Infants Cry Classification of Physiological State Using Cepstral and Prosodic Acoustic Features

Ramon L. Rodriguez and Susan S. Caluya
College of Computer Studies, National University, Philippines
rlrodriguez@national-u.edu.ph

***Abstract*—Infants cry to express their emotional, psychological and physiological states. The research paper investigates if cepstral and prosodic audio features are enough to classify the infants' physiological states such as hunger, pain and discomfort. Dataset from our previous paper was used to train the classification algorithm. The results showed that the audio features could classify an infant's physiological state. We used three classification algorithms, Decision Tree (J48), Neural Network and Support Vector Machine in developing the infant physiological model. To evaluate the performance of the infant physiological state model, Precision, Recall and F-measure were used as performance metrics. Comparison of the cepstral and prosodic audio feature is presented in the paper. Our findings revealed that Decision Tree and Multilayer Perceptron performed better both for cepstral and prosodic feature. It is noted the cepstral feature yielded better result compare with prosodic feature for the given dataset with correctly classified instances ranging from 87.64% to 90.80 with an overall kappa statistic ranging from 0.47 – 0.64 using cepstral feature.**

***Index Terms*—Audio Features; Audio Signal; Infants Cry; Infants State; Machine Learning Algorithms.**

I. INTRODUCTION

Infants cry to express their discomfort, pain, and hunger. Infant crying is a tool for communication and interaction with the external world. The crying sound produced during crying can be used in determining the emotional, psychological and physiological states of the infant. Cry signals have been investigated for a couple of years for medical condition and diagnosis purposes. A mother sometimes understands the baby by looking at other cues such as body language, facial expression and other contextual information related to the infant. As infants communicate through crying, it is quite important to know the physiological state of infants, so that mothers or people around them know how to address their needs. According to Zeman [1], infants begin expressing their emotions during the last half of the first year and oftentimes crying is used to signal about their discomfort or displeasure.

Many researches have studied the cries of infants for different purposes. Some study focused on the medical problem and diagnosis and other studies on emotional, physical and physiological states. In Fuhr, Reetz, and Wegener's study [2], they trained the supervised learning models of infant cry to know if the infant is healthy or not. Petroni et al. [3] attempted to investigate and classify the categories of an infant's cry, namely: 'pain', 'fear', and 'hunger' using three different kinds of Neural Networks: a simple feed-forward neural network, a recurrent neural network (RNN), and a time delay neural network (TDNN). The results of this study demonstrated that the highest classification rate was achieved using the simple feed-forward neural network. Moreover, Barajas-Montiel and Reyes-Garcia [4] used supervised algorithm for the classification of infant's cry from hunger and pain cries. Support Vector Machine (SVM) algorithm was used for their classification system and Mel Frequency Cepstral Coefficients (MFCC) audio feature was used. Orozco and Reyes-Garcia [5] used the linear predication method to extract the acoustic features from their infant cry dataset which are then fed to a feed-forward neural network recognition module.

The acoustic features of an infant's cry are of great importance in this field of signal processing. This acoustic signal contains valuable information about their physical, emotional and psychological condition, such as health, identity, gender and emotions according to Cohen and Lavner [6]. These information was used in developing models for infant cries using supervised learning algorithm. The recognition of the medical status of newborn infants, the classification whether the baby is healthy or not and infants physiological and emotional state are some common instances where the infant's cry can be used.

This paper will investigate if cepstral and prosodic audio features of infant cry alone can be sufficient to classify an infant's psycho-physiological state such as hunger, pain or discomfort.

II. REVIEW OF RELATED LITERATURE

This section discusses the related works on infants crying classification techniques and approaches. In the works of Saraswathy, Hariharan, Yaacob, and Khairunizam [7], the author presented a thorough review of the classification of infant cry. On the other hand, the study of Abdulaziz and Ahmad [8] compared both cepstral features MFFC and LPCC on automatic infant cry recognition system but focused on pain and not pain. Moreover, LFCC as features and KNN algorithm were used to classify hunger, sleepiness, and pain [9]. Another study by Sing, Mukhopadhyay, and Rao [10] used Epoch Interval Contour (EIC) and MFCC features in classifying infant cries. Several studies focused on the feature extraction of the infant's cry such as weighted segment-based two-dimensional linear frequency cepstral coefficients to characterize the time-frequency patterns within a long-range segment of the target signal [11], Discrete Fourier Transform coefficients [12], zero-crossing rate and fundamental frequency [13], and spatio-temporal box filter [14]. Infant cry classification using different supervised algorithms to detect the physiological and psychological states can be used in the medical field [15]. Another medical application for infants' cry study is the infant pathology classification using both dynamic and static MFCC audio features. Gaussian mixture model classification algorithm was used in [16]. In all the reviewed studies, supervised learning algorithm was used in the training of audio feature set. It is important to note that supervised learning algorithm yielded a better result. Fuhr, Reetz, and Wegener [2] presented the update and systematic review of the different supervised learning algorithm.

III. Research Method

This section presents the methodology used in the study. The research used the existing audio recording in the previous paper. Infant cries whose age ranges from 1 to 9 months were recorded. Out of 53 audio recording, only 18 recording were used in the training of the classifier to come up with a balanced dataset [17]. Audacity tools were used in the cleaning of the audio recording. In the previous paper [17], cepstral features (MFCC and LPCC) were extracted using MATLAB. Aside from cepstral acoustic features, prosodic acoustic features of the audio recordings were extracted using PRAAT. Prosodic features were time, intensity, formants and pitch. Both extracted acoustic features were saved as CSV file separately. Each instance for the two (2) dataset were labeled as pain, hunger and discomfort. The label of the instances was based on label of the audio clip.

WEKA software was used to train the system and to determine the infant physiological state model development for the two datasets on the cepstral audio features and the prosodic audio features. Decision Tree J48, Neural Network MLP and Support Vector Machine SMO were used to train the model for both datasets. In evaluating the model, Precision, F-Measure and Recall were used as a metric of evaluation. Cross-validation of 10 folds was used in training the system. The following is the short description of the classifiers and the performance measure.

J48 is a Decision Tree Algorithm implemented in WEKA. J48 classifier is a simple C4.5 decision tree for classification. It creates a binary tree. The decision tree approach is most useful in classification problem.

Multilayer Perceptron (MLP) is an implementation of Neural Networks algorithm in WEKA. This is the most prominent type of neural network which belongs to a class of networks called feedforward networks because they do not contain any cycle and the network's output depends only on the current input instance.

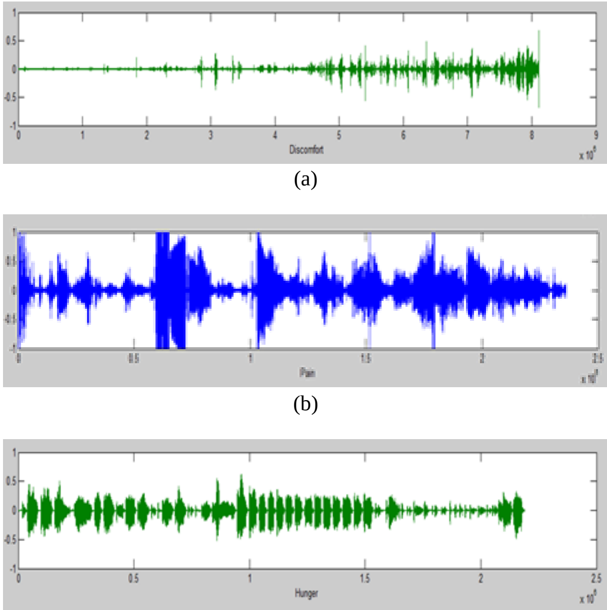
SMO is a support vector machine (SVM) implementation in WEKA. This algorithm uses linear model to implement nonlinear class boundaries.

To evaluate the model, Precision, Recall, and F-Measure were used. Precision is the probability that a class A is true among all that have been classified as class A. This is also referred to as the Positive Predictive Value (PPV). Recall is the proportion of examples which were classified as class A among all instances of class A. This is also referred to as the True Positive Rate or Sensitivity. F-Measure is the combined computation of Precision and Recall; it is the harmonic mean of Precision and Recall in which both are evenly weighted.

We also look at the True Positive (TP) Rate and False Positive Rate (one class classified as other class and the kappa statistics). These measures are determinants whether the physiological infant state model is good for classification.

IV. Results and Discussion

In this paper, 18 audio recordings were used to train the classification algorithm given the two datasets for the cepstral and prosodic features. The sample audio wave plot of the physiological state of the infants labeled as discomfort, pain and hunger is presented in Figure 1a, 1b, and 1c. Given the wave plot empirical test and visual inspection of the audio wave of the physiological state have shown that pain state can be easily detected. It shows that audio wave for discomfort and hunger has almost the same frequency, while for pain has a different audio wave plot. This is the reason that pain is easy to detect even for visual inspection.



(a) (b) (c)

Figure 1: Sample audio clip for (a) discomfort, (b) pain, and (c) hunger

Based on the inspection alone of the wave plot, it is suggested that prosodic acoustic features can be used to further enhance the accuracy of the model. The cepstral and prosodic features of each audio recording were extracted and the machine learning algorithm was used to know the pattern for each physiological state. Results show that the correctly classified instances range from 77.20% - 90.80%, using the cepstral and 64.76% - 83.87% using prosodic features based on the three classification algorithms. Among the three classification algorithms, the J48 decision tree implementation in Weka yielded higher accuracy result compared with Multilayer Perceptron and Support Vector Machine algorithms for both datasets. Table 1 shows the summary of the result of the three classification algorithms. The values of the Precision, Recall and F-Measure are the weighted average (WA).

Looking at the result, based on the evaluation metric, the Precision, Recall and F-Measure show that cepstral acoustic features are better compared with the prosodic features. The weighted average of the Precision ranges from 0.772–0.908. On the other hand, the prosodic features' weighted average of Precision ranges from 0.642–0.835. The infants' state can be detected based on cry. The Recall and F-Measure range from 0.772 to 0.908 for the cepstral features and 0.648–0.839 for the prosodic features. The model has an ability to distinguish the three physiological states from each other as shown in the performance measure result. It is noted that cepstral features yield better result compared with the prosodic features. The J48, SMO, and MLP detailed accuracy by class are shown in Table 2, 3 and 4.

Table 1
Summary Result of the Performance Measure Using Cepstral and Prosodic Features

Performance Measure	Classification Algorithm					
	J48 (Cepstral)	J48 (Prosodic)	SMO (Cepstral)	SMO (Prosodic)	MLP (Cepstral)	MLP (Prosodic)
Correctly classified instances	90.80%	83.87%	77.20%	64.76%	87.64%	67.92%
WA Recall	0.908	0.839	0.772	0.648	0.876	0.679
WA F-Measure	0.908	0.838	0.772	0.644	0.876	0.670
WA Precision	0.908	0.838	0.772	0.642	0.876	0.668

Table 2
J48 Detailed Accuracy by Class

Class	TP Rate (Cepstral)	TP Rate (Prosodic)	Precision (Cepstral)	Precision (Prosodic)	Recall (Cepstral)	Recall (Prosodic)	F-Measure (Cepstral)	F-Measure (Prosodic)
Hunger	0.893	0.855	0.900	0.853	0.893	0.882	0.896	0.876
Pain	0.903	0.750	0.901	0.761	0.903	0.855	0.902	0.856
Discomfort	0.929	0.882	0.924	0.876	0.929	0.750	0.926	0.755
Weighted Average	0.908	0.839	0.908	0.838	0.908	0.839	0.908	0.838

Table 3
SMO Detailed Accuracy by Class

Class	TP Rate (Cepstral)	TP Rate (Prosodic)	Precision (Cepstral)	Precision (Prosodic)	Recall (Cepstral)	Recall (Prosodic)	F-Measure (Cepstral)	F-Measure (Prosodic)
Hunger	0.725	0.748	0.720	0.709	0.725	0.748	0.728	0.728
Pain	0.833	0.449	0.782	0.522	0.833	0.449	0.808	0.483
Discomfort	0.755	0.665	0.805	0.644	0.755	0.665	0.779	0.755
Weighted Average	0.772	0.648	0.772	0.642	0.772	0.648	0.772	0.644

Table 4
MLP Detailed Accuracy by Class

Class	TP Rate (Cepstral)	TP Rate (Prosodic)	Precision (Cepstral)	Precision (Prosodic)	Recall (Cepstral)	Recall (Prosodic)	F-Measure (Cepstral)	F-Measure (Prosodic)
Hunger	0.868	0.748	0.863	0.695	0.860	0.772	0.865	0.731
Pain	0.875	0.394	0.871	0.529	0.875	0.394	0.873	0.452
Discomfort	0.876	0.668	0.896	0.734	0.886	0.768	0.891	0.751
Weighted Average	0.876	0.679	0.876	0.668	0.874	0.679	0.876	0.670

The detailed accuracy by class shows that pain physiological state is easier to distinguish from other physiological states such as hunger and discomfort, given the TP Rate, Precision, Recall, and F-Measure using cepstral feature. The result shows that for all performance measure, pain is consistent with higher value compared with other physiological states based on cepstral feature, but for prosodic feature, hunger and discomfort yields higher result compared with pain.

Three different classification algorithms were used in classifying infant physiological state from acoustic information: Decision Tree, Support Vector Machine, Neural Network. All these classification algorithms are WEKA implementations. Although all classification algorithms have good results, it is observable that Decision Tree performed better among all other classification algorithms for both cepstral and prosodic features with 90.80% and 83.87% accuracy results, respectively.

V. Conclusion

In this paper, we analyzed the infant cry to determine the physiological state such as pain, hunger, and discomfort. The goal of this paper is to determine what acoustic features can be used in detecting the physiological state of an infant. Infant-cry-signals provide valuable information to determine the psycho-physiological state of infants. The result shows that J48 Decision Tree and Multilayer Perceptron Neural Network and Support Vector Machine are classifiers that can be used in training the system for classification of psycho-physiological state of the infant for both cepstral and prosodic features. The consideration of gathering more audio recording on several types of pain and discomfort can further enhance the psycho-physiological model. Since the study did not consider the gender of the infant, this can be another point for analysis to further enhance the model. Another thing to consider is the recording of the hunger data.

Hence, the classification algorithm used in the study demonstrates a higher accuracy result for the given dataset. The combination of cepstral and prosodic features can be done to know if combining the feature will give a better result.

Acknowledgment

This study is an Internally Research Funded Project of the National University Research and Innovation Office (NU-RaIn) – Philippines.

References

[1] J. Zeman, "Emotional Development – Early Infancy, Later Infancy Months," Retrieved from http://psychology.jrank.org/pages/212/Emotional-Development.html.

[2] T. Fuhr, H. Reetz, and C. Wegener, "Comparison of Supervised-learning Models for Infant Cry Classification," *Int. J. Health Professionals*, vol. 2, no. 1, pp 4–15, 2015, doi: 10.1515/ijhp-2015-0005.

[3] M. Petroni, A.S. Malowany, C.C. Johnston, and B.J. Stevens, "A Comparison of Neural Network Architectures for the Classification of Three Types of Infant Cry Vocalizations," in *Proc. 17th Annu. Conf. of IEEE Eng. in Medicine and Biology Soc.*, Montreal (Quebec, Canada), Sep. 1995, vol. 2, pp. 975–976, doi: 10.1109/IEMBS.1995.575380.

[4] S. Barajas-Montiel and C. Reyes-Garcia, "Identifying Pain and Hunger in Infant Cry with Classifiers Ensembles," in *Proc. Int. Conf. on Comput. Intell. for Modelling, Control and Autom. and Int. Conf. on Intell. Agents, Web Technol. and Internet Commerce (CIMCA-IAWTIC)*, Vienna (Austria), Nov. 2005, vol. 2, pp. 770–775, doi: 10.1109/CIMCA.2005.1631561.

[5] J. Orozco and C.A. Reyes-Garcia, "Implementation and Analysis of Training Algorithms for the Classification of Infant Cry with Feed-Forward Neural Networks," in *Proc. IEEE Int. Symp. on Intell. Signal Process.*, Budapest (Hungary), Sep. 2003, pp. 271–276. doi: 10.1109/ISP.2003.1275851

[6] R. Cohen and Y. Lavner, "Infant Cry Analysis and Detection," in *Proc. IEEE 27th Conv. of Elect. and Electron. Engineers in Israel (IEEEI)*, Eilat (Israel), Nov. 2012, doi: 10.1109/EEEI.2012.6376996.

[7] J. Saraswathy, M. Hariharan, S. Yaacob, and W. Khairunizam, "Automatic Classification of Infant Cry: A Review," in *Proc. Int. Conf. on Biomed. Eng. (ICoBE)*, Penang (Malaysia), Feb. 2012, pp. 543–548, doi: 10.1109/ICoBE.2012.6179077.

[8] Y. Abdulaziz and S.M.S. Ahmad, "Infant Cry Recognition System: A Comparison of System Performance Based on Mel Frequency and Linear Prediction Cepstral Coefficients," in *Proc. Int. Conf. on Inform. Retrieval and Knowl. Manag.*, Shah Alam (Selangor, Malaysia), Mar. 2010, pp. 260–263, doi: 10.1109/INFRKM.2010.5466907.

[9] S.S. Jagtap, P.K. Kadbe, and P.N. Arotale, "System Propose for Be Acquainted with Newborn Cry Emotion Using Linear Frequency Cepstral Coefficient," in. *Proc. Int. Conf. on Elect., Electron., and Optimization Techn. (ICEEOT)*, Chennai (India), Mar. 2016, pp. 238–242, doi: 10.1109/ICEEOT.2016.7755094.

[10] A.K. Singh, J. Mukhopadhyay, and K.S. Rao, "Classification of Infant Cries Using Epoch and Spectral Features," in *Proc. Nat. Conf. on Commun. (NCC)*, New Delhi (India), Feb. 2013, doi: 10.1109/NCC.2013.6487999.

[11] M.J. Kim, Y. Kim, S. Hong, and H. Kim, "ROBUST Detection of Infant Crying in Adverse Environments Using Weighted Segmental Two-dimensional Linear Frequency Cepstral Coefficients," in *Proc. IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, San Jose, (CA, USA), Jul. 2013, doi: 10.1109/ICMEW.2013.6618321.

[12] S. Yamamoto, Y. Yoshitomi, M. Tabuse, K. Kushida, and T. Asada, "Recognition of a Baby's Emotional Cry Towards Robotics Baby Caregiver," *Int. J. Advanced Robot. Syst.*, vol. 10, no. 2, Feb. 2013, doi: 10.5772/55406.

[13] A.M. Prathibha, P. Raju, S. Halvi, and S.B. Satish, "An Eclectic Approach for Detection of Infant Cry and Wireless Monitoring of Swinging Device as an Alternative Warning System for Physically Impaired and Better Neonatal Growth," *World J. Sci. and Technol.*, vol. 2, no. 5, pp. 62–65, May 2012.

[14] P. Ruvolo and J. Movellan, "Automatic Cry Detection in Early Childhood Education Settings," in *Proc. 7th IEEE Int. Conf. on Develop. and Learn. (ICDL)*, Monterey (CA, USA), Aug. 2008, pp. 204–208, doi: 10.1109/DEVLRN.2008.4640830.

[15] L. Abou-Abbas, C. Tadj, C. Gargour, and L. Montazeri, "Expiratory and Inspiratory Cries Detection Using Different Signals' Decomposition Techniques," *J. Voice*, vol. 31, no. 2, pp. 259.e13–259.e28, Mar. 2017, doi: 10.1016/j.jvoice.2016.05.015.

[16] H.F. Alaie, L. Abou-Abbas, and C. Tadj, "Cry-based Infant Pathology Classification Using GMMs," *Speech Commun.*, vol. 77, pp. 28–52, Mar. 2016, doi: 10.1016/j.specom.2015.12.001.

[17] R. Rodriguez and S. Caluya, "Waah: Infants Cry Classification of Physiological State Based on Audio Features," in *Proc. 5th Int. Conf. on Soft Comput., Intell. Syst. and Inf. Technol. (ICSIIT)*, Kuta (Bali, Indonesia), Sep. 2017, pp. 7–10, doi: 10.1109/ICSIIT.2017.24.