Wavelet-based automatic cry recognition system for detecting infants with hearing-loss from normal infants

Mahmoud Mansouri Jam, Hamed Sadjedi

Department of Electrical Engineering, Shahed University, Thehran, Iran
E-mail: mansourijam@shahed.ac.ir

Published in *The Journal of Engineering*; Received on 17th September 2013; Accepted on 20th September 2013

Abstract: Infant cry is a multimodal and dynamic behaviour that it contains a lot of information. Goal of this investigation is recognition of two groups of infants by new acoustic feature that has not used in infant cry classification. The cry of deaf infants and normal hearing infants is studied. 'Mel filter-bank discrete wavelet coefficients (MFDWCs)' have been extracted as feature vector. Infant cry classification is a pattern recognition problem such as 'automatic speech recognition', which in signal processing stage the authors performed some pre-processing included silence elimination, filtering, pre-emphasising and, segmentation. After applying the discrete wavelet transform on the Mel scaled log filter bank energies of a cry signal frames, MFDWCs feature vector was extracted. The feature vector, MFDWCs, of each cry sample has large length, so they used principle components analysis to reduce in feature space dimension, after training of neural network as classifier, they achieved to 93.2% correction rate in cry recognition of test data set. This result shows better efficiency in comparison with previous familiarised approaches.

1 Introduction

Infant cry, as a useful acoustic signal, is a multimodal and dynamic behaviour that carries a lot of information about infant, such as hunger, pain, sleepiness or boredom. Mothers and specialists in the area of child care can distinguish their baby from others according to their baby cry. Such entity vividly clarifies an infant physiological anatomy and psychological conditions. If any disorder occurs with the infant, the cry may differ from the normal cry. This issue is main idea in recognition systems working based on infant cry [1, 2]. The basic hypothesis is cry vocalisations of hearing-impaired infants differ from those of their counterparts with normal hearing because of the lack of auditory feedback. To phonation, brain sends some commands to speech organs and they produce a phone. In the normal hearing case, the organ of hearing gives an external feedback of phone to the brain. If this feedback is missed, the created voice may differ from the original [3]. Therefore, it is possible to distinguish hearing disorder from produced voice. The pioneers of studying on infant cry issue were Wasz–Hockert *et al.* in Scandinavia in the 1960s, from the date onwards; others have investigated this issue such as [2–6]. However, clinical centres have not accepted the results these approaches so far, as a reliable diagnosis tool. On the other hand, it is necessary to study more effective approaches and results to satisfy clinical centres. In this Letter, the cry of infants with hearing-loss is compared with the cry of normal infants. In fact, the final goal of this investigation was diagnosis of babies coming to the world with severe hearing-loss or deafness with non-invasive method (more than 4000 babies with hearing-loss per year, only in Iran). In previous Letters, some acoustic features have been studied; such as pitch and formants [2, 3], 'Mel frequency cepstral coefficients (MFCCs)' and 'linear prediction coefficients (LPCs)' [4, 6]. Our goal is the evaluation of other appropriate features in infant cry classification, except the prior used features, so we evaluated a new approach by a feature extracted based on wavelet analysis and establish an automatic cry recognition (ACR) system by using artificial neural network (ANN) as classifier. Finally, we achieved better results in comparison to the previous features.

2 ACR process

'The ACR' process is a pattern recognition problem as same as 'automatic speech recognition'. Input of the ACR system is the acoustic signal of infant cry, and at the output, the kind of cry is obtained [4]. Generally, the process of ACR is performed in two steps. We called the first step as signal processing, whereas pattern classification was called to second step. In the first step, the goal is feature extraction, hence the cry signal needs some pre-processing. There is not important information above 4 kHz in cry signal, so the sampling rate was selected 8 kHz, according to Niquist rate. The recorded signal may include undesired effect such as background noise, echo etc. In the frequency field, the lowest main component is not <250–300 Hz, thus a high-pass filter with 200 Hz cut-off frequency is a suitable solution for reducing of noise effects. After signals normalisation, proposed feature vector was extracted. In second step, a part of feature vectors was used for training of classifier. Some classifiers in ACR system have been used in recent Letters, such as 'hidden Markov model (HMM)', Bayesian classifier, ANN etc. that in this Letter we have used ANN classifier because of these reasons: high capacity of learning, simplicity, various training algorithms and good result in infant cry researches [4, 6, 7]. Later on, a set of unknown feature vectors, the test patterns, are compared with the knowledge that the computer has to measure the classification output efficiency.

3 Feature extraction

To introduce a proper feature in infant cry analysis, we have used 'discrete wavelet transform (DWT)' of cry signal. The main advantage of wavelet transform against the other linear transforms, for example, 'discrete Fourier transform (DFT)' or 'discrete cosine transform (DCT)', is the ability to represent the signal in both time and frequency domains. It has been shown that the DWT can approximate time-varying non-stationary signals in a better way than DFT [8]. There are many families of wavelet basic function, but for this pattern recognition application, the 'Daubechies (db)' wavelet was chosen [9]. It has the following properties that make it a very attractive choice: time invariance; if the time series is time shifted then its wavelet packet coefficients are only time shifted. Fast computation: db wavelets have fractal-like self-similarity properties that lead to fast wavelet transform techniques. Sharp filter transition bands: db wavelets have very sharp transition bands, which minimises edge effects between frequency bands.

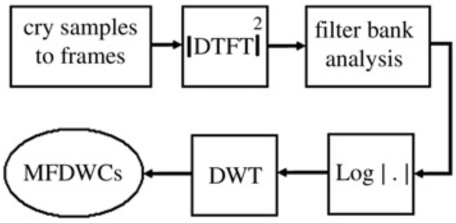

Fig. 1 *Extraction of MFDWCs*

By inspiration from human hearing where hearing ability is not equally sensible in linear frequency range, a filter bank with Mel frequency scale is used. As Fig. 1, we used 'Mel filter-bank discrete wavelet coefficients (MFDWCs) as the proposed features. The MFDWCs are obtained by applying the DWT to logarithm of energies in each filtered-bin of a cry spectrum. The difference between MFCCs and MFDWCs is the DCT of logarithm of energies in each filtered-bin of a cry spectrum is computed for MFCCs, but for MFDWCs the DWT is replaced to DCT.

4 Result of ACR implementation

For providing of infant cry database, we collected some signals from several clinics in Iran, and some signals were picked from the 'Baby Chillanto' database from 'Instituto Nacional de Astrofisica Optica y Electronica – CONACYT, Mexico'. We normalised all signals by 16 KHz sampling rate and each sample was quantised to 16 bit. Then, each signal is labelled by same infant property, that is, deaf and normal. We implemented our ACR system by cry signals of 11 hard hearing or deaf infants and 17 normal hearing infants without consideration of their gender. Length of signals was not equal, because of different situations in signal recording.

After silence elimination by energy threshold approach, we segmented each signal to segments with 1 s of length. We called each segment as 'cry sample', we obtained 326 cry samples for normal hearing infants and 326 cry samples for deaf infants. In the next step, we created 23 frames with 100 ms length from each cry sample after windowing by Hamming window with 60% of overlapping. These frames were passed from noise rejection filter and then pre-emphasise filter by frequency response $H_{(z)} = 1–0.95\ z^{-1}$. Then, the fast Fourier transform of each frame was computed, and then after passing from Mel filter bank with 14 sub-band, we took logarithm from energy of each sub-band of Mel passed frames that we called them as logarithm of filter bank energy (LFBE). Then, we obtained 16 DWT coefficients from LFBE, that this new coefficients known as MFDWC. We used db2 as wavelet mother function in this Letter. Each sample cry has 23 frames, so the final feature vector for each cry sample has 368 elements. The length of windows (100 ms), the number of filters in filter bank (14 filters) and the wavelet function have been optimised after several experiments.

The used classifier is 'multi-layer perceptron (MLP)' neural network, for training process, we need more cry sample by 368 elements in feature vector. On other side with size of this feature vector, computational cost is high. For overcoming to this problem, we used 'principle components analysis (PCA)'. PCA analysis reduced size of feature vector from 368 to 30 elements, so ANN can be trained in lower cost. Architecture of our used MLP consists of 30 input layers, 5 nodes in hidden layer and 2 nodes in output layer and scaled conjugate gradient method was used for network training. To obtain reliable results, we used permuted cross-validation method in training process. The 80% of data was dedicated to training phase and 20% of data for test phase, in separately. From data set of training phase, again, 80% of data was dedicated to training and 20% was dedicated to validation set. After ten-time of training process, we achieved to results as Table 1.

Table 1 Correction rate of ACR system

Infant class	Train data set, %	Test data set, %
deaf	94.7	91.2
normal	97.1	95.2
total	95.9	93.2

According to results in Table 1, we can observe the MFWDCs features are efficient features in classification of infant cry. In comparison with MFCCs and LPCs that had proper results in previous weeks, our approach has better results than results in [7, 4], respectively. We re-implemented their approaches and achieved 91.5 and 86.5% of correction rates, respectively, to MFCCs and LPCs.

5 Conclusion

The infant cry is a multimodal and dynamic behaviour that carries a lot of information. In this Letter, we could recognise deaf infants from normal hearing infants with extraction of MFDWCs. Our experiments preformed on 17 normal hearing infants and 11 deaf or hard hearing infants. We used MLP_ANN as classifier. By implementation of ACR system, we achieved relatively, high correct recognition rate (93.2%). We proposed new feature in cry classification that in prior works has not been used, results of this work showed the efficiency of MFDWCs, so we can use this approach in a non-invasive recognition system for early detection of deafness.

6 Acknowledgments

We like to thank Dr. Carlos A. Reyes-Garcia, Dr. Emilio Arch-Tirado and his INR-Mexico group, and Dr. Edgar M. Garcia-Tamayo for their dedication of the collection of the Infant cry database.

7 References

[1] Wasz-Hocket O., Lind J., Vuorenkoski V., Partanen T., Valanne E.: 'The infant cry, a spectrographic and auditory analysis' (Spastics International Medical Publications in Association with William Heinemann Medical Books Ltd., 1968)

[2] Varallyay G. Jr., Benyo Z., Illenyi A., Farkas Z., Kovacs L.: 'Acoustic analysis of the infant cry: classical and new methods'. Proc. 26th Annual Int. Conf. IEEE, EMBS, CA, USA, 2004

[3] Varallyay G. Jr.: 'Future prospects of the application of the infant cry in the medicine', *Period. Polytech. Ser. El. Eng.*, 2006, **50**, (1–2), pp. 47–62

[4] Reyes-Galaviz O.F., Reyes-Garcia C.A.: 'A system for the processing of infant cry to recognize pathologies in recently born babies with neural networks'. SPECOM'2004: Ninth Conf. Speech and Computer St. Petersburg, Russia, 20–22 September 2004

[5] Molaeezadeh S.F., Salarian M., Moradi M.H.: 'Type-2 fuzzy pattern matching for classifying hunger and pain cries of healthy full-term infants'. Proc. 16th Int. Symp. Artificial Intelligence and Signal Processing (AISP), 2012

[6] Reyes-Galaviz O.F., Cano-Ortiz S.D, Reyes-García C.A.: 'Evolutionary-neural system to classify infant's cry units for pathologies identification in recently born babies'. Proc. Special Session MICAI, 2008, pp. 330–335

[7] Orozco J., Reyes Garcia C.A.: 'Detecting pathologies from infant cry applying scaled conjugate gradient neural networks'. European Symp. Artificial Neural Networks, Belgium, 2003, pp. 349–354

[8] Tufekci Z., Gowdy J.N., Gurbuz S., Patterson E.: 'Applied mel frequency discrete wavelet coefficients and parallel model compensation for noise-robust speech recognition'. Speech Communication 48, Elesevier, 2006, pp. 1294–1307

[9] Daubechies I.: 'Ten lectures of wavelets' (SIAM, Philadelphia, 1992)