.
[3] K. Wermke, K. Lind, Development of the vocal fundamen
taneous cries during the first 3 months, Int. J. Pediatr. Ot
(2002) 97–104.
[4] M. Robb, H. GIlbert, Vocal fundamental frequency cha
hunger cries: birth to 12 months, Int. J. Pediatr. Otorhinol
237–243.
[5] E. Obediente, Fonética y Fonología, chapter Elementos
vol. 45–68, 2007.
[6] O. Wasz-Hockert, J. Lind, V. Vuorenkoski, E. Partanen, El
su significación diagnóstica, Cientí fico-Médica, Barcelor
[7] Fundamentos, Teóricos y Prácticos del Análisis del Lla
Reyes Garcí a and Sergio D, Cano Ortí z, 2009.
[8] E. Arch, M. Mandujano, L. Garcí a, C. Martí nez, C. Reyes, V
llanto del niño hipoacustico y del niño normo-oyente, C
(2004) 271–276.
[9] L. Sasvári, P. Gegesi-Kiss, P. Popper, Z.S.Z. Makói, 1st cry of
and cesarean delivery, Acta Pediatr. Hung. 16 (1975) 15
[10] K. Michelsson, H.T. de Barra, O. Michelsson, Focus Nonv
Research, Finley R. Lewis, 2007.
[11] T. Szende, J. Hirschberg, Pathological cry, stridor and coug
Childhood 58 (1983).
[12] Z. Benyó, G.J. Várallyay, A. Illényi, Melody Analysis in the
2009, pp. 11–14.

5. Conclusions

It has been proven that infant cries carry a great load of useful information, mainly related to the physical and psychological state of the baby. Based on these grounds, qualitative feature analysis from infant cry is proving its potential as a powerful non invasive tool to help the emission of opportune early diagnostics.

In the qualitative analysis. The correct detection of cry units is of vital importance for the success of the further stages of the analysis. As it was shown, with the selected thresholds, our proposed method is able to detect cry units even under noisy recordings. These established thresholds allow the elimination of silence, noise and inspiratory sounds from the crying samples. In general, our innovative proposed method will facilitate the automatic identification of qualitative features in infant cry.

In the near future we want to identify a larger set of qualitative features, like vibratos and those with a grater diagnostic force as recommended by expert physicians. We are working on a rule

[13] G.J. Várallyay, Z. Benyó, A. Illényi, Z. Farkas, L. Kovács, A
infantcry:classicalandnewmethods,in: C.Manfredi(Ed.
of Vocal Emissions for Biomedical Applications, 2004.
[14] R. Schönweiler, S. Kaese, S. Möller, A. Rinscheid, M. Pto
and self-organizingmaps: newcomputertechniques in t
of theinfant cry, Int. J. Pediatr. Otorhinolaryngol. 38 (19
[15] K. Michelsson, K. Eklund, P. Leepänen, H. Lyytinen, Cry
healthy 1 to 7dayold infants, Int.J. Phoniatr. Speech The
(2002).
[16] K. Wermke, W. Mende, C. Manfredi, P. Bruscaglioni, Dev
infant's crymelody and formants, Med. Eng. Phys. 24 (7-
[17] I.Saratxaga, I. Luengo, E. Navas, I. Hernández, J. Sanchéz
pitch en condiciones adversas, in: IV Jornada en Tecnolo
sidad del País Vasco-Euskal Herriko Unibertsitatea, 200
[18] J. Green, J. Irwin, G. Gustafson, Acoustic cry analysis, in:
(Eds.), Crying as a Sing, a Symptom and a Signal, Clin
Medicine, vol. 152, 2000, pp. 137–156.
[19] K. Michelsson, Cry Characteristics in Sound Spectro
College-Hill, Houston, 1980, 85–105.
[20] P. Boersma, Accurate short-term analysis of the fundam
harmonics-to-noise ratio of a sampled sound, in: Proce
of Phonetic Science, 1993, pp. 97–110.
[21] E. Amaro, Caracterización Automática del Llanto de Beb
Modelos de Clasificación, Master's Thesis, INAOE,Tonant
2008.
[22] S.Cano,D. Escobedo,M. Torres, A. Castillo, Análisis prelin