A Cry-Based Babies Identification System

Ali Messaoud and Chakib Tadj

Laboratoire de Traitement de l'Information et des Signaux,
École de Technologie Supérieure,
1100, Rue Notre-Dame Ouest, Montréal, Canada
ali.messaoud.1@ens.etsmtl.ca, chakib.tadj@etsmtl.ca

Abstract. Human biological signals convey precious information about the physiological and neurological state of the body. Crying is a vocal signal through which babies communicate their needs to their parents who should then satisfy them properly. Most of the researches dealing with infant's cry intend mainly to establish a relationship between the acoustic properties of a cry and the state of the baby such as hunger, pain, illness and discomfort. In this work, we are interested in recognizing babies only by analyzing their cries through the use of an automatic analysis and recognition system using a real cry database.

Keywords: Infant cry, classification, neural network, acoustic features.

1 Introduction

The human body is a source of many signals related to different functions such as cardiological and nervous systems. The analysis of biological signals is important for medical diagnoses and also for the study of various phenomena observed in the human body.

Although a baby has no explicit way of communication, he can inform the parents through his cry about a need to be satisfied. Experienced parents can analyze this signal correctly and act appropriately whereas others remain confused about it. In many cases, this parental perception is altered by individual or contextual factors [1-3]. Another important aspect of the subject is the ability of the mother to distinguish her infant's cry from others within the first days of life [4]. In previous studies [5-7], researchers have discovered that the acoustic features of a baby's cry change according to the physiological state of the baby. These findings were at the origin of the efforts made to develop "intelligent" fully automatic systems capable of analyzing infant's cry and decoding its underlying significance [8-10].

In this study, the main goal is to design an automatic system for the identification of a baby from his cry in an attempt to emulate a mother who is capable of recognizing her own child just by his cry. A comparison between the natural and the artificial recognition systems could contribute to a better understanding of the perception mechanism of infant's cry.

We followed a procedure composed of two main steps. First, we performed signal processing tasks resulting in the extraction of specific acoustic features characterizing the cry. Secondly, an automatic classifier based on neural networks was used to assign

A. Elmoataz et al. (Eds.): ICISP 2010, LNCS 6134, pp. 192–199, 2010.

an input cry to one of known babies in the database. It was trained and tested by the mean of real cries recorded for the purpose of this study. Figure 1 shows the general structure of the designed system.

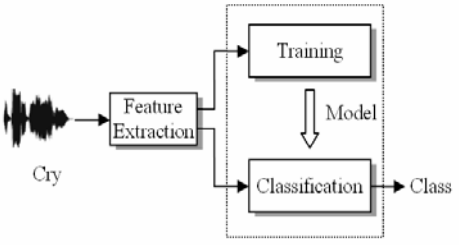

Fig. 1. Block diagram of cry-based recognition system

2 Methodology

2.1 Subjects

The cry database was constructed from the recording of 13 healthy babies aged less than 6 months. We dealt with 7 females and 6 males. These babies had no family support and therefore were part of a social program aiming to provide them special care and shelter during their first months of life before they get adopted. It is important to consider this social aspect in other analog studies especially if subjects without a family support and subjects with a normal family life are involved.

2.2 Acquisition Procedure

In this work, real infant's cries were used in the experiments. Therefore, it was necessary to maintain a uniform and well defined procedure during the acquisition of the cry sounds so as to guarantee a good quality of the final recordings. The procedure was based on subject-related, environment-related and material-related precautions.

The subject-related precautions consists in recording the babies in an "everyday" natural state like hunger or discomfort caused by wetness and avoid extreme cases such as undergoing a very painful stimulus. Hence, spontaneous cries were recorded mainly in a hunger context.

As for the environment-related precautions, a special care was given to the ambient conditions surrounding the baby being recorded. Noise was kept as low as possible by recording each baby separately and suspending the acquisition whenever the ambient noise was too high.

In order to achieve a good sound quality, we used the WS-310M Olympus digital voice recorder at a 44.1 KHz sample rate. During acquisition, the device was located such as to capture a maximum cry level while keeping the ambient noise very low (a high signal to noise ratio).

2.3 Segmentation

After acquisition, the digital cry sounds were transferred to a computer and stored in the WAV format. The raw cry recordings had a total duration of 158 min. They were treated by Adobe Audition software[1] to reduce the background noise picked up by the recorder. Then, we segmented these records manually into individual cry utterances discarding unusable ones. We obtained 1615 cry samples ready to be analyzed. In the next sections, the term "cry" denotes an individual cry utterance.

3 Acoustic Analysis

According to previous studies [8-10], a cry is characterized by a set of acoustic features which contain relevant information about the state of the baby and which can be perceived, analyzed and interpreted by the listener. Thus, the first step in this study is the extraction of these features.

3.1 Preprocessing

The inaccurate manual segmentation resulted in cries with imprecise start and end points (advanced start and delayed end). So, the "silent" portions at the edges of each cry were removed by considering energy less than 1 % of the maximum energy as "silence". As a result, all the cries in the database had synchronous start.

We have noticed in many other works [8-9] that the cry records were segmented on a fixed length basis (e.g. 1 second). This method leads to a loss of information due to the truncation of the cries longer than the chosen duration. Instead of this method, we adopted a variable-duration segmentation that resulted in cries with different durations.

The other preprocessing we implemented was the time scaling of all the cry waveforms to a unique length of 1 s using a phase vocoder technique.

Finally, we filtered the cry samples using a 4^{th} order low pass filter with a cutoff frequency of 3000 Hz.

3.2 Feature Extraction

The main goal of features extraction is the conversion of vocal waves into a set of values which represent the signal in a very compact way discarding irrelevant information. Various acoustic features such as pitch, intensity, jitter, etc. [3] have been proposed to characterize an infant cry.

Cepstral coefficients are the most widely used features for speech recognition. Cepstral analysis consists in calculating the short-term Fourier transform power spectrum, mapping this spectrum to a Mel scale using a filter bank and performing the Discrete Cosine Transform of the logarithm of the obtained spectrum [11]. The Mel Frequency Cepstral Coefficients (MFCCs) are the amplitudes of the resulting spectrum. The Mel scale used in this analysis is based on a model of the human auditory system.

[1] http://www.adobe.com/fr/products/audition/

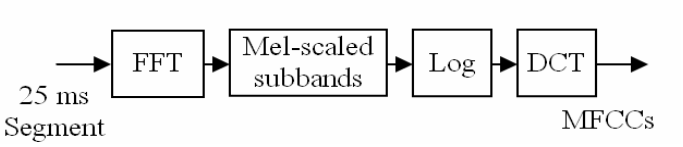

Fig. 2. MFCC extraction system

In order to perform such an analysis, each cry was divided into overlapping, Hamming windowed, short segments. A window duration of 25 ms and an overlapping amount of 20 % were adopted. From each window, 25 MFCCs were extracted. Figure 2 shows the block diagram of the MFCC extractor.

4 Classification

The recognition of a particular cry among a set of acoustically different cries was based on the acoustic characteristics (MFCCs) obtained by the feature extraction stage. These features were formatted as a vector representing the input of the system.

4.1 Neural Network

Among many neural network architectures, we adopted the probabilistic neural network (PNN). We made this choice because of the simplicity of the design of such a network and its suitability for classification problems. The PNN is a two layer network. In the first layer, the distances between the input vector and the training input vectors are computed. The second layer produces a vector containing the probabilities of belonging to the classes. A competitive transfer function takes theses probabilities as an input and produces the final classification result which corresponds to a value of 1 for the largest probability and 0 elsewhere. In our case, we used a network formed by 780 nodes in the first layer and 13 nodes in the second layer.

4.2 Classifier Input

The feature extraction stage produced for each cry a total number of 1225 MFCCs. Due to this large dimension of the input vector, we used a dimension-reducing technique which goal was to decrease the computational power. For each cry, the arithmetic average of the MFCCs was calculated along the time axis resulting in a vector formed by only 25 elements. In Fig. 3 we can see an example showing how the MFCC matrix of a cry is transformed into a simple 25-element vector (we used interpolation to smooth the variation of MFCC values). The simple visual analysis of the degree of similarity of this feature vector from one baby to another supports the fact of considering it as characteristic. Figure 4 shows the MFCC values for 4 babies (the mean of all the cries of each baby).

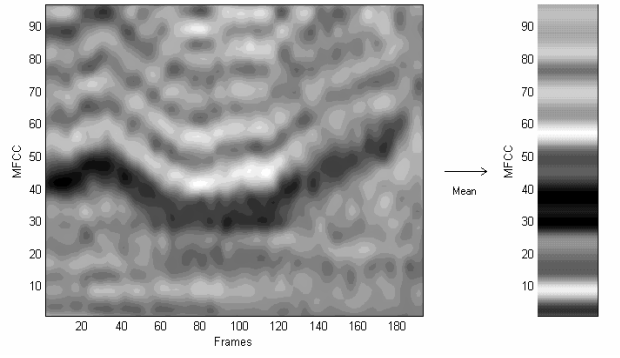

Fig. 3. Reduction of MFCC feature dimension

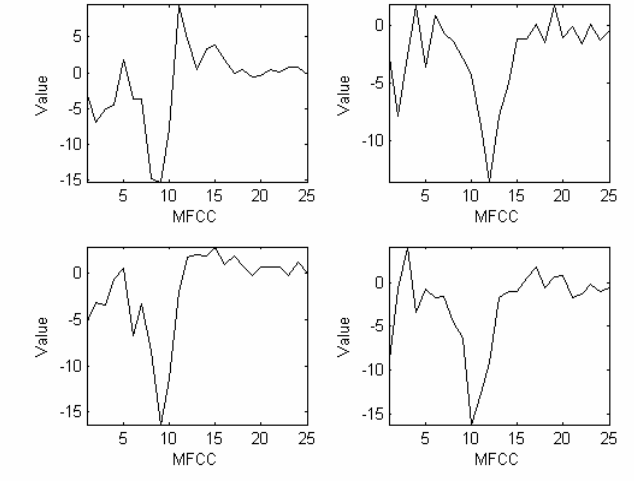

Fig. 4. Examples of MFCC curves of 4 babies

5 Results and Discussion

The designed system was implemented using the programming environment MATLAB. The experimental results presented in the next sections deal with the classification of 13 babies according to their characteristic cries modeled by the extracted MFCCs.

5.1 Preselection of Cry Samples

The segmentation stage resulted in 1615 cries not uniformly distributed on the 13 babies. The use of this raw distribution was likely to produce inaccurate classification results. So, before launching the classification process, we had to pass through a preselection step in order to obtain the same number of cries per class. The idea was to look for the class having the least number of cries and select randomly that number of cries from every class. We obtained 73 cries per class (a total number of 949 cries).

5.2 Cross Validation

In order to validate the designed classifier, we adopted a 7 cross validation technique that consisted in dividing the cry data set of each class into 7 subsets: 1 test set and 6 training sets (10 cries per subset for each class). Then, the classification was repeated 7 times by choosing each time one of the 7 subsets for testing and the remaining subsets for training.

5.3 Classification Performance

The performance of the classification is measured by the number of correctly recognized cries presented to the system. Table 1 gathers the performance specific to the 7 cross classifications. We can see from these results that the accuracy of the designed classifier is greater than 69.2 %.

Besides, in any classification problem, the classes to be predicted do not have the same specificity (the ability to be classified correctly). Therefore, we computed the accuracy of the classifier with respect to the 13 babies involved in the study. Table 2 presents these results which show different accuracy from one baby to another with a maximum performance of 90 % for the baby #12. The 40 % value of the first class indicates that the cries belonging to baby #1 were not easy to distinguish from other cries. A simple auditory inspection of the cries produced by baby #1 showed a relatively large variability of the perceived acoustic properties compared to other babies.

Table 1. Performance specific to the 7 classifications (%)

1	2	3	4	5	6	7
71.5	72.3	74.6	70	71.5	70.7	69.2

Table 2. Performance specific to the 13 classes (%)

1	2	3	4	5	6	7
40	68.5	65.7	61.4	84.2	70	70

8	9	10	11	12	13	
58.5	82.8	87.1	67.1	90	82.8	

The detailed classification performance results presented in the previous table can be used as a guideline in the improvement of the system performance. For instance, we would need to be concentrated on increasing the accuracy for baby #1 rather than baby #5 or baby #10 whose classification rates are already high. This improvement could be achieved for example by altering the content of the cry database of the concerned classes by performing the random preselection of cries described in section 5.1. The cry data set which gives the best results could be retained.

We present in table 3 the confusion matrix of the classifier which is the mean of the 7 confusion matrices. We can see in this table that the diagonal contains high values indicating a pretty good performance. The misclassification rate is represented by the rest of the matrix.

The overall performance of the classifier was calculated independently of the classes and the classification iterations by taking the mean of all the specific accuracy values. We obtained a global value of 71.4 % showing the high ability of the designed classifier to recognize individuals (babies) from their cry signals.

Table 3. Confusion matrix

	Predicted Class												
Actual Class	40.0	1.4	11.4	0	4.2	4.2	14.2	7.1	4.2	7.1	4.2	1.4	0
	2.8	68.5	1.4	0	0	8.5	4.2	1.4	4.2	8.5	0	0	0
	7.1	0	65.7	7.1	1.4	1.4	2.8	2.8	2.8	0	4.2	2.8	1.4
	0	0	2.8	61.4	20.0	0	1.4	0	0	0	12.8	0	1.4
	1.4	0	1.4	7.1	84.2	0	2.8	0	0	0	2.8	0	0
	2.8	2.8	1.4	1.4	0	70.0	5.7	2.8	11.4	1.4	0	0	0
	8.5	1.4	2.8	7.1	2.8	1.4	70.0	1.4	0	2.8	1.4	0	0
	2.8	0	1.4	0	0	2.8	4.2	58.5	11.4	12.8	2.8	2.8	0
	2.8	2.8	4.2	0	0	4.2	0	2.8	82.8	0	0	0	0
	1.4	0	0	0	0	0	2.8	2.8	2.8	87.1	2.8	0	0
	2.8	0	4.2	8.5	4.2	0	1.4	1.4	0	0	67.1	7.1	2.8
	0	0	0	0	0	0	0	0	0	0	8.5	90.0	1.4
	0	1.4	4.2	2.8	0	0	0	0	0	0	1.4	7.1	82.8

6 Conclusion

In the human body, biological signals convey rich information about various vital functions. A newborn, up to a certain age uses cry signals to communicate.

In this work, we were interested in using the information contained in the infant's cry to recognize a baby among 13 babies from his cry. Acoustic characteristics were extracted from a real cry database. We used Mel Frequency Cepstral Coefficients which were a good choice in this study. As for the classification, we chose a Probabilistic Neural Network known for its suitability for classification problems. The obtained results showed an overall performance of 71.4 %.

As a future work, we intend to enlarge the database and try other combinations of acoustic features to improve the accuracy.

Aknowlegements. We acknowledge the funding by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

1. LaGasse, L.L., Neal, A.R., Lester, B.M.: Assessment of infant cry: Acoustic cry analysis and parental perception. Mental Retardation and Developmental Disabilities Research Reviews 11, 83–93 (2005)
2. Wood, R.M., Gustafson, G.E.: Infant Crying and Adults' Anticipated Caregiving Responses: Acoustic and Contextual Influences. Child Development 72, 1287–1300 (2001)

3. Protopapas, A.: Perceptual differences in infant cries revealed by modifications of acoustic features. The Journal of the Acoustical Society of America 102, 3723–3734 (1997)
4. Linwood, A.: Crying and Fussing in an Infant. Gale Encyclopedia of Children's Health: Infancy through Adolescence. Thomson Gale (2006)
5. Fuller, B.F., Keefe, M.R., Curtin, M., Garvin, B.J.: Acoustic Analysis of Cries from "Normal" and "Irritable" Infants. West J. Nurs. Res. 16, 253 (1994)
6. Goberman, A.M., Robb, M.P.: Acoustic characteristics of crying in infantile laryngomalacia. Logopedics Phoniatrics Vocology 30, 79–84 (2005)
7. Green, J.A., Gustafson, G.E., McGhie, A.C.: Changes in Infants' Cries as a Function of Time in a Cry Bout. Child Development 69, 271–279 (1998)
8. Galaviz, O.F.R.: Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system. In: Gelbukh, A., de Albornoz, Á., Terashima-Marín, H. (eds.) MICAI 2005. LNCS (LNAI), vol. 3789, pp. 949–958. Springer, Heidelberg (2005)
9. Orozco, J., Garcia, C.A.R.: Detecting pathologies from infant cry applying scaled conjugate gradient neural networks. In: Proc. European Symposium on Artificial Neural Networks, pp. 349–354. d-side publi, Bruges-Belgium (2003)
10. Ortiz, S.D.C.: A radial basis function network oriented for infant cry classification. In: Sanfeliu, A., Martínez Trinidad, J.F., Carrasco Ochoa, J.A. (eds.) CIARP 2004. LNCS, vol. 3287, pp. 374–380. Springer, Heidelberg (2004)
11. Xu, M., Duan, L.-Y., Cai, J., Chia, L.-T., Xu, C., Tian, Q.: HMM-Based Audio Keyword Generation. pp. 566-574 (2005)