Computers in Biology and Medicine 167 (2023) 107626

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multi-modal analysis of infant cry types characterization: Acoustics, body language and brain signals

Ana Laguna [a,*,1], Sandra Pusil [a,1], Àngel Bazán [a], Jonathan Adrián Zegarra-Valdivia [b,c,d], Anna Lucia Paltrinieri [e], Paolo Piras [a], Clàudia Palomares i Perera [e], Alexandra Pardos Véglia [f], Oscar Garcia-Algar [e,g], Silvia Orlandi [h,i]

[a] Zoundream AG, Switzerland
[b] Global Brain Health Institute, University of California, San Francisco, CA, USA
[c] Achucarro Basque Center for Neuroscience, Leioa, Spain
[d] Universidad Señor de Sipán, Chiclayo, Peru
[e] Neonatology Unit, Hospital Clínic-Maternitat, ICGON, BCNatal, 08028, Barcelona, Spain
[f] Centro de Neuropsicología Alexandra Pardos, Madrid, Spain
[g] Department de Cirurgia I Especialitats Mèdico-quirúrgiques, Universitat de Barcelona, 08036, Barcelona, Spain
[h] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi"(DEI), University of Bologna, Bologna, Italy
[i] Health Sciences and Technologies Interdepartmental Center for Industrial Research (CIRI-SDV), University of Bologna, Bologna, Italy

ARTICLE INFO

Keywords:
Cry acoustics
EEG
NIRS
Body language
Newborns

ABSTRACT

Background: Infant crying is the first attempt babies use to communicate during their initial months of life. A misunderstanding of the cry message can compromise infant care and future neurodevelopmental process.
Methods: An exploratory study collecting multimodal data (i.e., crying, electroencephalography (EEG), near-infrared spectroscopy (NIRS), facial expressions, and body movements) from 38 healthy full-term newborns was conducted. Cry types were defined based on different conditions (i.e., hunger, sleepiness, fussiness, need to burp, and distress). Statistical analysis, Machine Learning (ML), and Deep Learning (DL) techniques were used to identify relevant features for cry type classification and to evaluate a robust DL algorithm named Acoustic MultiStage Interpreter (AMSI).
Results: Significant differences were found across cry types based on acoustics, EEG, NIRS, facial expressions, and body movements. Acoustics and body language were identified as the most relevant ML features to support the cause of crying. The DL AMSI algorithm achieved an accuracy rate of 92%.
Conclusions: This study set a precedent for cry analysis research by highlighting the complexity of newborn cry expression and strengthening the potential use of infant cry analysis as an objective, reliable, accessible, and non-invasive tool for cry interpretation, improving the infant-parent relationship and ensuring family well-being.

1. Introduction

Infant crying is the first proper attempt humans use to communicate during their initial months of life [1–3]. As a result, parents tend to interpret an infant's crying as a signal of alertness or need. However, new caregivers often feel confused about what crying means, and they may be unable to soothe the baby [4] which can result in a variety of mixed feelings [5]. Considering infants cry on average between 1.5 and 3h per day [4], the impact of infant crying on parents can vary between experiences of anxiety, depression, helplessness, anger, and frustration in response to infant crying, negatively affecting bonding and infant's parental perception [1,6]. Parents may even experience violent thoughts toward their newborn and later feelings of guilt and shame [4]. Considering parental response is crucial for the new dyadic relationship, deviations and/or misunderstanding of the cry message can compromise infant care and their future neurodevelopmental process [1]. Hence, it is important to find a meaning in the early cries to ensure the infants' well-being and health status.

Research [4] on sound spectrographic features of cry has been conducted since the 1960s. Further analysis [4] investigated how cry

* Corresponding author. Zoundream AG, Novartis Campus - SIP Basel Area AG, Lichtstrasse 35, 4056, Basel, Switzerland.
E-mail address: ana.laguna@zoundream.com (A. Laguna).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.compbiomed.2023.107626
Received 4 July 2023; Received in revised form 14 September 2023; Accepted 23 October 2023
Available online 25 October 2023
0010-4825/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

duration and intensity may vary across contexts. Both time and frequency domain approaches led to the conclusion that cry sounds convey a level of distress or urgency of the newborn's need that offers some clues as to the specific cause of crying [4]. Moreover, coupled with contextual information, the sounds of crying may be highly informative, facilitating an accurate (even though complex) interpretation for the caregiver or healthcare professional [4].

In recent decades, the field of automated analysis of infant cry has advanced significantly, attracting the attention of an increasing number of researchers, clinicians, and computer scientists [7–9]. Moreover, the advent of Deep Learning (DL) techniques has facilitated the development of more reliable and precise Artificial Intelligence algorithms for analyzing infant cries [10–17].

While significant advances have been achieved in infant cry research, the interpretation of infant cries still remains partially unanswered and subjective. There is an ongoing debate in the field regarding the detection and classification of different cry types. While some studies have shown promising results, others remain inconclusive [12–16]. Therefore, it is evident that future research should focus on multimodal data collection to understand infant cries by concurrently assessing diverse newborn measures. The parallel infant cry analysis coupled with neurophysiological measures as near-infrared spectroscopy (NIRS), electroencephalography (EEG) and simultaneously behavioral signals (extracted from facial expressions and body movements using video recordings), arises as a promising pioneer approach to provide scientific and objective evidence to support and validate literature on cry acoustics and its analysis [4].

The primary aim of this study is to characterize different newborns' cry sounds based on acoustic features, neurophysiological and behavioral signals. The secondary goal is to determine the most relevant features that allow the distinction of different crying reasons based on Machine Learning (ML) within a multimodal dataset. Lastly, we aim to demonstrate that DL approaches, such as the one presented in this paper (i.e. the Acoustic MultiStage Interpreter (AMSI [18])), developed by our research group, are automatic, effective and reliable tools for interpreting infant cries and assessing the well-being of newborns.

Therefore, we hypothesize that what is occurring at an acoustic level can be reinforced with behavioral and brain neurophysiological pattern analysis. Moreover, by analyzing sound features of infant cries, AMSI [18] is able to identify valuable acoustic biomarkers that are indicative of a newborn's wellness status. Through this multimodal analysis approach, we gain a better comprehension and a more accurate interpretation of the complex phenomenon of infant crying. This could have significant implications for bonding and also implying important outcomes for newborns' healthcare based on a reliable understanding of the first human communicative attempt.

2. Methods

2.1. Participants

The study included 38 healthy full-term newborns that were part of a previous dataset [19], and recruited at the maternity ward of the Hospital Clínic of Barcelona (mean gestational week 39.34 $\pm$ 1.33, recording age 15.54 $\pm$ 21.46 days after birth, 21 males/17 females, head circumference 34.41 $\pm$ 1.21 cm, birth weight 3132.91 $\pm$ 404.29 g). APGAR scores were collected at 1, 5 and 10 min: 8.86 $\pm$ 0.713, 9.83 $\pm$ 0.56 and 9.965 $\pm$ 0.185 respectively. Infants had been assessed by board-certified neonatologists and diagnosed as healthy term newborns with no major congenital abnormalities or illness since birth. Newborns under medication and/or with congenital malformations, chromosomal abnormalities, hypoxic-ischemic encephalopathy, intraventricular hemorrhage greater than grade 2, and any other type of brain damage, congenital heart disease, siblings with autism spectrum disorders or other neurodevelopmental disorder were excluded from this study.

2.2. Ethical considerations

The study was conducted following the Institutional Research Ethics and the Declaration of Helsinki. Formal ethical approval was granted by the Local Ethical Committee, Hospital Clínic of Barcelona (Ref: NeuroCry/HCB/2021/0843). The consent form documents the study's aims, nature, and data acquisition procedures. Anonymization and data confidentiality was maintained throughout the study, and all parents agreed and signed the informed consent prior to participation.

2.3. Procedure

Data collection was performed during the standard routine of newborn nursing (before and post feeding, during some necessary medical procedures, etc.). Synchronized EEG, NIRS, audio, and video recordings were acquired for each newborn, lying down comfortably in a cot in the hospital maternity ward. Each session lasted from 20 to 120 min when newborns were calm-awake or crying. Cries were never induced for the purpose of this study, as spontaneous vocalizations are part of normal infant behavior. During the recording sessions, different cry types were defined as changes in the newborn's status generated by different scenarios (i.e., hunger, sleepiness, fussiness, need to burp, stress, pain, etc.), yielding in the following cry conditions: *resting*, *hungry*, *sleepy*, *fussy*, *burp,* and *distress*.

2.4. Audio analysis

Data acquisition. Audio acquisition and processing analysis were described in Ref. [19]. Briefly, newborn cries were recorded using a ZOOM H1N™ recorder (with a unilateral microphone) positioned 30 cm from the infant's mouth. Audio recordings were stored on a Waveform Audio Format (WAV) file with a double channel audio track, with a sampling frequency of 48 kHz and 24-bit resolution.

Data processing.

Segmentation. Audio recordings were manually segmented into cry episodes (CEs – the amount of time the infant cries in each audio recording divided by silence periods). Then, CEs were manually segmented into cry units (CUs - individual cry patterns within a CE separated by an expiration phase). Visual spectrographic analysis was carried out using iZotope RX 7 Audio Editor™. Both the segmentation and the qualitative assessment of every CEs and CUs into different cry types have been carefully reviewed by two authors (AL, PP) through visual inspection of the acoustic features following what has been done in previous research articles [9,20,21]. An additional author (SO) randomly reviewed annotated CUs achieving a level of agreement of 100% on the cry annotation. Cries without unanimous agreement among the experts were excluded from further analyses. Afterwards, newborns' states or behaviors were identified in every CE using the following criteria [4,22]:

1. *resting*: no CEs, pause, or resting periods with silent audio recordings, the newborn is not crying but awake.
2. *hungry*: pre-feeding arousal state acoustically characterized by short, rhythmic, symmetrical, harmonic structured and intensity sound patterns with pitch range of 200–600Hz.
3. *sleepy*: post-feeding state, when the infant starts feeling tired characterized by softer, quieter sounds with a falling melody prolonged in duration, showing smaller amplitude or variability.
4. *fussy*: baby unsettled, low level intermittent protest cries similar to whimper or whine, not reaching the point of full-blown crying and separated by periods of silence.
5. *burp*: whining or noise during or post-feeding phases characterized by a sound of effort related to the need to expel something out of the body.
6. *distress*: stressed infant showing more acoustically urgent and intense CEs composed of high spectral content CUs, disphonation, increased

fundamental frequency (F0), and instability (greater amplitude, etc.).

Feature extraction. Audio processing of each CU was conducted through Praat software [23] using a band-pass filter between 200 and 1200 Hz to compute the F0 and a low-pass filter of 10,000 Hz to compute the spectrum [24]. We then computed the F0 descriptive statistics (maximum, minimum, mean, and standard deviation defined as maxF0, minF0, F0 and stdF0 respectively). Other typical voice quality parameters related to the phonation were also included, such as jitter, shimmer, and harmonic-to-noise ratio (hnr) [25]. Basic crying melody shapes, the trend of F0 over time (F0 melody), were defined as falling, rising, symmetric, plateau, and complex using 1st and 2nd derivatives based on F0 [26]. CU durations were also extracted and analyzed. Each feature was extracted for each CU and averaged in each cry type (i.e. *hungry*, *sleepy*, *fussy*, *burp,* and *distressed*).

2.5. EEG analysis

Data acquisition. Neurophysiological data were acquired using an 8-channel ANT Nëo Monitor eego™ (ANT Neuro, Germany – CE mark MDD 93/42/EEC, CE class Iia, FDA 510(k) in USA) at a sampling frequency of 512 Hz. Electrodes were placed according to the 10–20 international standard positioning system and were re-referenced to the average reference. Sensors' impedance was kept below 10 kΩ. Electrodes were fixed at key locations, including F3, F4, C3, C4, P3, P4, T7, and T8, ensuring consistency and compatibility with widely accepted conventions.

Data processing. The dataset was analyzed using Matlab r2022a with the Brainstorm Toolbox [27]. To remove the power line contamination and low frequency artifacts, a band-pass filter between 1 and 45 Hz was applied. A careful visual inspection of EEG data was performed to detect artifacts produced by newborn movements and/or the hardware. After that, bad channels were identified and interpolated using spherical splines [28]. The remaining artifact-free data were segmented into 4-s epochs according to the audio segmentation criteria mentioned before.

EEG data analysis was performed for the following classical frequency bands: delta (δ = 1–4Hz), theta (θ = 4–8Hz), and alpha (α = 8–12Hz). To avoid contamination from muscle activity, higher frequencies from the beta to gamma range were not included in the analysis. Additionally, the normalized power spectrum density (PSD) of each EEG sensor was computed using Welch's periodogram method with the Neural Toolbox [29].

2.6. NIRS analysis

Data acquisition. NIRS data acquisition was performed using a Root O3™ (Masimo, USA - CE mark G1 092076 0013 Rev. 00). This device collects data every 2 s using 4 wavelengths (730/760/805/880 nm) with 70% representing venous blood and 30% arterial blood. Root O3™ provides a regional hemoglobin oxygen saturation (rSO2) ranging from 0 to 100%. Due to the limited available space on the newborn's head, we used a single NIRS forehead sensor to enable measuring rSO2. Functional arterial hemoglobin oxygen saturation (SpO2) is continuously and non-invasively monitored with a fingertip sensor on the newborn.

Data processing. rSO2 and SpO2 data were analyzed in Python 3. A preprocessing step was applied to eliminate rSO2 and SpO2 values with a standard deviation lower than 0.5 to reduce errors from the acquisition process. Additionally, the interquartile range (1.5*IQR) method was used to remove outliers. Then, clean data were segmented into *resting, hungry, sleepy, fussy, burp, and distress* based on the audio segmentation criteria. The 15 s preceding and following each segment were discarded [30]. In addition, to reduce noise and errors derived from newborn's movements, SpO2 mean values lower than 80 [31] and rSO2 lower than 50 [32] were removed.

2.7. Facial expression & body movement analysis

The COMFORT scale [33,34] was used to qualitatively evaluate the high-quality video recordings acquired of facial expressions and body movements of the newborns during each session and CE. Two experts reviewed and assessed the newborns individually according to the COMFORT scale for each CE on the video. In case of disagreement between the experts, a third reviewer was asked to present their evaluation. The aspects evaluated include six sections: alertness, agitation, crying, body movements, muscular tone, and facial tension. Each section can be rated from 1 (calm infant) to 5 (stressed infant), and the total score of each CE ranging from 5 to 30, with larger score values indicating a higher arousal threshold.

2.8. Statistical analysis

Statistical analysis was performed using Matlab r2022a (MathWorks, USA - mathworks.com) and SPSS22 (IBM, USA - https://www.ibm.com/es-es/products/spss-statistics). The Shapiro-Wilk test was applied to verify that data were not normally distributed. Also, due to the nature of the data collection, which consisted of spontaneous cry recordings during the newborn's daily routine, the segments of the six different conditions were not balanced. As such, we randomly selected a representative number of segments for each signal feature (audio, EEG, NIRS); after that, an outlier removal (interquartile range) procedure was applied to each variable. Thus, balance audio data, EEG, NIRS, and the COMFORT scale were processed to be compared between all conditions *(resting, hungry, sleepy, fussy, burp, and distress)* using ANCOVA analysis; age was included in our model as a covariate. Dunn-Sidak tests were used for post hoc comparisons, plus a bootstrapping procedure repeated 1000 times to correct for normality. Results are reported in terms of mean ± standard error mean and statistically significant p-values are coded as follows: ▲ $p<=0.001$, ■ $p < 0.01$ and ✱$p < 0.05$.

2.9. Machine Learning and Deep Learning analysis

Two approaches have been implemented to classify cry types. The first approach involved training different well-known ML algorithms (AdaBoost [35], Random Forest [36], and Logistic Regression [37]) based on the multimodal dataset. The following are the specific features computed for each data signal: cry acoustics (duration, F0, minF, maxF0, stdF0, hnr, jitter, shimmer, F0 melody), EEG (PSD for delta, theta and alpha bands for each electrode position), NIRS (rSO2 and SpO2) and COMFORT scale scores (alertness, agitation, crying, body movements, muscular tone, and facial tension). They were introduced in the models to classify five classes (i.e., *hungry, sleepy, fussy, burp, distress*) within a total of 156 CEs. 70% of the data was used to train the models and 30% for validation.

The comparative analysis of the outcomes obtained from the three ML models allows us to determine the most relevant set of features for cry classification. In particular, we identified three distinct feature subsets: 1) a dataset with the audio features; 2) a dataset without audio features (NIRS, EEG, and COMFORT scale); 3) the whole multimodal dataset (audio cry, NIRS, EEG, and COMFORT scale).

Finally, for visualization purposes, a 2D graph of the CE clusters was accomplished with the most relevant features using the dimensionality reduction t-distributed Stochastic Neighbor Embedding (t-SNE) [38] model.

The second approach used the images of the spectrogram of each CU as input after applying the Fourier transform to the raw audio signal. We used a realistic cry dataset composed of 5002 CUs from the 156 CEs as test data to evaluate the applicability of AMSI [18], a pre-trained Convolutional Neural Network (CNN) [39] method developed by our research team. AMSI [18] breaks down audio recordings into a series of internal stages to extract the meaning of a baby's cry from cepstral features. The output is the multiclass classification of the CE within five possible labels: *hungry, sleepy, fussy, burp,* and *distress*. Further information regarding the algorithm can be found under patent registration [18].

3. Results

3.1. Acoustic features on different cry types

We found statistically significant differences between the different cry types among the audio features. Fig. 1A shows the differences across cry types in terms of duration of CUs. The ANCOVA analysis on duration indicated a significant effect of the cry type ($F = 14.095$, $p < 0.001$), an effect of the age ($F = 10.14$, $p = 0.002$), and an interaction effect ($F = 2.603$, $p = 0.035$). The duration of crying events in the *sleepy*, *fussy*, *burp*, and *distress* CU categories were found to be longer compared to those in the *hungry* category, which displayed lower values. Also, Fig. 1B shows F0 changes through the different cry categories. A significant effect of the cry type ($F = 15.354$, $p < 0.001$) was found, along with an interaction effect ($F = 9.167$, $p < 0.001$) and no effect of age ($F = 0.851$, $p = 0.357$). Higher mean F0 and minF0 values can be observed for *hungry* and *fussy* while *distress, burp,* and *sleepy* exhibited lower values. Moreover, higher values of maxF0 and stdF0 can be observed for *burp* and *distress* and lower values for *hungry*, *sleepy,* and *fussy*.

Additionally, Fig. 1C presents changes in hnr for each cry type. We found a statistically significant effect of the cry type ($F = 47.435$, $p < 0.001$), no effect of the age ($F = 1.586$, $p = 0.209$) and an interaction effect ($F = 9.507$, $p < 0.001$). *Hungry*, *sleepy* and *fussy* presented higher values of hnr compared to *burp* and *distress*. For jitter *(Fig. 1D)* and shimmer *(Fig. 1E)*, we found a significant effect of the cry type ($F = 59.063$, $p < 0.001$ and $F = 87.144$, $p < 0.001$), an effect of the age ($F = 19.670$, $p < 0.001$ and $F = 9.595$, $p = 0.002$) and an interaction effect ($F = 5.090$, $p < 0.001$ and $F = 17.119$, $p < 0.001$). Also, lower values of jitter and shimmer were observed for *hungry, sleepy* and *fussy* compared to *burp* (where shimmer showed the highest values) and *distress*.

Additionally, we analyzed different F0 melody shapes for the different cry types. The ANCOVA analysis revealed that there was no effect of the cry type on the F0 melody ($F = 2.223$, $p = 0.066$), a significant effect of the age ($F = 8.386$, $p = 0.004$) and no interaction effect ($F = 1.031$, $p = 0.391$). Fig. 1F shows the distribution of the computed melodies. We observed a more falling trend for *fussy* and *sleepy* and *distress*, complex/symmetric for *hungry* and *burp*, which also showed a rising tendency.

3.2. Neurophysiological patterns for cry types

The PSD analysis unveiled different topological patterns for all cry conditions and frequency bands (Fig. 2). In delta band power (Fig. 2Aδ), the *hungry* category showed a higher left fronto-parietal PSD distribution, while *sleepy* and *fussy* presented only a higher power distribution in the left frontal region. Furthermore, the *burp* category showed a higher PSD distribution in the right fronto-temporal region. Lastly, *distress* displayed a similar PSD distribution all over the scalp for each frequency band.

In the theta band PSD of *hungry* (Fig. 2Aθ), we found a higher PSD distribution in the right fronto-parietal region. Also, *sleepy* and *fussy* showed a higher PSD in the centro-posterior area for theta band, while *burp* exhibited a higher PSD distribution in the left centro-temporal region. *Distress* reflected a similar PSD distribution all over the scalp.

In the alpha band (Fig. 2Aα), the *hungry* category depicted a higher right fronto-parietal PSD distribution, while *sleepy* and *fussy* showed a higher centro-posterior PSD distribution like in the θ band. *Burp* displayed activation of the left hemisphere, and *distress* showed a similar pattern all over the scalp.

Lastly, the PSD distribution of the *resting* condition reflected higher values in the left parietal for the delta band power, in the fronto-central for theta band power, and in the right fronto-central region for the alpha band power.

On the other hand, Fig. 2B showed the statistically significant differences in PSD for each condition, electrode, and frequency band. In the delta band power (Fig. 2Bδ), we found that most of the cry types' PSD values decreased compared to the *resting,* except for the *burp* whose PSD values increased. In theta band (Fig. 2Bθ), when compared to *resting,*

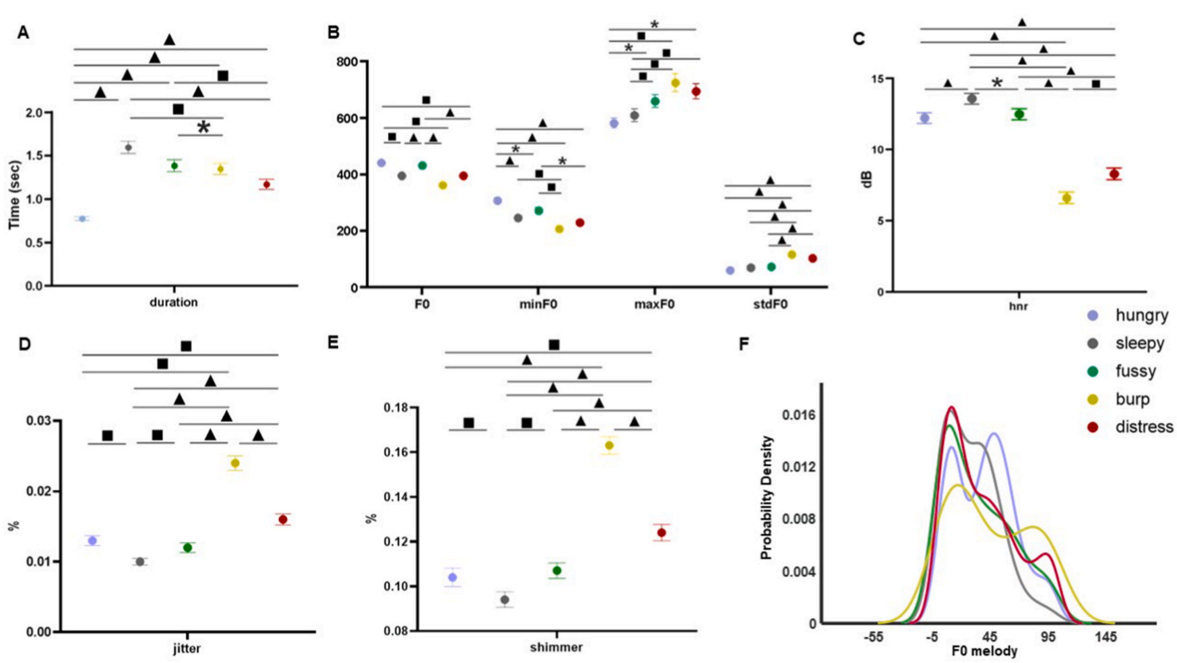

Fig. 1. Differences for each cry type on audio features. ANCOVA testing the effect of cry type and age and Dunn-Sidak were used for pairwise comparisons of the different cry types: *hungry* (n = 102 segments), *sleepy* (n = 111 segments), *fussy* (n = 101 segments), *burp* (n = 95 segments) and *distress* (n = 104 segments), n was balanced using random sampling. The figures show the different cry features studied: A. duration. B. F0 and its descriptive statistics. C. hnr. D. jitter. E. shimmer. F. F0 melody. Data are presented as the mean ± standard error mean and statistically significant p-values are coded as follows:▲ p<=0.001, ■ p < 0.01 and ✱ p < 0.05.

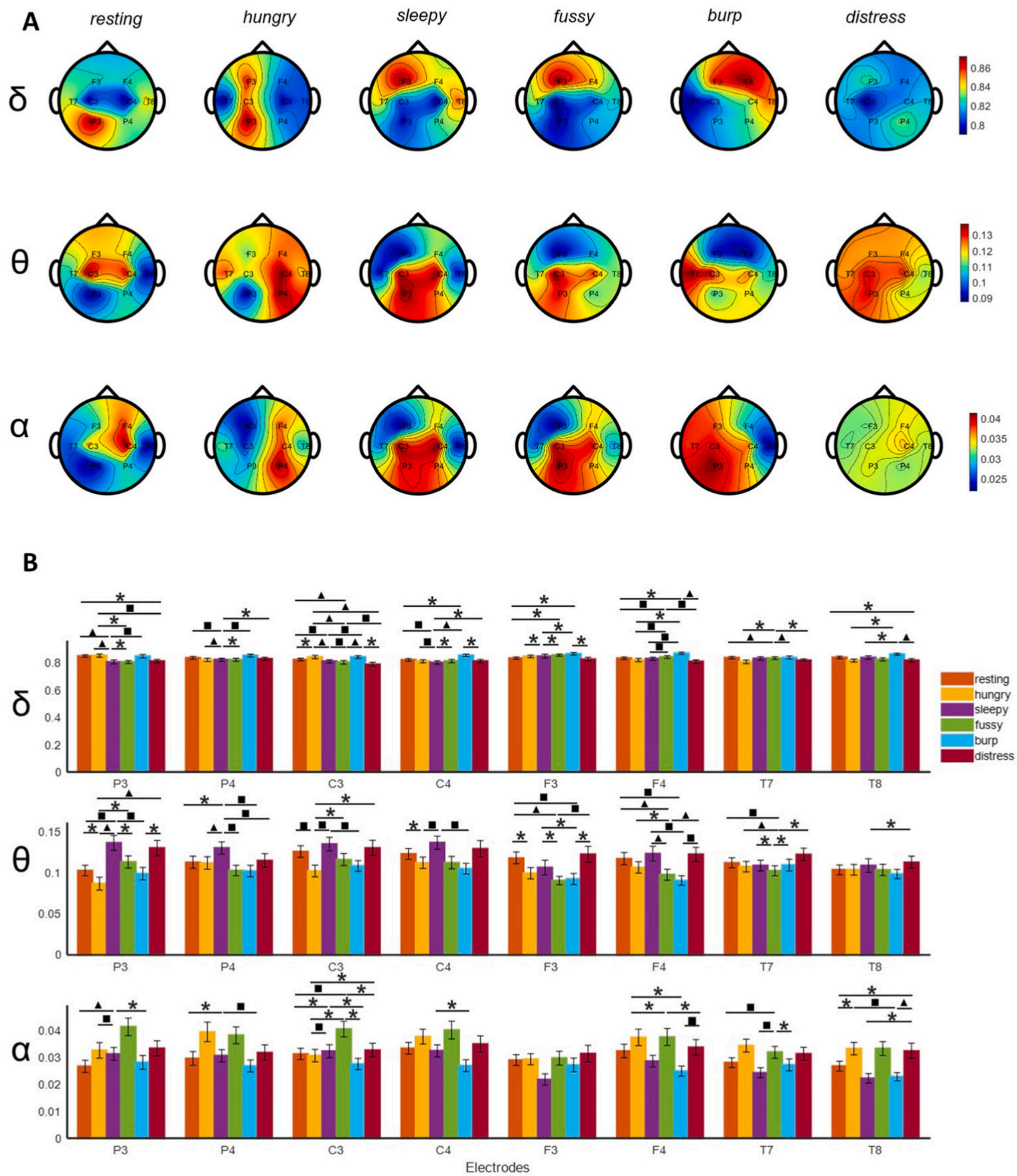

Fig. 2. A. Topographic EEG maps of PSD distributions for δ, θ, and α bands for *resting, hungry, fussy, burp and distress*. The upper portion of each map shows the nose (frontal area), and the lower side shows the occipital side. B. Differences in PSD for *resting* (n = 87), *hungry* (n = 79), *sleepy* (n = 81), *fussy* (n = 83), *burp* (n = 88), and *distress* (n = 75) conditions were obtained by applying an ANCOVA test with age as a covariate and a Dunn-Sidak test (for post-hoc comparisons) plus a bootstrapping procedure. Data are presented as the mean ± standard error mean and statistically significant p-values are coded as follows: ▲ p<=0.001, ■ p < 0.01 and ✱ p < 0.05.

sleepy and *distress* PSD values increased. *Hungry*, *fussy* and *burp* power values decrease compared to the *resting* condition. Finally, for the alpha band (Fig. 2Bα), PSD values for *hungry*, *fussy* and *distress* increased compared to *resting*.

3.3. Brain oxygenation changes during different cry types

We analyzed the differences among cry types for the SpO2 and rSO2 features. The ANCOVA analysis revealed a significant effect of the cry type on SpO2 ($F = 13.735$, $p < 0.001$) and rSO2 ($F = 104.93$, $p < 0.001$), an effect of the age on SpO2 ($F = 6.469$, $p = 0.011$) and an interaction effect on rSO2 and SpO2 ($F = 8.356$, $p < 0.001$; $F = 68.142$, $p < 0.001$). Fig. 3A displays the statistically significant differences for pairwise comparisons between the different cry types. Compared to *resting*, SpO2 and rSO2 values decreased for all the cry types. For rSO2 values, *hungry* presented lower regional oxygenation values than the rest of the categories. For SpO2, *burp* exhibited lower oxygen values compared to the other cry types.

3.4. Behavioral assessment of cry types

Fig. 3B shows the differences between all items within the behavioral

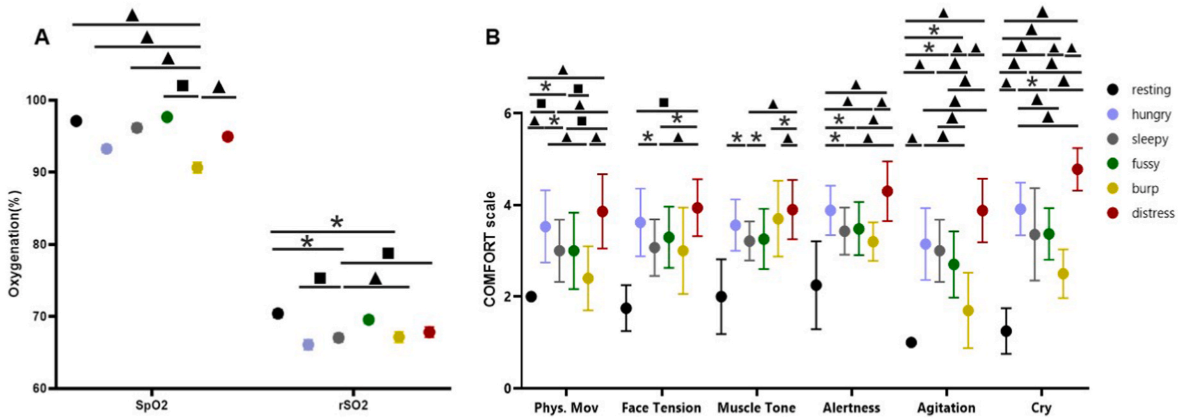

Fig. 3. A. ANCOVA testing the effect of cry type and age on NIRS features: SpO2 and rSO2, and pairwise comparisons (Dunn-Sidak test) of cry types: *resting, hungry, sleepy, fussy, burp and distress* (n = 121 segments). B. Comparisons of the COMFORT scale scores among cry types (*resting*: n = 4 segments, *hungry*: n = 46 segments, *sleepy*: n = 17, *fussy*: n = 32, *burp*: n = 12 and *distress*: n = 57). Physical Movements, Facial Tension, Muscular Tone, Alertness, Agitation, and Cry scores are reported. An ANCOVA test with age as a covariate, Dunn-Sidak test (for post-hoc comparisons), and bootstrapping procedure were used. Data are presented as mean ± standard error mean and statistically significant p-values are coded as follows:▲ p<=0.001, ■ p < 0.01 and ✱ p < 0.05.

assessment scored with the COMFORT scale for all cry categories. Compared to the *resting* condition, higher scores were found for the cry types on every item of the COMFORT scale. In general, the *distress* category presented higher scores in all items compared to the other cry types. The ANCOVA analysis revealed a significant effect of the cry type on Physical Mov (F = 6.304, p < 0.001), Face Tension (F = 6.145, p < 0.001), Muscle Tone (F = 6.357, p < 0.001), Alertness (F = 11.126, p < 0.001), Agitation (F = 17.496, p < 0.001) and Cry (F = 26.040, p < 0.001). There was no significant statistical difference in age and interaction on all the variables of the COMFORT scale.

3.5. *Relevance of audio features in the classification of cry types*

The examination of the feature importance assigned by each ML model applied to distinguish cry types based on a multimodal dataset is essential for the interpretation of our results. We found that the most relevant features for cry interpretation within the three ML models using audio, EEG, NIRS and COMFORT scale features correspond to the following audio features: duration, hnr, jitter, shimmer, minF0, mean F0, F0 melody (mean test accuracy for all models of 71.5% and the area under the curve (AUC) 90.4% (more details on the classification performance metrics per model including specificity, precision, recall and f1-score are reported in Table 1SA and Table 1SB of the Supplementary Material). Some other features related to body language from the COMFORT scale, such as facial tension and crying were also found significant. In Fig. 4 we can observe the clusters, identified by feeding a t-SNE model with the features mentioned above.

Regarding the comparison of the different subsets of features, we found that when including audio features within the dataset all the evaluation metrics improved. In particular, the mean accuracy of the ML models increased by around 13%, AUC 6%, specificity 2%, precision 22%, and recall and f1-score 19%. Specifically, when only audio features were used, an accuracy of 74% and AUC of 90.3% were found. When audio features were excluded, the classification accuracy decreased to 61% and AUC to 84% (more details on the classification performance metrics per model and dataset including specificity, precision, recall and f1-score are reported in Table 1SA and Table 1SB in the Supplementary Material).

Lastly, when comparing the ML to the DL method, the mean classification accuracy achieved by the DL model AMSI [18] was 92.0%. These results can be translated into a 18% increase in test accuracy using cepstral features and CNNs. For more details on the classification performance refer to Table 1SA and Table 1SB in the Supplementary Material.

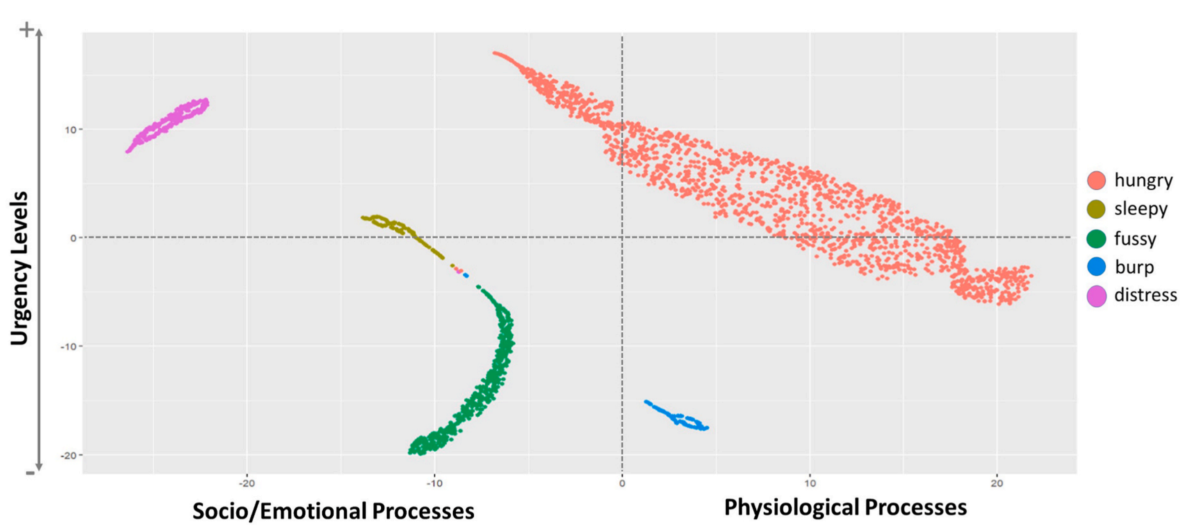

Fig. 4. A. Visualization of the cry type clusters found using a t-SNE model.

4. Discussion

Limited quantitative research has been conducted to understand infant cry as a communication response and complex neurophysiological and behavioral functions [17]. To the best of our knowledge, there is no literature regarding multimodal data collection concurrently analyzing infant cry acoustics with reliable and objective measures.

Thence, this study represents a pioneering analysis of infant cry in combination with neurophysiological and behavioral signals, comparing acoustic cry features (e.g., F0, jitter, shimmer, F0 melody) to EEG power spectrum, NIRS regional hemoglobin oxygenation, physiological features (e.g., body oxygen saturation), body movement features (e.g., muscle tone), facial expression characteristics (e.g., facial tension) during different newborn needs (i.e., *resting*, *hungry*, *sleepy*, *fussy*, *burp,* and *distress*).

Our primary findings showed how every condition is characterized by different acoustics, neurophysiological and behavioral patterns. In the case of *hungry*, a condition that represents a survival physiological need ensuring nutrition and hydration, our acoustic patterns matched the existing literature [4,40] and described this condition as urgent, constant, rhythmic and short in duration, intense/loud (mean F0) but not high-pitched (maxF0). We also found a prevalence of rising/falling or symmetric shapes for melody patterns of *hungry*. Regarding oxygenation, our results showed that *hungry* had a direct and significant negative impact on newborn oxygenation (central and body) also observed in a similar study evaluating prior, during and after feeding states in newborns [41]. As for brain activity, *hungry* showed a decrease in theta band which is reflected in the literature when analyzing infants' EEG activity before feeding [42]. Additionally, we found an increase in alpha PSD distribution related to feeding cries, which has been linked to hunger in EEG in adult studies [43]. This PSD increase in the alpha band in newborns is usually associated with a relaxed but awake state of mind and possibly reflects a state of heightened alertness or arousal as the infant seeks nourishment [44,45]. This restlessness tendency was also observable in the results from the COMFORT scale, where physical movements, facial expressions, muscle tone, and agitation are highest compared to the rest of the conditions (except *distress*) indicating a state prior to feeding that reflects large-amplitude and jerky movements at the age of 2 and 10 weeks [42]. Overall, *hungry* in newborns can cause a variety of facial expressions and body movements aimed at getting the attention of caregivers and indicating a need for food, among them: rooting reflex, sucking motions, fidgeting or squirming, or arching the back.

The results of the *distress* category matched the conclusions found in Ref. [19]. This condition represents the state of an infant suffering physically (e.g., pain) or emotionally (e.g., overstimulation) and from an acoustic perspective, our results showed that distressed cries were more erratic, with fewer pauses, prolonged in duration and high-pitched (maxF0). Lower hnr, higher jitter, and higher shimmer values compared to non-distressed cries reflect more noise, less harmonic structure, and increased variation in frequency and amplitude. All these acoustics measurements indicate a loss of stability in the vocal folds caused by a stressful situation similar to literature in adults [46]. Regarding oxygenation, *distress* in newborns presented a less direct impact on oxygenation compared to other bottom-up mechanisms such as *hungry* and *burp*, but higher than other more central mechanisms such as *sleepy* and *fussy*. While distress can trigger a physiological response that affects heart rate [19], newborns' distress responses are still developing and may not be as pronounced as in older children or adults [47]. However, according to our results, an infant in *distress* reflected significant changes in the brain in theta and alpha PSD [19], indicating increased arousal. These findings reinforce the existing literature suggesting that EEG alpha activity may be a useful measure of stress in newborns [48]. The restlessness tendency was also clear in the COMFORT scale assessment, where the results for this condition presented the highest values pointing out to newborn's discomfort and emotional distress highlighting increased motor activity as an expression of a high discomfort level or emotional distress expressed through the body [49] characterized by common facial expressions and body movements (i.e., facial grimaces, flailing limbs, arching the back, or clenched fists).

The *burp* condition reflects a gastrointestinal physiological need to be expelled, even in the form of air or solid. This condition presented common characteristics to the audio frequency features found in *distress* (even though the spectral patterns are completely different), due to the strain put on the vocal folds leading to more hoarseness and other vocal quality problems [50]. According to our results in the *burp* condition, infants with gastrointestinal issues produced longer cries (compared to *hungry* and *distress*) with more variable pitch and intensity reflecting vocal folds instability. These cries presented lower hnr, meaning that the cry may have more noise components and fewer harmonics. Additionally, *burp* exhibited higher jitter and shimmer values, which indicate more irregularities in the frequency and amplitude of the cry signal. Compared to *distress*, cries were less continuous and more sparse in time with more unvoiced segments. Melody was characterized by more rising shapes, putting most of the strength at the end, like expressing the act of expelling something. Oxygenation levels analyzed for our *burp* condition showed the lowest values compared to the other conditions, eliciting that gastrointestinal issues such as reflux could interfere with proper breathing and decreased oxygen saturation levels [51]. From a brain activity perspective, our results showed a similar tendency to the *resting* condition on delta power and less activity in theta or alpha bands as gastrointestinal actions like bowel movements involve physical activity, such as abdominal muscle contractions but less brain activity levels. This tendency was confirmed by the COMFORT scale results, where *burp* values were also the closest to the *resting* condition showing lower alertness levels or body movements, but more tension or muscle tone presence due to body muscles contraction or facial tension as an unconscious response to the effort required to expel something out of the body [52]. Generally, this condition looks more like whining related to a physiological complaint than an actual urgent cry, where the arousal state is not very high. This low alertness can be linked to the fact that *burp* usually happens after a feed, which is when newborns usually tend to relax.

The *sleepy* condition represents a fatigued infant not able to fall asleep. From an acoustic perspective, we found the longest patterns in duration, with prolonged monotonous cries presenting a clear falling melody. These cries also exhibited the highest hnr, meaning that the cry has fewer noise components and more harmonics, in combination with lower jitter and shimmer values, indicating fewer irregularities in the frequency and amplitude of the cry signal indicating reduced arousal [53]. Body oxygenation levels were not highly impacted, similar to the *resting* condition. However, our results for brain oxygenation showed alterations when compared to *resting*, indicating that the need to sleep may be linked to reduced cortical oxygenation [54]. Furthermore, we found an increase in theta band related to sleepiness in newborns, which is associated with drowsiness, alertness reduction, and sleep induction found in similar studies in children [55,56] or adults [57,58]. Body language also reinforces the baby's calmness in the *sleepy* and *fussy* categories showing lower scores on the COMFORT scale on muscle tone, indicating a more relaxed state.

The *fussy* condition is a catch-all category whose interpretation from the caregivers will really depend on the contextual information available. Even if most of the time is related to the demand for attention or contact, it could also be the cause of a diaper change, uncomfortable position, temperature regulation, etc. Acoustically, *fussy* has similar characteristics in duration and F0 melody to *sleepy* (as a tired infant also presents fussiness), but sounds more like whining than an actual cry [4]. F0 for *fussy* is closer to *hungry* and its melody is flat or falling with more unvoiced segments. Regarding oxygenation, *fussy* is the closest category to the *resting* condition, highlighting an awake but calm state. However, the increase found in alpha power reinforces the alert level, looking for attention, contact and/or comfort [44,45,48].

In general, both *fussy* and *sleepy* conditions showed very similar acoustics, body language and oxygenation characteristics also present in similar topographic activations in the brain signals. The big difference is that *fussy* EEG activity showed more alertness reflected through alpha increased compared to *sleepy* which indicates more drowsiness presented in theta bands. Even though none of those conditions seems extremely urgent (compared to *hungry or distress*), they express an intent to communicate certain physiological needs or emotions very important to be covered for the comfort, well-being, and proper development of the newborn.

Overall, our results demonstrated our primary objective showing that there are statistically significant differences among cry acoustics, EEG, NIRS, facial expression, and body movements features characterizing the different causes of crying in a newborn (*hungry*, *sleepy*, *fussy*, *burp,* and *distress*).

Regarding our secondary aim, the pioneer multimodal data analysis based on ML highlighted that features involved in communication (body language and voice) are the most relevant ones to express the cause of cry, matching also with adults' communicative approach [59]. As shown in a recent literature review [16], several articles tried to identify five cry types using ML approaches using only acoustic data [60,61]. However none of these studies used a multimodal approach, being the idea behind this analysis, to identify the most important features (among audio, EEG, NIRS, facial expressions and body movements) to better understand the causes of infant crying. Thus, the most significant impact comes from audio quantitative features (e.g., duration, hnr, jitter, shimmer, etc.) that considerably improve all the performance metrics over the rest of the signals, emphasizing the cry acoustic analysis as a powerful biomarker to assess infant's well-being.

Interestingly, the most important audio features found with the ML analyses allowed us to characterize the cry types based on urgency levels (Fig. 4, *y-axis*) and brain (socio/emotional processes) and body (physiological processes) mechanisms (Fig. 4, *x-axis*) representing an overview over the results of the statistical analysis of acoustical, neurophysiological and behavioral signals. In this context, both *distress* and *hungry* (Fig. 4, *upper quadrant*) showed an intense and noisy cry, indicating the need for urgent external/physiological attention for *hungry (*Fig. 4, *right upper quadrant*) and emotional care or internal cognitive regulation for *distress (*Fig. 4, *left upper quadrant*). They also share a higher brain activation, eventually due to a newborn's cognitive regulation attempt to solve this arousal state. *Burp* (Fig. 4, *right bottom quadrant*) requires low urgency levels with an important body activity/effort presence. Both *burp* and *hungry* (Fig. 4, *right quadrant*) showed lower levels of brain oxygenation and a similar muscular tone pattern. This suggests that the physiological states that trigger these cries may involve a more automated response rather than a conscious effort. Regarding *burp* and *fussy* (Fig. 4, *bottom quadrant*), both showed a less urgent communicative pattern, suggesting a less agitated state compared to *distress* or *hungry*. However, they differ in brain activation and oxygenation, and body movement, possibly due to the fact that being *fussy* involves cognitive or emotional regulation, while *burp* cries are more related to physiological and practical care. Lastly, *fussy* (Fig. 4-*left bottom quadrant*) and *sleepy* (Fig. 4, *left-middle quadrant*) are very similar in their communicative pattern (i.e., long cries) indicating a not-so-urgent but emotional care needed or an internal cognitive regulation. However, *sleepy* is more urgent than *fussy*, which probably indicates that the *fussy* state is less aroused than the *sleepy* state. These categories also showed similar low oxygenation brain level and low body movement highlighting an awake but calm state and physiological processes that do not precise particular parts of the body compared to *hungry* and *burp*.

Finally, to address our third aim, we validated the DL algorithm AMSI [18] based on cepstral features and CNN using the collected infant cries as a test set achieving an accuracy of 92%, which represents a 18% improvement in performance over the best accuracy obtained with the ML models. Previous studies on DL for cry type classification reported accuracies between 89% (comparing feeding, pain, sleep cries) [14] and 94% (comparing pain, hunger, discomfort, burp, belly pain cries) using CNN [62] and CNN-RNN (recurrent neural networks) [63] with spectrographic features. However, none of these studies used naturalistic data recordings. Also, previous studies [14,60,61] used a limited sample size and did not provide robust classification metrics to validate their models.

5. Limitations

Nonetheless, there are some limitations in our exploratory study. The foremost limitation concerned the small sample size and the low density of EEG (i.e., only 8 electrodes were recorded) and NIRS (only one frontal electrode was used) systems. Deep subcortical brain structures have been associated in some studies [64,65] with emotional processing and crying but with the EEG system, applied in this study, we did not have access to these structures. Despite this constraint, we found that brain activity patterns showed statistically significant differences in the different cry types, proving how crying analysis can add relevant information to understanding the earlier human communication process. Furthermore, EEG and NIRS data collection and analysis while crying can be quite challenging due increase in noise artifacts, environmental noise, excessive movement, and muscle activity of the infant. Also, the pulse oximeter, as the fingertip sensor, was not always able to detect a clear signal, for example, when the infant was agitated, the signal was polluted with noise artifacts. In our specific scenario, the restriction of infant movement becomes notably intricate, as our intent is to assess all variables within a naturalistic environment. Consequently, this inherent limitation prompts a deliberate selection of methodological strategies designed to enhance the signal's quality. Another limitation of our study is related to the montage of the EEG and NIRS implies the infant wearing a cap and forehead sensor. Despite the lightweight and flexible nature of these components, it is important to acknowledge that they may potentially induce discomfort in the infant, possibly resulting in more fussy cries. Lastly, we were not able to collect balanced data samples for each condition due to the nature of spontaneous crying and naturalistic recording contexts. To mitigate the effect of noisy audio recordings a careful detection and removal of artifacts was carried out ensuring a clean dataset for every feature. Future investigations should include longitudinal and follow-up studies utilizing multimodal datasets, thereby enabling an exploration of the evolving neurodevelopment of infants and the potential influence of this brain development on the patterns elucidated in this study.

We recognize that continuously monitoring newborns' brain physiological patterns, facial expressions, and body movements to classify cry types may not be a practical long-term solution. Therefore, our study puts forth the concept that acoustic analysis alone can significantly enhance the interpretation of newborns' needs and emotions. In real-world scenarios, relying on a single audio signal is undeniably more feasible than deploying a comprehensive array of clinical equipment (such as EEG, NIRS, etc.) for the same purpose. Our research highlights the potential of audio cry analysis as a valuable tool for assessing an infant's well-being, both in clinical and home settings, owing to its practicality and non-invasive nature as a vocal biomarker.

Despite our limitations, our results represent a preliminary step toward the development of a reliable tool for cry analysis. We were able to verify our hypothesis and confirm that what is occurring at an acoustic level can be reinforced with behavioral and brain neurophysiological patterns leading to a better understanding of the human infant cry never accomplished until now. We also confirmed the importance and potential use of cry acoustics as a biomarker for infant well-being assessment. Lastly, we proved the effectiveness of the audio DL algorithms for cry characterization and interpretation.

6. Conclusion

Hence, we set a precedent on cry analysis research, stating that infant cry expression of a need and/or emotion triggers a complex communicative process involving neurophysiological and behavioral patterns that are able to characterize and differentiate cry types associated with distinct arousal states in the newborn.

Moreover, we demonstrated that acoustics play a key role in the interpretation of the reasons for crying over the rest of the infant's features within a multimodal approach using ML. Thus, we also validated the strength of the audio DL AMSI technology as a tremendous aid for the understanding of what is communicated by an infant's crying.

In conclusion, our results support the potential use of infant cry analysis as an automatic, promising, objective, accessible, and non-invasive tool to improve the infant-parent relationship ensuring the family's well-being and proper newborn development. At the same time, this study elicits the implicit consideration for further research in infant cry as a clinical biomarker, also supporting clinicians in the assessment of the infant's health status.

Funding

This research has been funded by Zoundream AG together with the Swiss State Secretariat for Education, Research and Innovation (SERI), the Swiss Innovation Agency (Innosuisse) and the NEOTEC grant (SNEO-20211305) received from the Spanish Center for Technological Development and Innovation (CDTI).

Author contributions

AL: Conceptualization, Funding acquisition, Methodology, Data processing, Formal analysis, Manuscript – original draft preparation. SP: Conceptualization, Methodology, Data processing, Formal analysis, Manuscript – original draft preparation and preparing figures. AB: Methodology, Data processing, Formal analysis, Manuscript – original draft preparation. JAZV: Advice on Data processing, Manuscript – review & editing. ALP: Site facilitator, Conceptualization, Data collection, Methodology. PP: Data processing. CP: Data collection. APV: Manuscript – review. OGA: Site facilitator, Funding acquisition. SO: Advice on Data analysis pipeline and Result discussion, Manuscript – review & editing. All present authors contributed to the article and approved the submitted version.

Summary

In this manuscript, our multi-modal analysis combines infant cry audio recordings and neurophysiological signals such as electroencephalography (EEG) and near-infrared spectroscopy (NIRS) as well as facial expression and body movements from 38 healthy term newborns recruited at Hospital Clínic (Barcelona). Statistical analysis was conducted on a multimodal dataset to characterize and differentiate cry types during five different infant conditions (i.e., hunger, sleepiness, fussiness, burp, distress).

Additionally, Machine Learning (ML) was used within this multimodal dataset to determine the most relevant features that allow the distinction of different crying reasons. Finally, we want to demonstrate the effectiveness of audio Deep Learning (DL) algorithms as a valuable tool for interpreting infant cries and assessing the well-being of newborns.

Our findings show significant differences for the different cry types based on acoustics, EEG, NIRs, facial expressions, and body movements. Acoustical features and body language were found to be the most relevant ML features to support the cause of crying. Moreover, the DL algorithm tested on the naturalistic cry database achieved an accuracy rate of 92%.

With our study we were able to confirm that what is occurring at an acoustic level can be reinforced with behavioral and brain neurophysiological patterns leading to a better understanding of human infant cry never accomplished until now. Also, we were able to confirm the importance and potential use of cry acoustics as a tremendous tool for infant well-being assessment and also to confirm that audio DL algorithms are the best state of the art solution for cry analysis interpretation.

Declaration of competing interest

The authors declare competing interests (Funding, Employment or Confidentiality interests) in relation to the work described herein.

Ana Laguna, Sandra Pusil, Àngel Bazán and Paolo Piras are employed by Zoundream AG. Ana Laguna is also a co-founder of the company and owns stock in Zoundream AG. Silvia Orlandi, Alexandra Pardos Véglia and Jonathan Adrian Zegarra-Valdivia receive compensation for the collaboration as members of the scientific advisory board of Zoundream AG. Clàudia Palomares' salary is funded by Zoundream AG through Fundació Clínic. Anna Lucia Paltrinieri and Oscar Garcia-Algar declare no potential conflict of interest.

Acknowledgements

Zoundream AG, a health tech startup specializing in cry analysis, has been the sponsor of this study in collaboration with the Hospital Clínic Barcelona. The authors thank all the newborns' families and the Hospital Clínic - Maternitat for their trust and kind cooperation during data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.107626.

References

[1] L.L. LaGasse, A.R. Neal, B.M. Lester, Assessment of infant cry: acoustic cry analysis and parental perception, Ment. Retard. Dev. Disabil. Res. Rev. 11 (2005) 83–93.

[2] G. Gremmo-Feger, in: Un autre regard sur les pleurs du nourrisson, 2007, pp. 15–17.

[3] K. Wermke, et al., Fundamental frequency variation in crying of Mandarin and German neonates, J. Voice Off. J. Voice Found. 31 (2017) 255.e25–255.e30.

[4] Crying as a Sign, a Sympton, & a Signal: Clinical Emotional and Developmental Aspects of Infant and Toddler Crying, Cambridge University Press, 2000, p. 228, viii.

[5] L. Girard, Regard anthropologique et signification des pleurs du nourrisson, Cah. Puéricultrice 261 (2012).

[6] S. Oldbury, K. Adams, The impact of infant crying on the parent-infant relationship, Community Pract. J. Community Pract. Health Visit. Assoc. 88 (2015) 29–34.

[7] B. Reggiannini, S.J. Sheinkopf, H.F. Silverman, X. Li, B.M. Lester, A flexible analysis tool for the quantitative acoustic assessment of infant cry, J. Speech Lang. Hear. Res. JSLHR 56 (2013) 1416–1428.

[8] S. Jeyaraman, et al., A review: survey on automatic infant cry analysis and classification, Health Technol. 8 (2018) 391–404.

[9] C. Manfredi, et al., Automated detection and classification of basic shapes of newborn cry melody, Biomed. Signal Process Control 45 (2018) 174–181.

[10] S. Ntalampiras, Audio pattern recognition of baby crying sound events, J. Audio Eng. Soc. 63 (2015) 358–369.

[11] X. Yu, et al., An investigation into audio features and DTW algorithms for infant cry classification, in: Proceedings of the 2019 6th International Conference on Biomedical and Bioinformatics Engineering, Association for Computing Machinery, 2020, pp. 54–59, https://doi.org/10.1145/3375923.3375929.

[12] R. Tuduce, M. Rusu, H. Cucu, C. Burileanu, Automated Baby Cry Classification on a Hospital-Acquired Baby Cry Database, 2019, p. 346, https://doi.org/10.1109/TSP.2019.8769075.

[13] H.B. Sailor, H. Patil, Auditory filterbank learning using ConvRBM for infant cry classification, in: *Interspeech 2018* 706–710, ISCA, 2018, https://doi.org/10.21437/Interspeech.2018-1536.

[14] C.-Y. Chang, J.-J. Li, Application of deep learning for recognizing infant cries, in: 2016 IEEE International Conference on Consumer Electronics-Taiwan, (ICCE-TW), 2016, pp. 1–2, https://doi.org/10.1109/ICCE-TW.2016.7520947.

[15] K. Ashwini, P.M.D.R. Vincent, K. Srinivasan, C.Y. Chang, Deep learning assisted neonatal cry classification via support vector machine models, Front. Public Health 9 (2021), 658–658.

[16] C. Ji, T.B. Mudiyanselage, Y. Gao, Y. Pan, A review of infant cry analysis and classification, EURASIP J. Audio Speech Music Process. 2021 (2021) 1–17.

[17] S. Orlandi, C.A. Reyes Garcia, A. Bandini, G. Donzelli, C. Manfredi, Application of pattern recognition techniques to the classification of full-term and preterm infant cry, J. Voice Off. J. Voice Found. 30 (2016) 656–663.

[18] P.F. Ingraito, A. Laguna Pradas, A Computer-Implemented Method of Providing Data for an Automated Baby Cry Assessment. EU Patent EP4179533, 2020.

[19] A. Laguna, et al., How can cry acoustics associate newborns' distress levels with neurophysiological and behavioral signals? Front. Neurosci. 17 (2023), 1266873.

[20] K. Wermke, W. Mende, Musical elements in human infants' cries: in the beginning is the melody, Music. Sci. 13 (2009) 151–175.

[21] C. García, A.A.T. García, M.A.R. Díaz, On the Development of a Non-invasive Pathologies Identification System by Qualitative and Quantitative Characterization of Infant Crying and the Application of Intelligent Classification Models to Be Used in Rural Environments, 2018.

[22] O. Wasz-Höckert, K. Michelsson, J. Lind, Twenty-five years of scandinavian cry research, in: B.M. Lester, C.F. Zachariah Boukydis (Eds.), Infant Crying: Theoretical and Research Perspectives, Springer US, 1985, pp. 83–104, https://doi.org/10.1007/978-1-4613-2381-5_4.

[23] P. Praat Boersma, A system for doing phonetics by computer, Glot Int. 5 (2002).

[24] L. Rautava, et al., Acoustic quality of cry in very-low-birth-weight infants at the age of 1 1/2 years, Early Hum. Dev. 83 (2007) 5–12.

[25] J.P. Teixeira, C. Oliveira, C. Lopes, Vocal acoustic analysis – jitter, shimmer and HNR parameters, Procedia Technol 9 (2013) 1112–1122.

[26] C. Manfredi, et al., Automated analysis of newborn cry: relationships between melodic shapes and native language, Biomed. Signal Process Control 53 (2019), 101561.

[27] F. Tadel, S. Baillet, J.C. Mosher, D. Pantazis, R.M. Leahy, Brainstorm: a user-friendly application for MEG/EEG analysis, Comput. Intell. Neurosci. 2011 (2011) 1–13.

[28] F. Perrin, J. Pernier, O. Bertrand, J.F. Echallier, Spherical splines for scalp potential and current density mapping, Electroencephalogr. Clin. Neurophysiol. 72 (1989) 184–187.

[29] J.M.O. Toole, G.B. Boylan, NEURAL: Quantitative Features for Newborn EEG Using Matlab, 2017, https://doi.org/10.48550/arxiv.1704.05694.

[30] S. Orlandi, L. Bocchi, G. Donzelli, C. Manfredi, Central blood oxygen saturation vs crying in preterm newborns, Biomed. Signal Process Control 7 (2012) 88–92.

[31] Y.-C. Lu, et al., Reevaluating reference ranges of oxygen saturation for healthy full-term neonates using pulse oximetry, Pediatr. Neonatol. 55 (2014) 459–465.

[32] C. Lian, P. Li, N. Wang, Y. Lu, W. Shangguan, Comparison of basic regional cerebral oxygen saturation values in patients of different ages: a pilot study, J. Int. Med. Res. 48 (2020), 030006052093686–030006052093686.

[33] B. Ambuel, K.W. Hamlett, C.M. Marx, J.L. Blumer, Assessing distress in pediatric intensive care environments: the COMFORT scale, J. Pediatr. Psychol. 17 (1992) 95–109.

[34] E. Olsson, et al., The use and reporting of neonatal pain scales: a systematic review of randomized trials, Pain 162 (2021) 353–360.

[35] Freund, Y. & Schapire, R. E. A short introduction to boosting. J. Jpn. Soc. Artif. Intell. 14, 771–780.

[36] L. Breiman, Random forests, Mach. Learn. 451 (45) (2001) 5–32, 2001.

[37] Model-building strategies and methods for logistic regression, in: Applied Logistic Regression, John Wiley & Sons, Ltd, 2013, pp. 89–151, https://doi.org/10.1002/9781118548387.ch4.

[38] L. van der Maaten, G. Hinton, Viualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[39] K. O'Shea, R. Nash, An Introduction to Convolutional Neural Networks, 2015, https://doi.org/10.48550/arxiv.1511.08458.

[40] C.A. Reyes-Garcia, O.S. Cano, E. Arch-Tirado, Fundamentos Teóricos y Prácticos del Análisis de Llanto Infantil. Una herramienta auxiliar para diagnóstico no invasivo, 2009.

[41] S. Niaz, et al., Variation in oxygen saturation by pulse oximetry during and after breastfeeding among healthy term neonates during early postnatal life at tertiary care hospital, Cureus 13 (2021), e16564.

[42] K. Paul, J. Dittrichová, H. Papousek, Infant feeding behavior: development in patterns and motivation, Dev. Psychobiol. 29 (1996) 563–576.

[43] A.V. Sulimov, [EEG spectral analysis in the hunger dominant in man], Zh. Vysshei Nervn. Deiatelnosti Im. I. P. Pavlova 45 (1995) 1095–1100.

[44] E.V. Orekhova, T.A. Stroganova, I.N. Posikera, Alpha activity as an index of cortical inhibition during sustained internally controlled attention in infants, Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 112 (2001) 740–749.

[45] T.A. Stroganova, E.V. Orekhova, I.N. Posikera, EEG alpha rhythm in infants, Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 110 (1999) 997–1012.

[46] C.A. Tomlinson, K.R. Archer, Manual therapy and exercise to improve outcomes in patients with muscle tension dysphonia: a case series, Phys. Ther. 95 (2015) 117–128.

[47] Cicchetti, Dante. Developmental Psychopathology, Volume 2: Developmental Neuroscience. (Wiley).

[48] G. Berretz, J. Packheiser, O.T. Wolf, S. Ocklenburg, Acute stress increases left hemispheric activity measured via changes in frontal alpha asymmetries, iScience 25 (2022), 103841.

[49] N. Lester, L.E. Nebel, A. Baum, Psychophysiological and behavioral measurement of stress, in: W.R. Avison, I.H. Gotlib (Eds.), Stress and Mental Health: Contemporary Issues and Prospects for the Future, Springer US, 1994, pp. 291–314, https://doi.org/10.1007/978-1-4899-1106-3_11.

[50] G.T. Schneider, M.F. Vaezi, D.O. Francis, Reflux and voice disorders: have we established causality? Curr. Otorhinolaryngol. Rep. 4 (2016) 157–167.

[51] S.R. Orenstein, D.M. Orenstein, Gastroesophageal reflux and respiratory disease in children, J. Pediatr. 112 (1988) 847–858.

[52] V. Poylin, T.E. Cataldo, Anal physiology: the physiology of continence and defecation, in: S.R. Steele (Ed.), The ASCRS Textbook of Colon and Rectal Surgery, 44, Springer International Publishing, 2016, p. 37, https://doi.org/10.1007/978-3-319-25970-3_3, et al.

[53] X. Li, et al., Stress and emotion classification using jitter and shimmer features, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '07, 2007, 4 IV-1081-IV–1084.

[54] S. Miyata, et al., Impaired cortical oxygenation is related to mood disturbance resulting from three nights of sleep restriction, Sleep Biol. Rhythm 13 (2015) 387–394.

[55] S. Fattinger, S. Kurth, M. Ringli, O.G. Jenni, R. Huber, Theta waves in children's waking electroencephalogram resemble local aspects of sleep during wakefulness, Sci. Rep. 7 (2017), 11187.

[56] T. Winkelman, et al., Resting EEG architecture is atypical in children with ASD and sleep problems, INSAR 9 (2018) 15.

[57] C. Cajochen, R. Foy, D.-J. Dijk, Frontal predominance of a relative increase in sleep delta and theta EEG activity after sleep loss in humans, Sleep Res. Online SRO 2 (1999) 65–69.

[58] A.M. Strijkstra, D.G.M. Beersma, B. Drayer, N. Halbesma, S. Daan, Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta (4–8 Hz) frequencies in the human resting awake electroencephalogram, Neurosci. Lett. 340 (2003) 17–20.

[59] E. De Stefani, De Marco, D. Language, Gesture, and emotional communication: an embodied view of social interaction, Front. Psychol. 10 (2019), 2063.

[60] S. Bano, K.M. RaviKumar, Decoding baby talk: a novel approach for normal infant cry signal classification, in: 2015 International Conference on Soft-Computing and Networks Security, ICSNS, 2015, pp. 1–4, https://doi.org/10.1109/ICSNS.2015.7292392.

[61] M.V.V. Bhagatpatil, An automatic infant's cry detection using linear frequency cepstrum coefficients (LFCC), Int. J. Sci. Eng. Res. 5 (2014) 1379–1383.

[62] E. Franti, I. Ispas, M. Dascalu, Testing the universal baby language hypothesis - automatic infant speech recognition with CNNs, in: 2018 41st International Conference on Telecommunications and Signal Processing, TSP), 2018, pp. 1–4, https://doi.org/10.1109/TSP.2018.8441412.

[63] T.N. Maghfira, T. Basaruddin, A. Krisnadhi, Infant cry classification using CNN – RNN, J. Phys. Conf. Ser. 1528 (2020), 012019.

[64] L.M. Bylsma, A. Gračanin, A.J.J.M. Vingerhoets, The neurobiology of human crying, Clin. Auton. Res. 29 (2019) 63–73.

[65] A. Etkin, T. Egner, R. Kalisch, Emotional processing in anterior cingulate and medial prefrontal cortex, Trends Cognit. Sci. 15 (2011) 85–93.