# Why is my Baby Crying? An in-depth Analysis of Paralinguistic Features and Classical Machine Learning Algorithms for Baby Cry Classification

Rodica Ileana Tuduce*, Horia Cucu, and Corneliu Burileanu
Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Romania
*Corresponding author (E-mail: rodica.i.tuduce@speed.pub.ro)

***Abstract*—As human beings, we begin interacting with the world by expressing our basic needs through crying. Parents strive to identify and timely address these needs before hysterical crying sets in. However, first-time parents usually fail, and this leads to frustration and feelings of helplessness. In this context, our work focuses on creating an automatic system able to distinguish between different infant needs based on crying. We extract various sets of paralinguistic features from the baby-cry audio signals and we train various rule-based or statistical classifiers. We evaluate and in-depth compare the results and obtain up to 70% accuracy on the evaluation dataset.**

***Keywords—baby cries; ComPareE; emobase2010; DUNSTAN; openSMILE; WEKA***

## I. INTRODUCTION

Crying is part of a behavioral system improving survival of helpless newborn by eliciting an emotional and behavioral response in their caregivers [1, 2] ensuring subsequent alleviation of their immediate needs. The interpretation of those cries varies with experience, such that decoding a baby's needs can be a guessing game for new parents who, in time, seem to develop an understanding of what the communicated needs are. Priscilla Dunstan first labeled these cries [3] and her classification has since been widely accepted by the pediatric community.

Assessment of a newborn's first cry provides a glimpse of the neurological development and overall medical status of the neonate. These sounds have been used for decades experienced neonatologists and nurses around the world to identify Sudden Infant Death Syndrome (SIDS), substance exposure in utero and various chromosomal abnormalities [4, 5]. For example, a high-pitched cry can signal a baby with acute bilirubin encephalopathy, brain damage, or if sounding like the cry of a cat – "cri du chat syndrome" - a genetic condition in which part of the arm of chromosome 5 is severed, associated with distinctive facial features, small head size, low birth weight and weak muscle tone. In contrast, a low-pitch cry can identify the Cornelia De Lange Syndrome [6], associated with slow growth, low set ears, small upturned noses and behavioral problems.

Generally, when babies are born they are capable of five distinguishable types of crying, documented by P. Dunstan [3]. As they develop, they become capable of increasingly complex crying. Around 2 months, they can engage in intimate communication crying [7] and around 6 months they can even display "fake crying" behavior designed to manipulate their caregiver [8], considered to be a key milestone for both emotional and intellectual development. Our work focuses on distinguishing the 5 primal cries.

Baby cry classification is part of the field of affect recognition, similar to emotion recognition from speech [10]. In emotion recognition, features such as pitch, breathiness, jitter, shimmer, Mel Frequency Cepstral Coefficients (MFCCs), formants and spectral statistics [11] are known to be effective, but new features are continuously proposed [12]. Recently, Parlak et al. showed the importance of looking at multiple features sets in emotion recognition [9].

Due to the difference between infant and adult vocal tracts [13], previous studies raised the question with the accuracy of some features extracted from baby cries [14] such as fundamental frequency. However, previous work on baby cry classification has focused only on a handful of individual features such as frequency, pitch, MFCC, Linear Predictive Codes and Linear Prediction Cepstral Coefficients [14, 15]. Therefore, an analysis of baby cries looking at the change of a wide array of features is needed.

On the classifier side, baby cry studies [14, 16, 19] (including our own previous research [17, 18]) use only a handful of classifiers at a time.

In our previous work, which is part of the same research project[1], we used the classical MFCCs and classifiers such as Gaussian Mixture Model – Universal Background Model (GMM-UBM) and i-vectors, we performed experiments on two baby cry datasets: Dunstan [17] and SPLANN [18, 23].

In this paper, we explore baby cry characteristics by looking at various paralinguistic feature sets, proven to be successful in tasks such as emotion recognition [9], deception, sincerity and native language from speech recognition [20]: the emobase2010 feature set [24] and the INTERSPEECH Computational Paralinguistics Challenge (ComParE) feature set [17]. Going further, we extended our approach by analyzing these features and using them to train the various classifiers available in the Waikato Environment for Knowledge Analysis (WEKA) [22].

This work has been funded by University Politehnica of Bucharest, through the Excellence Research Grants Program, UPB GEX 2017. Identifier: UPBGEX2017, Ctr. No. 32/2017.

[1] SPLANN research project (2014 – 2017) aimed to develop an automatic infant crying recognition system; project consortium: SOFTWIN SRL (coordinator), Speech and Dialogue Research Lab and The Emergency Clinical Hospital "Sf. Pantelimon".

The rest of the paper is organized as follows: Section II looks at the baby cry database, while Section III briefly describes the feature extraction procedure and the two feature data sets. Section IV presents the baby cry recognition experiments and machine learning classifiers. Section V compares the results and Section VI is dedicated to our conclusions.

## II. Baby Cry Study Design

### A. Baby Cry Database

The five types of baby cries denote the following problems: eairh (colic), heh (diaper), eh (burp), neh (hunger) and owh (tiredness). The total number of baby cry .wav files we used is 82, as detailed in Table I, adapted from [11]. The DBL contains baby cries with sampling frequencies of 8 kHz, 16 kHz and 44.1 kHz. We choose to use the highest, 44.1 kHz. Even though the overall number of samples is small, labeling is reliable and the cries chosen to be represented in the recordings are not ambiguous. Previous experiments showed that classification results are more consistent on DBL corpus, than on SPLANN corpus [18, 23] (the corpus collected in the SPLANN project).

### B. Baby Cry Feature Extraction

The Munich open Speech and Music Interpretation by Large Space Extraction (openSMILE) tool [24] enables extraction of large audio feature spaces in real-time. OpenSMILE allows for various configurations used for distinct purposes, such as speech emotion recognition, sincerity recognition, or in our case baby cry classification. It has since become the standard feature-extraction tool for the annual ComParE challenge [20].

To pinpoint the effective feature set required for baby cry recognition, we used the openSMILE feature extraction tool. The first feature set is tailored for emotions. The second feature set is more general and more popular.

TABLE I. BABY CRY TYPES ON THE DUNSTAN DATABASE

| Cry type | # Babies | # Cries | Duration [s] |
|---|---|---|---|
| Flatulence (eairh) | 10 | 12 | 78 |
| Eructation (eh) | 11 | 13 | 60 |
| Discomfort (heh) | 12 | 13 | 43 |
| Hunger (neh) | 23 | 25 | 60 |
| Tiredness (owh) | 16 | 19 | 121 |
| | Total | 82 | 362 |

TABLE II. CLASSIFIERS

| Category | Classifiers |
|---|---|
| Bayes | C1 – Naive Bayes, C2 – Bayes Net. |
| Functions | C3 – Dl4jMlp, C4 – MLPClassifier, C5 – SMO. |
| Lazy | C6 – IBk, C7 – LocalKNN, C8 – LWL. |
| Meta | C9 – END, C10 – IterativeClsfOpt, C11 – LogitBoost, C12 – MultiClassClsfUpd, C13 – RandomCommittee C14 – RandomSubSpace, C15 – RotationForest. |
| Misc | C16 - CHIRP |
| Rules | C17 - FURIA, C18 - NNge, C19 - PART, C20 - Rough Set, C21 – VFDR. |
| Trees | C22 - CDT, C23 - FT, C24 - Hoef, C25 - J48, C26 - J48Consolidated, C27 - J48Graft, C28 - LMT, C29 - RandomForest, C30 – SimpleCart. |

#### 1. *emobase2010*

The emobase2010 is an emotion classification feature set, containing a carefully-tailored set of 1582 features out of the 6552 features (75% less features) of the previous emotion feature set (emo_large). Previous emotion recognition studies have found it to be more accurate [9]. It has been used successfully in various emotion-based studies, such as detecting pain in speech [25].

#### 2. *ComParE 2013*

This is the gold standard feature selection set comprising 6373 features. Schuller et al. [21] provides a detailed description of the set. Though it contains a more complex list of features, it also drives a much longer processing time with feature potentially not needed in baby-cry analysis. When using an .arff file containing several thousand audio files, such as those from the SPLANN database [23], some computations are simply impossible. This is the reason why we chose to conduct our proof of concept study on the smaller, DBL database.

## III. Automated Method Description

### A. Baby Cry Classification Algorithms

When it came to classification algorithms, we decided to have an exhaustive, “brute-force” approach: we evaluated all classical machine learning algorithms in WEKA version 3.8 machine learning tool [22]. WEKA enables the analysis of a large body of data and finding the relevant information, thus allowing for a data-driven approach, opening the door for more accurate predictions.

From the vast array of methods offered by the tool, we experimented with about 45 classifiers. We report the results obtained by the best 30, belonging to seven method classes, as illustrated in Table II.

WEKA is an open-source Java application developed by the University of Waikato in New Zealand. It allows for a wide variety of algorithms to be applied on a preformatted data set. To assess the performance of both feature sets (ComParE, emobase2010), duplicate experiments were run. Each algorithm performance was assessed based on predefined baseline thresholds and limits specified by various parameters. Altering those complex parameters should be the basis of future research. For the purpose on this paper, only baseline settings were used.

### B. Data Splitting

Due to the small size of the DBL database (82 cries, 362 seconds), we decided to use the 10-fold cross validation technique. An alternative would have been to split the data into static training (90%) and evaluation (10%) sets, but due to the small size of the dataset this option would have resulted into statistically inconsistent results. According to the 10-fold cross validation technique, we divided the data set into 10 groups and for each of the 10 runs, nine groups of data served as the training set, while the other one was used for evaluation. We performed 10 iterations of the above procedure to be able to compute a statistically-relevant average of the classification accuracy.

IV. EXPERIMENTAL RESULTS

We present the analysis of the results of the best classifiers from each class. Because we consistently select the default options, we report the baseline results.

## A. Bayes Classifiers

As illustrated in Fig. 1, within the Bayes group of classifiers, the best results were consistently obtained for the emobase2010 feature set. The best Bayes classifier has proven to be C2 (Bayes Net). It obtained an accuracy of 57% if the ComParE feature set was used and a relative accuracy improvement of 21% (to reach an average absolute accuracy of 69%) over the case when the emobase2010 feature set was used.

## B. Function Classifiers

As with the Bayes Classifiers, all Function classifiers C3-C5 report higher relative accuracy improvement for the emobase2010 feature set: 5%, 6% and 8%. The best classifier in this group (see Fig. 1) is C5 (SMO), obtaining 63% accuracy for ComParE and 68% for emobase2010.

## C. Lazy Classifiers

The three best classifiers of the Lazy method class obtained relatively mediocre results. Depicted on the right side of Fig. 1, the best result of the class was C7 (Local KNN) with 59% mean accuracy on the emobase2010 feature set. C8 (LWL), the second result of the group achieves 57% on the ComParE feature set. Overall, 2 out of 3 classifiers (C6 and C7) obtained improved results on the emobase2010 feature set 14% and 7%.

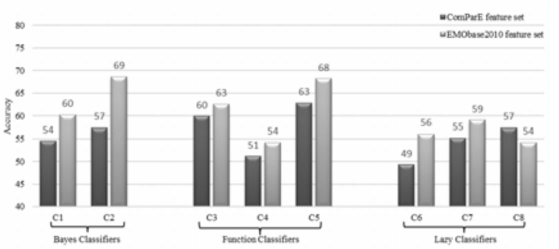


Fig. 1. Accuracy of 3 classifier classes: Bayes, Functions and Lazy

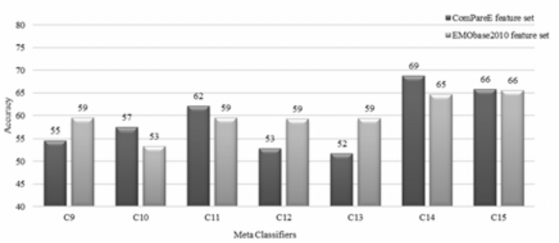


Fig. 2. Accuracy of Meta Classifiers

### 1) Meta Classifiers

The Meta Classifier class contains 7 classifiers, depicted in Fig. 2 below. Four out of seven classifiers (C10, C11, C14 and C15) obtained higher results on the ComParE feature set. The best result of the class is of 69% mean accuracy obtained by C14 (RandomSubSpace), for the ComParE feature set. The second classifier in the class is C15 (Rotation Forest).

## D. Misc Classifiers

C16 (CHIRP) shown in Fig. 3, the only classifier of the Misc category, obtained over 50% accuracy, shows only a marginal difference in accuracy between the 2 feature sets.

## E. Rule Classifiers

As shown in Fig. 3, the best obtained result from the rule-based classifier class is C21 (VFDR) with 62% on emobase2010, which as explained previously, is a very fast performing classifier. C21 sees a 15% relative improvement with emobase2010 vs. ComParE, perhaps because the emotion-based features require less computation, the match becoming apparent in an algorithm requiring only one run through the data. Overall for the rule-based category, 3 classifiers C17, C18 and C21 obtain superior results: 12%, 6% and 15% increased accuracy respectively on the emobase2010 dataset; while the 2 more complex classifiers C19 and C20 perform better on the larger feature set, seeing an approximate 4% change in accuracy.

## F. Tree Classifiers

This class contains the best overall results of the entire experiment, depicted in Fig. 4. It shows an apparent reverse of trend from the above reported classes, with 7 out of 9 classifiers obtaining an improved result on the ComParE feature set. The exceptions are C26 and C29. The best 2 results of the paper are obtained from this class, C28 (LMT) and C23 (FT), both using the ComParE feature set, obtaining 71% and 69% mean accuracy respectively.

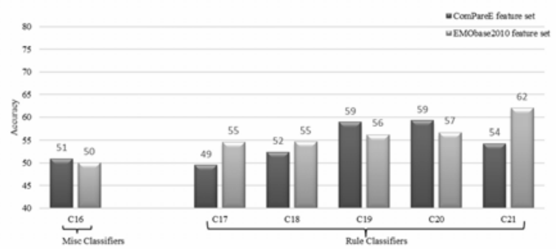


Fig. 3. Accuracy of 2 classfier classes: Misc and Rules

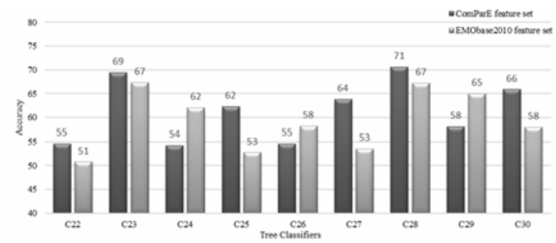


Fig. 4. Accuracy of Tree Classifiers

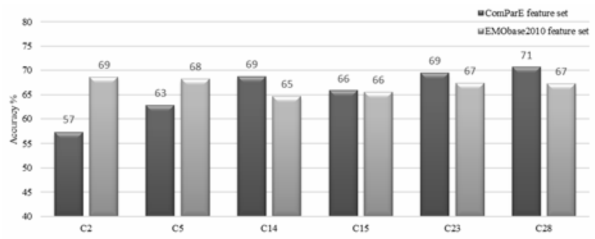


Fig. 5. Best overall classifiers

### G. *Overall Comparison*

Fig. 5 displays the best overall classifiers: C2 (Bayes Net) reaching 69% for the emobase2010 feature set, C5 (SMO), C14 (RandomSubSpace), C15 (RotationForest), C25 (J48) and C28 (LMT). The results are close, with an outlier in C2 for the ComParE feature set. The best overall classifier is C28 with 71% accuracy for ComParE feature set, tied with C14 and C23(FT) for the emobase2010 feature set. Part of the class of Tree classifiers, C28 is capable of learning from massive data streams, which is helpful in this case.

## V. CONCLUSION

Despite dealing with the entire feature sets and baseline algorithmic options, 20 out of the 30 classifiers achieve an accuracy of over 60%, making the classifiers and feature set pairs relatively successful. The success of one emotion tailored feature set (emobase2010) vs. general feature set (ComParE) is evenly split, with 15 classifiers obtaining a better result with one over the other. Overall, the highest accuracy results are obtained for the ComParE feature set with the LMT classifier. Taking into account that the emobase2010 feature set contains approximately 75% less features, the closeness of the obtained results in some instances is encouraging. Interestingly, we obtained a lower score for C9, which uses a combination of a ND classifier and J48, then for C25 containing the J48 classifier alone.

Observed trends can be used as guidelines for different feature sets to be used with certain algorithms, or method types (such as the emobase2010 feature set for Bayes and Function based algorithms.)

Future work will evaluate in-depth the best algorithms by altering the available options. Moreover, best feature selection experiments will be performed for the algorithms which obtained the highest accuracy.

## REFERENCES

[1] R.G. Barr, N. Fairbrother, J. Pauwels, J. Green, M. Chen, and R. Brant, “Maternal frustration, emotional and behavioural responses to prolonged infant crying", Infant Behavior and Development, vol. 37, no. 4, pp. 652-664, 2014.

[2] H.C. Lin, and R. McFatter, "Empathy and distress: Two distinct but related emotions in response to infant crying," Infant Behavior and Development, vol. 35, no. 4, pp. 887-897, December 2012.

[3] P. Dunstan, “Calm the crying: The Secret Baby Language,” Penguin Publishing Group, pp. 16-17, 2012.

[4] H.L. Golub, and M.J. Corwin, "Infant Cry: A Clue to Diagnosis," Pediatrics, vol. 2, no. 197, p. 69, 1982.

[5] L.L. LaGasse, A.R. Neal, and B.M. Lester. "Assessment of infant cry: acoustic cry analysis and parental perception," Developmental Disabilities Research Reviews 11, 1, pp. 83-93, 2005.

[6] M. Uluba. H. Uzun, and D.A. Senses, “A newborn with Cornelia de Lange syndrome: a case report,” Cases J, 1:329, 2008.

[7] H. Nakayama, "Emergence of amae crying in early infancy as a possible social communication tool between infants and mothers," Infant Behavior and Development, vol. 40, pp. 122-130, 2015.

[8] H. Nakayama, “Development of infant crying behavior: A longitudinal case study,” Infant behavior & development, vol. 33, no. 10, pp. 463-71, 2010.

[9] C. Parlak, B. Diri, and F. Gürgen, “A cross-corpus experiment in speech emotion recognition,” in SLAM@ Interspeech, pp. 58-61, 2014.

[10] M. Chóliz, E. G. Fernández-Abascal, and F. Martínez-Sánchez, “Infant Crying: Pattern of Weeping, Recognition of Emotion and Affective Reactions in Observers,” The Spanish Journal of Psychology, 15, 3, pp. 978-988, 2012.

[11] Yi-Chun Ko, "A study on feature sets for emotion recognition from speech signals," PhD Thesis, Univ. Maryland, College Park, 2015.

[12] B. Yang, and M. Lugger, “Emotion recognition from speech signals using new harmony features,” Signal Processing, Vol. 90, Issue 5, pp. 1415-1423, 2010.

[13] M.P. Robb, and A.T. Cacace, "Estimation of formant frequencies in infant cry," Int. J. of Pediatric Otorhinolayngology, Vol. 32, pp. 57-67, 1995.

[14] R. Lavner, Y. Cohen, D. Ruinskiy, and H. IJzerman, “Baby cry detection in domestic environment using deep learning,” IEEE Int. Conf. on the Science of Electrical Engineering, pp. 1-5, 2016.

[15] M. Hariharan, S. Chee, L. Sazali, and Y. Sazali, “Analysis of Infant Cry Through Weighted Linear Prediction Cepstral Coefficients and Probabilistic Neural Network,” Journal of medical systems, 36, pp. 1309-15, 2010.

[16] J. Saraswathy, M. Hariharan, S. Yaacob, and W. Khairunizam, “Automatic classification of infant cry: A review,” Int. Conf. on Biomedical Engineering (ICoBE), pp. 543-548, 2012.

[17] I-A Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, “Baby Cry Recognition in Real-World Conditions,” Int. Conf. on Telecommunications and Signal Proc., Austria, pp. 315-318, 2016.

[18] I.-A. Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Automatic Methods for Infant Cry Classification," In Int. Conf. on Communications (COMM), Bucharest, Romania, 2016.

[19] R. Cohen, and L. Yizhar, "Infant cry analysis and detection," Conv. of Electrical & Electronics Engineers in Israel (IEEEI), pp. 1-5, 2012.

[20] B. Schuller, et al. “Deception, Sincerity & Native Language,” In Interspeech Computational Paralinguistics Challenge, pp. 2001-2005, September 2016.

[21] B. Schuller, "The Interspeech 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism," in Proceedings of Interspeech, 2013.

[22] E. Frank, M. A. Hall, and I. H. Witten, “The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, 2016.

[23] M.S. Rusu, S. Diaconescu, G. Sardescu, and E. Bratila, “Database and system design for data collection of crying related to infant’s needs and diseases,” Int. Conf. on Speech Technology and Human-Computer Dialogue SpeD 2015, Bucharest, Romania, 2015.

[24] F. Eyben, F. Weninger, F. Groß, and B. Schuller, “Recent developments in opensmile, the Munich open-source multimedia feature extractor,” ACM Int. Conf. on Multimedia, Spain, pp. 835–838, 2013.

[25] Y. Oshrat, A. Lerner, A. Cohen, and M. Avigal, “The Fingerprints of Pain in Human Voice,” Israel, 2014.