Literature Review on Infant Cry Emotion Recognition

1 Literature Review - Draft

Most informative articles so far:

- *Why is My Baby Crying? An In-Depth Analysis of Paralinguistic Features and Classical Machine Learning Algorithms for Baby Cry Classification*
  `https://ieeexplore.ieee.org/abstract/document/8441363`
- *Baby cry recognition in real-world conditions*
  `https://ieeexplore.ieee.org/document/7760887`
- *Analysis of Infant Cry Through Weighted Linear Prediction Cepstral Coefficients and Probabilistic Neural Network*
  `https://link.springer.com/article/10.1007/s10916-010-9591-z`
- *Baby Cry Detection: Deep Learning and Classical Approaches*
  `https://link.springer.com/chapter/10.1007/978-3-030-31764-5_7`

1.1 Dunstan Baby Language

Priscilla Dunstan (2012), in *Calm the Crying: The Secret Baby Language*, was one of the first to propose understanding babies' "speech" through reflex sounds before full crying. Her theory defines five universal pre-cry sounds, each associated with a specific need: hunger (Neh), burp (Eh), gas (Eairh), discomfort (Heh), and sleepiness (Owh). Most emotional baby cry classification studies rely on these five types and group them into emotional categories. Some choose to simplify into four.

We suggest adopting these five or four emotional states as the main classes, depending on the quality and type of data available.

1.2 The SPLANN Project

The SPLANN project (`https://www.dcae.pub.ro/en/proiecte/7/splann/`) is a Romanian initiative aimed at the development of automated methods for infant cry recognition. It includes the creation of an annotated database with over 13,000 cry samples recorded in hospitals. The database includes emotional categories such as colic, hunger, discomfort, pain, fatigue, and pathological crying, using rigorous labeling and recording standards.

1.3 Main Cry Databases

Most infant cry databases, including Dunstan and especially SPLANN, are not openly accessible and require direct communication with dataset creators.

SPLANN is the most robust and well-annotated, with approximately 13,373 samples recorded in Romanian hospitals. It includes categories such as colic, hunger, discomfort, pain, fatigue, and pathology.

Other databases such as **Baby Chillanto** (around 2,268 samples across five clinical classes) are also commonly cited and require formal requests for access.

Public alternatives, like the **Donate A Cry corpus** or the **Baby Cry Sense Dataset** available on GitHub or Kaggle, are accessible but come with less reliable labels.

1.4 Research Trends

The literature highlights two main research directions:

- **Cry detection vs. background noise:** Based on Google's crowd-sourced infant cry datasets, this task focuses on distinguishing baby cries from other ambient sounds. It is essential to include noise samples to train robust emotion classifiers.
- **Emotion classification:** Most studies aim to differentiate normal and pathological crying, or classify emotional needs (e.g., hunger, pain, sleepiness) according to the five Dunstan states. The papers listed above serve as strong references.

1.5 Feature Extraction

To classify baby cries emotionally, feature extraction is required. The article *Why is My Baby Crying?* describes the use of:

> The Munich open Speech and Music Interpretation by Large Space Extraction (openSMILE) tool enables extraction of large audio feature spaces in real-time. OpenSMILE allows for various configurations used for distinct purposes, such as speech emotion recognition or baby cry classification. It is the standard feature extraction tool for the ComParE challenge.

Popular configurations include ComParE (6,373 features) and emobase2010 (around 1,582 features), both widely used for speech emotion recognition and baby cry classification.

In SPLANN, Best Feature Selection (BFS) was applied to the ComParE set, reducing it to around 329–552 features. The most informative features were MFCC percentiles and fundamental frequency (F0).

1.6 Models and Classification Approaches

In the article *Why is My Baby Crying?*, the following classical models were applied to MFCC, F0, jitter, shimmer features:

- K-Nearest Neighbors (KNN)

- Support Vector Machines (SVM)
- Gaussian Mixture Models (GMM)
- Bayesian Networks

These methods achieved around 70% accuracy on small datasets like Dunstan or SPLANN.

More recent methods include CNNs, RNNs, and Graph Convolutional Networks (GCNs), with superior performance, even under low-data scenarios.

A recent study reports 96.39% accuracy in classifying five cry types (hunger, discomfort, abdominal pain, burp, fatigue) using MFCC and traditional ML methods: `https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1337356/full`

Current strategies also explore **self-supervised learning** (SimCLR, contrastive pre-training) and frameworks like **InfantCryNet**, which combine multi-head pooling, knowledge distillation, and quantization for efficient deployment on mobile devices.

1.7 Conclusions and Directions

A promising baseline is to combine CNNs with MFCCs, according to the reviewed literature.

As an innovation, we can incorporate uncertainty estimation, for instance using Bayesian learning.

It is critical to include noise data, along with emotional labels, to train robust models.

Dataset selection or collection should focus on: **audio + emotional label + noise**, enabling robust multi-emotional classification.