Infant Crying Cause Recognition using Conventional and Deep Learning based Approaches

Shivam Sharma
IIIT Sri City
630 Gnan Marg, Chittoor
Andhra Pradesh, India
shivam.sharma@iiits.in

P. Viswanath
IIIT Sri City
630 Gnan Marg, Chittoor
Andhra Pradesh, India
viswanath.p@iiits.in

Vinay Kumar Mittal
Ritwik Software Technologies Pvt. Ltd.
Road No 45, Jubilee Hills
Hyderabad, Telangana, India
drvinaykrmittal@gmail.com

Abstract

The significance of the meaningful information about the psycho-physiological state, contained within infant cries is well established, but the computational task of cry-cause recognition is yet to gain attention. In this work, features F_0 contour, sub-band spectral energy and MFCCs using the data-set ICSD2, collected especially for this study are examined for crying characterization. Recognition task is approached using k-nearest neighbour (k-NN), feed-forward (FFNN) and convolutional neural networks (CNN). A 55 dimensional feature-set is assessed for classification using nearest neighbour and feed-forward neural network. FFNN based classification outperforms not only the primary classifications evaluated using manual forward sequential feature selection, but also several State-of-the-Art results for similar tasks. Also examined is the performance of a convolutional neural network (CNN) towards learning suitable feature representations using MFCC and delta MFCC as inputs. Although with limited evaluation and output, insights towards effectively utilising taller filters and data-augmentation for a data-set with an imbalanced category-wise count are obtained. The results provide evidence in favour of MFCC based functionals along-with feed-forward neural network based cause recognition. The discussion provided in this work can possibly provide motivation towards tasks related to infant cry-cause recognition or acoustic feature representation learning using deeper networks.

1 Introduction

The advent of the research on paralinguistic speech analysis is traced back to the middle of the 20^{th} century, notably through the statements given by Crystal, that defined it as "vocal factors involved in paralanguage" (Crystal, 1974). Although, with no formal connection to the linguistics, the potential of the meaningful information contained within the acoustics of an infant cry, has already started convincing medical practitioners, parents, care-givers, etc., about its diagnostic importance.

Major work in Paralinguistic research, for instance academic challenges are being directed towards objectives like speech emotion recognition (Schuller et al., a,b). Although these efforts streamline state specific discoveries for paralinguistic, but as an outcome they lead to limited growth for understanding relevant techniques and useful resources for the applications that do not involve linguistic information within the acoustic signal, and hence the development remains largely application specific. It is this aspect that the work done in this paper attempts to address.

Tools like short-time spectrograms and cross-correlograms have been used towards majority of the infant cry acoustic analysis, leading to spectral and inter-segmental cross-correlation analysis in (Neustein, 2010; Petroni et al., 1994; Sharma et al., 2017). Attempts involving Fundamental frequency (F_0) contour by implementing Welch's method, autocorrelation, FFT, and modZFF, have been observed to be crucial for characterising excitation source within an infant cry in (Petroni et al., 1994; Cohen and Lavner, 2012; Sharma et al., 2017; Sharma and Mittal, 2017b; Yegnanarayana and Murty, 2009; Mittal, 2016b,a). ZFF along-with dominant frequency analysis is used to analyse shouted speech in (Mittal and Vuppala,

Proc. of ICON-2018, Patiala, India. December 2018 ©*2018 NLPAI*, pages 20–27

2016; Mittal and Yegnanarayana, 2013). Also, evaluation of base-line features like pitch, formants and MFCCs, along-with popular classification techniques like SVM and k-NN, including the neural network based classifiers like feed-forward and convolutional neural networks as well in (Galaviz and García, 2005; Reyes-Galaviz et al., 2008; Sahak et al., 2010; Zabidi et al., 2010; Cohen and Lavner, 2012; Lavner et al., 2016), have resulted in insightful observations.

Spectral information from higher order cumulants, along-with base-line features like MFCC, LPC and PLP, was observed to elucidate the non-linearity towards classifying Normal vs. Pathological cry sounds in (Chittora and Patil, 2015). Authors in (Orozco et al.) attempted to classify infant cries into Hunger or Pain, by individually evaluating linear prediction coefficients and signal intensity using a feed-forward neural network. A duration-thresholding based pre-processing step of cry sound segmentation along-with sequential forward floating feature selection approach, was taken up in (Chang et al., 2017) and compared with the results from (Abdulaziz and Ahmad, 2010) for evaluating spectro-temporal features for a dataset with 490 cry samples, towards cry-cause classification as Hunger, Lack of sleep or Pain.

A pressing concern in this field is the shortage of publicly available datasets of infant cries for categorical studies. Another direct challenge posed is about the disparateness of the categories being studied. The work has been done for a variety of causes ranging from pathologies like Asthma to disorders like Asphyxia, Ventricular septal defect (VSD), Upper respiratory tract infection (URTI), etc., in (Wahid et al., 2016; Chittora and Patil, 2015). This diversifies not only the utility of the characteristic feature-set, but also the understanding about the approaches suitable for a class specific study. All these challenges are resonated with the limitations imposed by the unavailability of infant cry dataset in public domain. It is this lack of understanding and common frameworks with respect to the resources and techniques, that the fundamental approach adopted in this work and the observations made thereof, towards infant cry acoustic analysis and cause recognition, is motivated from.

The rest of the paper is organised as follows. Signal processing methods used and features examined are stated in Section 2. Infant cry corpus collection and organization is discussed in Section 3. This is followed by the description of the experimental setup in Section 4. Acoustic analysis of the infant cries is described in Section 5. The observations related to the evaluation of crying cause classification are stated in Section 6. Section 7 provides a detailed account of the key results obtained. Finally, the paper is summarized and concluded in Section 8.

2 Signal processing methods used and Features examined

2.1 Signal processing methods used

1. *Short-time Fourier analysis:* Frequency domain processing of a signal, considered for short-time durations (Oppenheim et al., 1989).
2. *Autocorrelation:* Provides a measure of self-similarity over time (Haykin, 1989).
3. *Filter-bank spectral analysis:* Critical frequency band based spectral content analysis (Bourlard and Dupont, 1997) and (Lyons, 2012).
4. *Cepstral analysis:* Obtaining Cepstral coefficients as features representing speech sound production system characteristics in quefrency domain (Chang et al., 2017).

2.2 Features explored

1. F_0 *contour:* Functionals like Standard-deviation (dev_{F_0}) and Mean $(mean_{F_0})$ of F_0 contour are computed at the cry segment level.
2. *Sub-band spectral energy (SSE):* The ratios between the 4^{th} sub-band to 1^{st} and 2^{nd} sub-bands, i.e., $(\epsilon_{X_{4:1}})$ and $(\epsilon_{X_{4:2}})$, commonly denoted as SSE_r are computed.
3. *Mel frequency cepstral coefficients:* A total of 52 coefficients, including MFCCs $(MFCC)$, delta MFCCs $(\Delta MFCC)$, and their Standard-deviations, $(MFCC_{dev})$ and $(\Delta MFCC_{dev})$ respectively, with 13 coefficients each are computed.

3 Infant Cry Corpus (IIIT-S ICSD2)

An infant cry dataset IIIT-S ICSD2, having a total of 104 subjects (50 Male and 54 Female), age ranges for whom lie between 2 days and 6 years

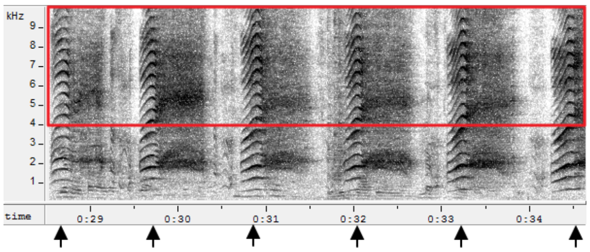

Figure 1: Spectrogram for Pain cry, depicting arched excitation contour (Marked by arrows) and consistently high spectral energy (Marked by rectangle box).

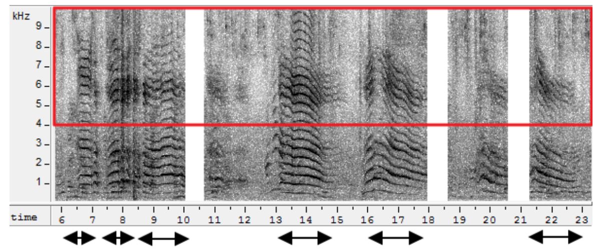

Figure 2: Spectrogram for cry due to Environmental change, depicting monotonous excitation contour (Marked by arrows) and lesser spectral energy in higher spectrum (Marked by rectangle box).

is collected for this study. From a total of 7 categories noted for the data-collection at the clinic, 4 are specifically focussed upon, for acoustic analysis and cause classification. Amongst these, *Pain* and *Stranger's anxiety* are the most prominent reasons that induced crying during the hospital visits for majority of infants. Whereas, *Discomfort* and *Environmental change* were other important cry categories for which the data is collected. On-going and historical medical conditions, parent's inputs and infant's present health status as adjudicated by the doctor, formed the basis of ground truth categorization. Due to the category-wise constraint of cry sample count in the data set, these categories are aggregated to form higher level classes, *Severe* and *Non-severe*, for the study in this work. Further corpus details can be referred from (Sharma and Mittal, 2017a).

4 Experimental setup

4.1 Acoustic Analysis

Cry signals are recorded at 48 kHz and 24 bit coding rate. The short-time analysis of the cry signal is done by considering Frame size of 30 ms and frame shift of 10 ms. Denoising of the computed fundamental period is done using median filtering of order 5. For filter-bank analysis, 6 sub-bands covering the successive spectrum ranges of 1 kHz each, starting from 100 Hz up-to 6 kHz are considered. The order of the Mel scale cepstrum is set as 13. Feature extraction and conversion into vector format is done using MATLAB R2017a and Python routines.

4.2 Cause Classification

Classifier models are implemented using statistics and machine learning, parallel processing and neural network toolboxes, in MATLAB R2017a.

The neuron count in the *Hidden layer* of the *Feed-forward neural network* has been empirically evaluated, based upon the following conventions (Heaton, 2008),

- $I_{nc} < H_{nc} < O_{nc}$,
- $H_{nc} = \frac{2}{3}(I_{nc} + O_{nc})$ and
- $H_{nc} < 2^{*} I_{nc}$.

where, I_{nc}, H_{nc} and O_{nc} are the neuron count for input, hidden and output layers respectively. *Convolution neural network* architecture based evaluation is done using Keras routines with Tensor flow as backend in IPython environment. Adopted from the task on *environmental sounds classification* (ESC) (Salamon and Bello, 2017), the CNN has 4 convolution layers, each followed by a relu activation layer, with max-pooling of size (2,2) and dropout after every set of 2 conv-activation layers, dropout being 15 and 20 % respectively. Next is a fully connected layer with 256 neurons with relu activation and a 50 % dropout. Finally the output layer neurons are activated using softmax function.

5 Acoustic analysis of Infant Cry

The primary experiments are focussed upon validating the acoustic characteristics observed from the cry spectrograms, as shown in Fig. 1 and 2. It is observed that the cry signals having majority of the cries with relatively *less* deviation of the F_0 contour and more stability (Fig. 2), are mostly observed for the causes that are *less severe* in nature, as can be observed from the F_0 contour plotted in the Fig. 4 (sub-plot (b)) for an Environmental change case. Cries due to *Discomfort* and *Environmental change* fall under this category. Whereas, the presence of arc-shaped excitation contours within a cry event as depicted in

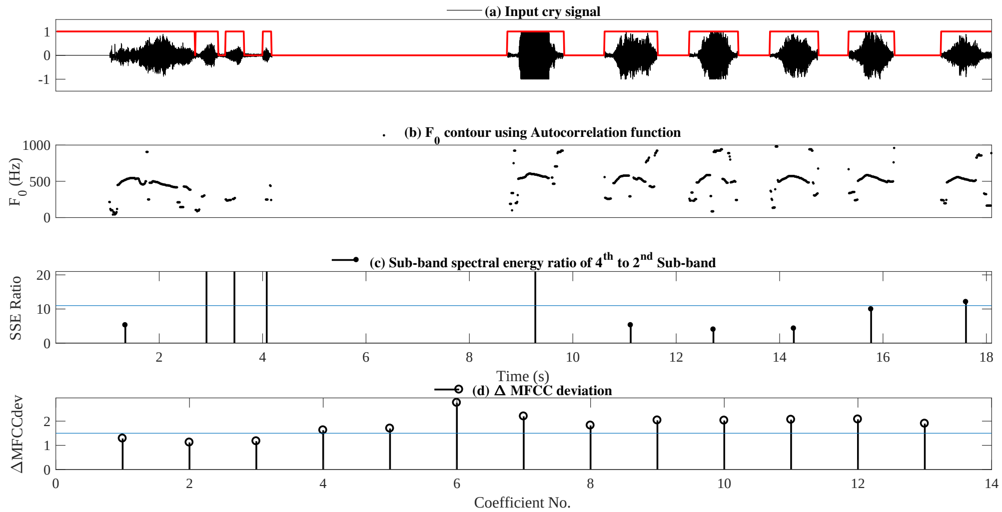

Figure 3: Comparison of (a) input signal, (b) F_0 contour (using autocorrelation), (c) Sub-band spectral energy ratio of 4^{th} to 2^{nd} sub-bands and (d) Δ $MFCC_{dev}$ for Pain cry.

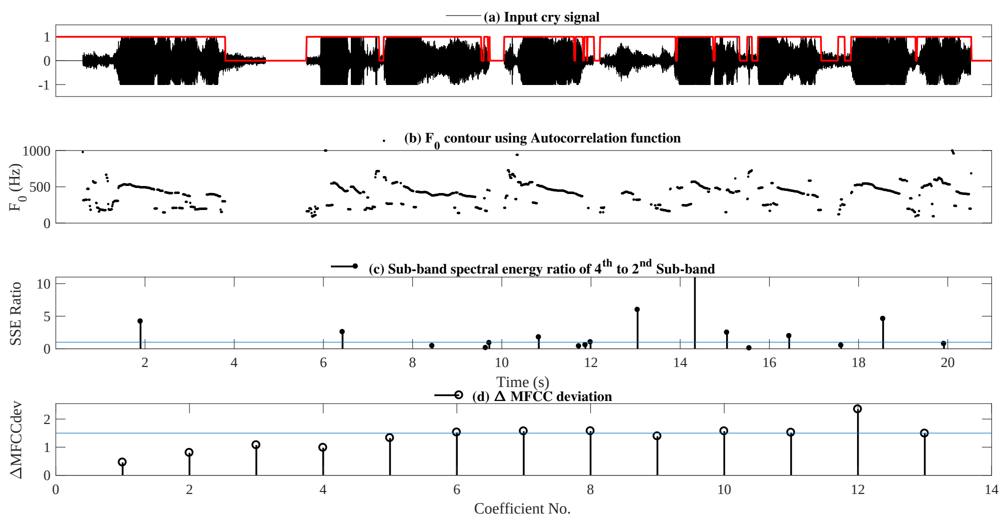

Figure 4: Comparison of (a) input signal, (b) F_0 contour (using autocorrelation), (c) Sub-band spectral energy ratio of 4^{th} to 2^{nd} sub-bands and (d) Δ $MFCC_{dev}$ for cries due to for Environmental change.

Fig. 1 indicate causes that are *severe* in nature. Cries due to *Pain* and *Stranger's anxiety* exhibit such behaviour, which can be validated from the F_0 contour of a Pain cry, as shown in Fig. 3 (subplot (b)).

For the *Pain* category, the spectral intensity appears to be more across the spectrum, as compared to that of the *Environmental change*. This effect can also be observed from the cases depicted in Fig. 1 and 2, for Pain and Environmental change respectively, wherein the spectral intensity is consistently observed to be either equivalent or more, for higher sub-bands as compared to the lower ones for *severe* categories as that for *pain*. Whereas, the same is observed to fade away for the core cry segment progressions for the *non-severe* cases like *Environmental change*. The presence of higher spectral intensity within the 4^{th} filter-bank for *pain* category as observed from the experiments, is also demonstrated by comparing the filter-bank energy ratios of 4^{th} sub-band to 1^{st} and 2^{nd} sub-bands, for 85 Severe and 30 Non-

Figure 5: Confusion matrix for the 2 layered Feed-forward neural net based classification, with 90 hidden neurons.

sever cases. The average ratios for Severe category are found to be higher, at 6 and 39 respectively, whereas they are observed to be much lower at 2 and 12 for Non-severe categories. The decrease in the spectral intensity for the cries from the latter set of categories with respect to the sub-band spectral energies could possibly be attributed to the effect of *hypo-phonation* induced from the voice effects like *soft shrill*. Sub-band spectral energy ratios, plotted for the cases for Pain and Environmental change cries, as shown in Fig. 3 and 4 (sub-plots (c)) respectively, with the average ratio value for the former category being significantly *higher* than that for the latter, distinctly characterize cries as either *Severe* or *Non-severe*.

Although, infants are hardly capable of mimicking the linguistic vocalizations being actively used in their surroundings in their infancy, they do develop profound effects of paralinguistic like intonations that characterize their cultural backgrounds (Mampe et al., 2009). In an attempt to capture such time-varying spectral characteristics, MFCCs are found to divulge the cause specific characteristics effectively when subjected to distinctive classification. The qualitative difference can be easily observed by comparing the $\Delta MFCC_{dev}$ plots (Fig. 3 and 4, sub-plots (d)) for the cases from Pain and Environmental change categories, the average value of which is observed to be *higher* for *Severe* as compared to that for *Non-severe* cases. This implies greater modulations within vocalizations in the core crying regions of the bouts, captured by the dynamic functionals modelling the time-varying system characteristics.

6 Crying cause classification

6.1 Using conventional machine learning techniques

The cry-cause recognition by classification is first attempted using technique k-nearest neighbour. It is observed that $mean_{F_0}$, dev_{F_0}, SSE_r, are not capable enough of giving good overall accuracy, without compromising on true prediction for class Non-severe. Whereas, MFCC features are observed to be facilitating predictions with at least 30 % true positive rates for class Non-severe with $\Delta MFCC_{dev}$ giving the highest rate of 34 %, which is significantly greater than the performance for the rest of the features, without using any ensemble classification technique. $\Delta MFCC_{dev}$ is found to be successful in facilitating decent classification for cry sounds as either Severe or Non-severe, with an overall accuracy of 80 % and true positive classification rates of 96 % and 34 % for both classes respectively, with Cosine k-NN based classification. It is important to note that this performance is obtained while evaluating the data-set with significantly skewed instance distribution, dominated by *Severe* class instances. This effect is taken into consideration by evaluating ensemble techniques like RUS boosting that takes such imbalance into consideration while evaluating the classification, which resulted in a base-line performance of 65 %. In addition to this, the problem of instance imbalance is further addressed by performing data-augmentation for audio data-set, while evaluating CNN based classification, which is discussed in further sub-sections.

6.2 Using Feed-forward neural network

Neural network based classification performs with average overall accuracy of 87.64 % with acceptable true positive prediction rates of 92.96 % and 68.68 % for the respective classes. Empirical evaluation elucidated that with the increasing no. of hidden layer neuron count and epochs, for which the neural network converges, the cross entropy error reduces. Neural network with 90 Neurons in

Table 1: Comparison of CNN based classification performances; (a) Configurations detail format: height×width (kernel), n1-n2-n3-n4 (No. of filters in 4 layers) and (b) Datasets (Augmented).

(a) Configurations	*(b) Datasets*	*(c) Accuracy (%)*	*(c) Val. (%)*	*(d) Test (%)*
2×2, 32-32-64-64	Cry-3	57	68	63
	ESC-12	49	66	62
4×3, 12-12-16-16	Cry-3	62	60	65
Average		*56*	*65*	*63*

Table 2: Comparison of best results of several State-of-the-Art for 2/Similar class and the *Proposed* (3^{rd} observation) approach; Classifier abbreviations (c) Feed-forward neural network (FFNN), Scaled conjugate gradient (SCG) and Radial basis function network (RBFN); References in (a): [1](Orozco et al.), [2,5](Wahid et al., 2016), [3](Abdulaziz and Ahmad, 2010) and [4]*(Ours)*.

(a) Database (No. of infants)	*(b) Features*	*(c) Classifier*	*(d) Accuracy (%)*
47 Hunger, 47 Pain[1]	LPC, Intensity	FFNN (SCG)	74.70
350 Hunger, 192 Pain[2]	MFCC, LPCC, Dynamics	RBFN	86.54
88 Pain, 88 Non-Pain[3]	MFCC, LPCC	FFNN (SCG)	91.43
85 Severe, 30 Non-severe[4]	*Pitch, SSE, MFCC, Dynamics*	*FFNN (SCG)*	***93.90***
879 Deaf, 157 Normal[5]	MFCC, LPCC, Dynamics	RBFN	99.42

the hidden layer outperformed all other configurations converging at 35th epoch, giving 93.9 % accuracy rate, with 95.3 % and 90 % as true positive rates for the respective classes, the confusion matrix for which can be observed from Fig. 5.

6.3 Using convolutional neural network (CNN)

Primary objective of evaluating a CNN is to examine the spatial pattern recognition capability along-with popular data-augmentation technique, while addressing the concerns of imbalanced dataset, for the task of infant cry-cause recognition. The original data sub-set is augmented 4 times using techniques like adding white noise, shifting the sound, followed by stretching using the factors 0.8 and 1.2. Also evaluated is the ESC-10 dataset (Piczak, 2015), to examine feature learning towards robust inter-class classification. Main data-sets evaluated are *Cry-3* (600), having Normal (200), Severe (200) and Non-severe (200) cry instances with 50 min. of recordings, and ESC-12 (2400) *with* Cry-3 instances along-with 9 *additional* environmental sound classes (210 min.).

Classification without any data-augmentation of-course resulted in poor performance with 34 % test accuracy. Data-augmentation helped increase the performance by approx. 30 %, which is significant. The key performances from the experimentation can be observed from the Table. 1, with configuration details specified as {*height×width (kernel), n1-n2-n3-n4 (No. of filters in 4 layers)*} and augmented data-sets. Second observation having *tall* filters with dimensions 4x3 and No. of filters as 12, 12, 16 and 16, with the Cry-3 (augmented) dataset, giving an overall accuracy of 62 % with 60 % validation and 65 % test accuracy outperforms other similar evaluations.

7 Results and Discussion

The results from feed-forward neural network are compared with some popular State-of-the-Art approaches (Table 2). With the available set of resources for cry analysis, a simple approach to observe and classify the infant cries as either *Severe* or *Non-severe* in this work, establishes performance up-to 93.9 %, outperforming conventional approaches and several similar/2-class classification attempts made earlier.

CNN based results do not produce optimal performance results, as against the ones already reported in (Zabidi et al., 2017), while significant fine-tuning of network hyper-parameters still being required with the current setup. It does give insights regarding the efficacy of data augmentation for such scenarios, and using small *but taller* filters, that capture both the localized and spectral acoustic patterns, relevant in case of infant cry sounds analogous to the additional effect of temporal progression of speech utterances too.

8 Summary and Conclusion

A multi-class dataset of infant cries is used towards the infant cry-cause analysis and classification. Excitation source and production system characterising features are evaluated. It is established, that F_0 contour, sub-band spectral energy and MFCCs can distinctly characterize cries w.r.t. different causes and severity.

The significant differences in the average pitch, excitation contour patterns and spectral intensity variations across the frequency spectrum and the filter banks thereof, all for different cry-causes under consideration is established. Non-neural network based classifications yield $\approx$ 50 % true positive rates, which provided baseline for the current work. Whereas, MFCCs and related derivatives have shown promising performances with an average classification accuracy of $76.9\ \%$, and also the highest accuracy of $80\ \%$ for $\Delta\ MFCC_{dev}$, suggesting the utility of the time-varying deviation in the rate of change of the *system characteristics* represented by the $MFCC$ coefficients. The required non-linearity is observed to be modelled best by the feed-forward neural networks with accuracy up-to $93.9\ \%$. Convolutional neural networks are observed to learn discriminative feature representations that helped provide improvement upon the initial baseline obtained.

The qualitative analysis of the cry acoustics led to several observable patterns that can be refined using better cry signal processing. Such patterns are also significantly being explored within the speech recognition community that is involved with utilizing the localized spatial pattern recognition using the convolutional neural networks, but primarily towards the task of automatic speech recognition or speech emotion recognition. Acoustic signal like cry, which is devoid of any linguistic content but full of unconventional non-linguistic utterances, can also be investigated using such techniques.

Acknowledgments

The authors are thankful to Dr. Nizam (M.B.B.S-DCH), Dr. Bhavya (M.D.-Ped.) and Dr. Venkat (M.D.-Ped.), the nursing staff of Pranaam Hospital, Madinaguda, Hyderabad, and parents of the infants for supporting in collection of the cry samples needed for the study.

References

Y. Abdulaziz and S. M. S. Ahmad. 2010. Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients. In *2010 International Conference on Information Retrieval Knowledge Management (CAMP)*, pages 260–263.

Hervé Bourlard and Stéphane Dupont. 1997. Subband-based speech recognition. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 2, pages 1251–1254. IEEE.

Chang, Chang Chuan-Yu, Chuan-Wang, Kathiravan, Lin S, Chen, Chen, and Szu-Ta. 2017. Dag-svm based infant cry classification system using sequential forward floating feature selection. *Multidimensional Systems and Signal Processing*, 28(3):961–976.

A. Chittora and H. A. Patil. 2015. Classification of normal and pathological infant cries using bispectrum features. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 639–643.

Rami Cohen and Yizhar Lavner. 2012. Infant cry analysis and detection. In *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, pages 1–5. IEEE.

D Crystal. 1974. In: Sebeok, t. (ed.), paralinguistics. In *Current Trends in Linguistics*, volume 12, page 265295, The Hague. Mouton.

Orion Fausto Reyes Galaviz and Carlos Alberto Reyes García. 2005. Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system. In *MICAI 2005: Advances in Artificial Intelligence*, pages 949–958, Berlin, Heidelberg. Springer Berlin Heidelberg.

Simon Haykin. 1989. *An Introduction to Analog and Digital Communications*. John Wiley & Sons, Inc., New York, NY, USA. Pp. 652.

Jeff Heaton. 2008. *Introduction to Neural Networks for Java, 2Nd Edition*, 2nd edition. Heaton Research, Inc.

Y. Lavner, R. Cohen, D. Ruinskiy, and H. Ijzerman. 2016. Baby cry detection in domestic environment using deep learning. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5.

James Lyons. 2012. Mel frequency cepstral coefficient (mfcc) tutorial.

Birgit Mampe, Angela D. Friederici, Anne Christophe, and Kathleen Wermke. 2009. *Newborns' Cry Melody Is Shaped by Their Native Language*, volume 19.

V. K. Mittal. 2016a. Discriminating features of infant cry acoustic signal for automated detection of cause of crying. In *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5.

V. K. Mittal. 2016b. Discriminating the infant cry sounds due to pain vs. discomfort towards assisted clinical diagnosis. In *Proc. SLPAT 2016 Workshop on Speech and Language Processing for Assistive Technologies*, pages 37–42.

V. K. Mittal and A. K. Vuppala. 2016. Changes in shout features in automatically detected vowel regions. In *2016 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5.

V. K. Mittal and B. Yegnanarayana. 2013. Effect of glottal dynamics in the production of shouted speech. *The Journal of the Acoustical Society of America*, 133(5):3050–3061.

Amy Neustein, editor. 2010. *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics*. Springer, New York.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck. 1989. *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs. Pp. 71.

Jose Orozco, Carlos A. Reyes Garcia, Luis Enrique Erro, and Tonantzintla Puebla Mxico. Detecting pathologies from infant cry applying scaled conjugate gradient neural networks.

M. Petroni, M.E. Malowany, C.C. Johnston, and B.J. Stevens. 1994. A new, robust vocal fundamental frequency (f0) determination method for the analysis of infant cries. In *Computer-Based Medical Systems, 1994., Proceedings 1994 IEEE Seventh Symposium on*, pages 223–228.

Karol J Piczak. 2015. Environmental sound classification with convolutional neural networks. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, pages 1–6. IEEE.

O. F. Reyes-Galaviz, S. D. Cano-Ortiz, and C. A. Reyes-Garca. 2008. Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies. In *2008 Seventh Mexican International Conference on Artificial Intelligence*, pages 330–335.

R. Sahak, W. Mansor, Y. K. Lee, A. I. Mohd Yassin, and A. Zabidi. 2010. Orthogonal least square based support vector machine for the classification of infant cry with asphyxia. In *2010 3rd International Conference on Biomedical Engineering and Informatics*, volume 3, pages 986–990.

Justin Salamon and Juan Pablo Bello. 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283.

Bjorn Schuller, Stefan Steidl, and Anton Batliner. a. The interspeech 2009 emotion challenge. In *INTERSPEECH*, pages 312–315. ISCA.

Bjorn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus R. Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, Marcello Mortillaro, Hugues Salamin, Anna Polychroniou, Fabio Valente, and Samuel Kim. b. The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In *INTERSPEECH*, pages 148–152. ISCA.

S. Sharma and V. K. Mittal. 2017a. Infant cry analysis of cry signal segments towards identifying the cry-cause factors. In *TENCON 2017 - 2017 IEEE Region 10 Conference*, pages 3105–3110.

S. Sharma and V. K. Mittal. 2017b. A qualitative assessment of different sound types of an infant cry. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, pages 532–537.

Shivam Sharma, Pruthvi R. Mykala, Rajasree Nalumachu, Suryakanth V. Gangashetty, and Vinay Kumar Mittal. 2017. A study on acoustic features of infant cry signal for different causes of crying. *3rd international workshop on Affective Social Multimedia Computing (ASMMC), INTERSPEECH 2017*. Stockholm.

N. S. A. Wahid, P. Saad, and M. Hariharan. 2016. Automatic infant cry classification using radial basis function network. In *Journal of Advanced Research in Applied Sciences and Engineering Technology*, volume 4 of *1*, pages 12–28. ISSN (online): 2462-1943.

B. Yegnanarayana and K. Sri Rama Murty. 2009. Event-based instantaneous fundamental frequency estimation from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(4):614–624.

A. Zabidi, W. Mansor, L. Y. Khuan, I. M. Yassin, and R. Sahak. 2010. The effect of f-ratio in the classification of asphyxiated infant cries using multilayer perceptron neural network. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 126–129.

A Zabidi, Ihsan Yassin, H A. Hassan, Nurlaila Ismail, M M. A. M. Hamzah, Zairi Rizman, and Husna Zainol Abidin. 2017. Detection of asphyxia in infants using deep learning convolutional neural network (cnn) trained on mel frequency cepstrum coefficient (mfcc) features extracted from cry sounds. 9.