Real-Time Baby Crying Detection in the Noisy Everyday Environment

Lee Sze Foo
Universiti Tunku Abdul Rahman
leesze96@1utar.my

Wun-She Yap
Universiti Tunku Abdul Rahman
yapws@utar.edu.my

Yan Chai Hum
Universiti Tunku Abdul Rahman
humyc@utar.edu.my

Zulaikha Kadim
MIMOS Berhad
zulaikha.kadim@mimos.my

Hock Woon Hon
MIMOS Berhad
hockwoon.hon@mimos.my

Yee Kai Tee
Universiti Tunku Abdul Rahman
teeyeekai@gmail.com

***Abstract*— Baby crying detection is an important component in child monitoring, diagnostics, as well as emotion detection systems. This study proposed a real-time baby crying detection algorithm that monitors the noisy environment for baby crying on a second-by-second basis. The algorithm detected baby crying through five acoustic features – average frequency, pitch frequency, short-time energy (STE) acceleration, zero-crossing rate (ZCR), and Mel-Frequency cepstral coefficients (MFCCs). The thresholds for each feature in classifying an audio segment as "crying" were set by extracting and examining the distribution of the features of noise-free crying and non-crying samples collected from an audio database freely available on the Internet. Later, the algorithm was tested using noisy crying and non-crying samples downloaded from YouTube, where an accuracy of 89.20% was obtained for the offline testing. In order to test the robustness and performance of the designed algorithm, online testings were also conducted using three customly composed noisy samples containing both crying and non-crying segments. The online accuracy obtained was 80.77%, lower compared to the offline testing which was mainly caused by the extra noise introduced by the experimental settings. With more advanced equipment, it should be possible to increase the online testing to be closer to the offline testing accuracy, paving the way to use the designed algorithm for reliable real-time second-by-second baby crying detection.**

Keywords—baby crying, real-time, live, crying detection

I. INTRODUCTION

In the first months of a baby's life, crying is a child's major means of communicating needs or distress. Typically, parents or caregivers are able to recognize a baby's cry with ease and great accuracy. However, this requires the constant monitoring and attention of the caregivers. Having an automated baby crying detection system would not only be able to alleviate the workload of caregivers but also can form an essential component of other baby-crying related applications such as diagnostic systems [1], [2], emotion detection [3], [4], etc. Consequently, the study of automatic baby crying detection algorithms has been the subject of research for many years [5]–[13].

Existing detection algorithms achieve this by analyzing and extracting useful audio features such as pitch and short-time energy from the baby crying sound [7], [14]. Recently, the use of deep learning has also been rapidly implemented in baby crying detection algorithms, particularly convolutional neural network (CNN) algorithms [9], [13] whereby the one dimensional audio signal is converted into a spectrogram – a two dimensional plot of frequency versus time of the audio signal, enabling end-to-end learning and eliminating the need of manual feature selection.

For real-time baby crying detection, the algorithm is required to be fast in analyzing the audio signals to classify the data and to alert the parent in the event of a baby crying. Although the use of deep learning has proven to be relatively accurate in baby crying detection with accuracy of around 85% - 90% [8], its use in this case – real-time detection, may not be suitable due to the long processing time required by the machine learning technique. In addition, the majority of the pervious works focused on analysis of pre-recorded audio data, some of which were recorded in controlled clinical or laboratory environments to better isolate the baby crying signal from the external noise [10], [11]. This also makes it unclear whether the accuracy obtained in noise-free experiments could be generalized to more complicated conditions in our daily lives. For example, when there are background sounds such as human talking and rain, or constant noise coming from the electrical appliances like fan or air-conditioners which may corrupt the recorded audio signal and affect the detection.

In this study, a real-time baby crying detection algorithm was proposed for use in the noisy everyday environment. To achieve the real-time detection target and to decrease the latency due to long processing time required by the more complicated machine learning techniques, this study focused on using only important acoustic baby crying features for the detection. The proposed algorithm first extracted and analyzed temporal and spectral features of the audio signal, and then classified the signal based on the extracted information. The feature thresholds for the classification were learnt via labelled audio data obtained from freely available online database and sources. The proposed algorithm was tested both offline and online (in real-time) using noisy crying and non-crying samples recorded in the everyday environment, e.g. at home or school, to investigate the accuracy, reliability and feasibility of real-time automatic baby crying monitoring.

II. METHODS AND MATERIALS

The algorithm performed analysis on an audio signal on a second-by-second basis to reduce latency while maintaining sufficient amount of information for analysis and detection. There were four stages in the proposed algorithm. Firstly, the input voice data went through preprocessing stage. Subsequently, the audio signal was framed and five features were extracted from the audio signal, namely, average frequency, pitch frequency, maximum short-time-energy (STE) acceleration, zero-crossing rate (ZCR) and Mel-frequency cepstral coefficients (MFCCs) features which will be defined in the following sections. These features were chosen as they were able to distinguish the unique acoustic characteristics of baby crying [7], [15], [16]. Fig. 1 shows the flowchart of the proposed crying detection algorithm.

A. Preprocessing

In the preprocessing stage, any stereo audio signal was

978-1-7281-5313-1/20/$31.00 ©2020 IEEE

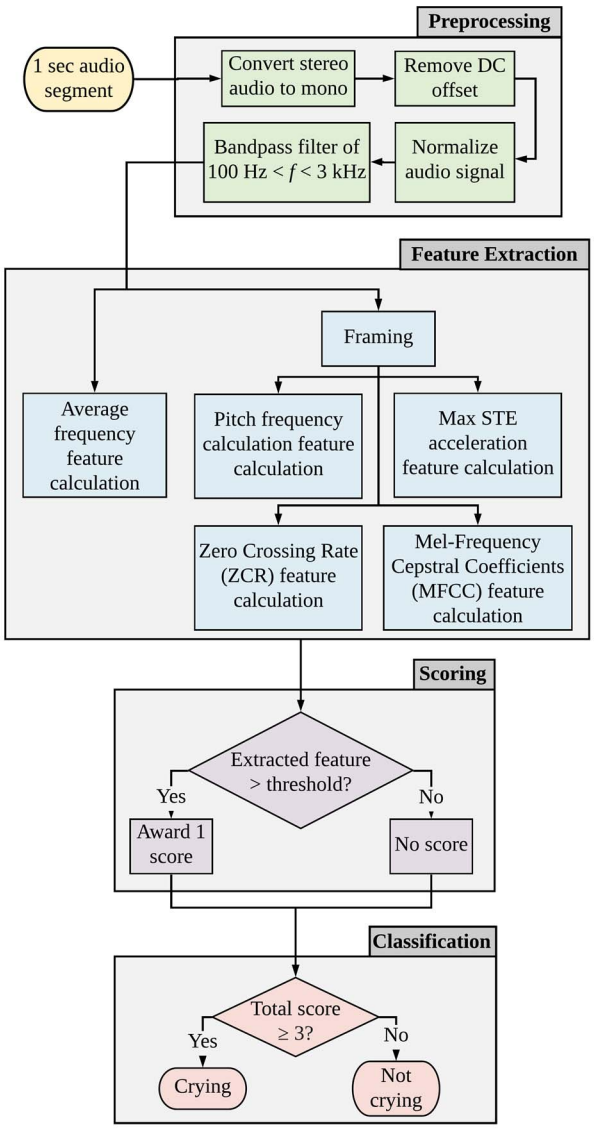

Fig. 1: Proposed algorithm for baby crying detection.

converted into mono by taking the average of the two channels. Next, the direct current (DC) offset of the raw audio signal was removed and normalized to ensure a consistent feature extraction in the following stages. After that, a digital bandpass filter with a lower limit of 100 Hz and upper limit of 3 kHz the audio signal was applied to reduce noise and non-vocal sounds in the signal [10].

B. Feature Extraction

After the preprocessing stage, the pre-processed one-second segments were further framed into 50 ms frames, with no overlapping, except when stated otherwise below:

i. Frequency

The frequency spectrum of the input sound, *S*(t) was calculated using the Fast Fourier Transform (FFT). From the frequency spectrum, the highest magnitude of the investigated frequency range, mag_{max} was determined, and the frequency band where the audio signal mostly stayed, termed as feature average frequency, was calculated as the sum of magnitudes of the frequency spectrum that were larger than 25% of mag_{max}:

$$Average\ Frequency_n = \frac{\sum freq_n * mag_n}{\sum mag_n}, \quad (1)$$

where $mag_n > 0.25\ mag_{max}$.

ii. Short-Time Energy (STE) and Max STE Acceleration

The short-time energy (STE) of an audio signal was calculated as below:

$$E(n) = \sum_{n=1}^{N} S^2(n)\,. \quad (2)$$

During a crying episode, there was usually a silent period intermittent for the baby to breathe. This occurrence might be detected via the maximum rate of change of the STE, term as STE acceleration, and was calculated as:

$$Max\ STE_{acc} = max\ (E(n+1) - E(n)). \quad (3)$$

iii. Zero Crossing Rate

Zero crossing rate (ZCR) is the rate at which the amplitude of a signal crosses the zero axis in the specific frame:

$$\mathrm{ZCR}_n = \frac{1}{N}\sum_{m=0}^{N-1} \left|\mathrm{sign}\big(x(n-m)\big) - \mathrm{sign}\big(x(n-m-1)\big)\right| w(m), \quad (4)$$

where $\mathrm{sign}\big(x(m)\big) = 1,$ if $x(m) \geq 0$

$\mathrm{sign}\big(x(m)\big) = -1.$ otherwise

ZCR is useful for voiced sound detection because there are noticeably fewer zero crossings in voiced speech compared to unvoiced sounds [16].

iv. Pitch Frequency

Fundamental frequency is one of the most important features for baby crying detection. A baby's cry originates from rhythmical transitions between inhalation and exhalation as a result of vocal cord vibrations that produce periodic air pulses [9]. The period of these pulses correspond to the fundamental frequency of the crying.

The fundamental frequency was calculated by first determining the fundamental period – the time between successive vocal fold openings. The fundamental frequency of the phonation (rate of vibration) was then determined as the inverse of the fundamental period [14]. To find the fundamental period, R, of the audio signal, short-time-autocorrelation was used:

$$R_n(k) = R_n(-k) = \sum_{m=-\infty}^{\infty} [x(m)x(m-k)][w(n-m)w(n-m+k)]. \quad (5)$$

In this implementation, pitch frequency feature referred to the number of frames in which the fundamental frequency exceeded 200 Hz.

v. Mel-Frequency Cepstral Coefficient (MFCC)

Cepstral coefficients are the most widely used features for speech recognition. Since human hearing has non-linear tone perception, Mel scale frequencies are used to represent the low frequencies (below 1000 Hz) in a linear scale, and the high frequencies (above 1000Hz) in a logarithmic scale, to

better emulate human hearing [14].

Firstly, the audio signal was split into overlapping frames of 50 ms duration with 10 ms overlaps. Then, the discontinuity between consecutive frames was reduced by implementing a Hamming window defined as [8]:

$$w_n = 0.5\left(1 - \cos\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \tag{6}$$

where w_n is the windowing function, and N is the number of samples per frame.

FFT was then applied on the windowed signal and the power spectrum of the signal was computed. From there, the Mel filter-bank was calculated:

$$\text{Mel}(f) = 2595\log_{10}(1 + \frac{f}{700}). \tag{7}$$

Lastly, discrete cosine transform is applied on the Mel log amplitudes, resulting in amplitudes of the spectrum also known as MFCCs:

$$\text{MFCC}_n = \sum_{k=0}^{n-1} \log(S_k)\cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{k}\right], \tag{8}$$

where $n = 1,2,\dots,K$, and S_k is the output power spectrum of filters; K was chosen to be 12 in this study. For the proposed algorithm, the 4th to 12th coefficients, corresponding to frequencies within the bandpass filter range, were summed up and normalized by N to give the MFCC feature.

C. Scoring and Classification

Upon completing feature extraction, one score was awarded to every extracted feature that exceeded the threshold set for the respective feature (elaborated in Section II-D). If the total score of the audio signal was equal to or more than 3, the signal was classified as "crying" sound. Otherwise, it was labelled as "non-crying".

D. Experiments

The crying detection algorithm was implemented using MATLAB (Mathworks, Natick, MA, USA). A total of three experiments were carried out.

i. Training Using Clean Samples

Firstly, "pure" features of both crying and non-crying sounds were extracted from relatively noise-free crying and non-crying samples. The distribution of the extracted features from the crying and non-crying samples were used to set the thresholds for each feature such that crying and non-crying sounds could be distinguished clearly from one another. These thresholds set were used for the implementation of the crying detection algorithm for all subsequent experiments.

ii. Offline Testing Using Noisy Samples

The second experiment was an offline crying detection test, where the crying detection algorithm was applied on noisy children crying samples as well as non-crying samples to test the accuracy of the proposed algorithm in the noisy everyday environment. In this experiment, MATLAB was used to open the audio samples and the algorithm was applied directly onto the samples in MATLAB. From the algorithm predictions, the accuracy of the crying detection algorithm was calculated as:

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total number of samples}}, \tag{9}$$

where True Positives (TP) refers to the number of crying segments that were accurately predicted as "crying"; True Negatives (TN) refers to the number of non-crying segments that were accurately predicted as "non-crying".

In addition, the sensitivity and specificity of the algorithm were calculated following (10) and (11) respectively:

$$\text{Sensitivity} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}, \tag{10}$$

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}}, \tag{11}$$

where False Negatives (FN) refers to the number of crying segments falsely predicted as "non-crying", and False Positives (FP) refers to the number of non-crying segments falsely predicted as "crying". In this case, sensitivity refers to the ability of the algorithm to correctly predict crying segments, while specificity refers to its ability to correctly predict non-crying segments.

iii. Online Testing Using Custom Noisy Samples

The third experiment was on implementing the algorithm in real-time second-by-second detection. In this experiment, the algorithm's real-time detection performance was determined using three customly composed noisy samples, each containing both noisy "crying" and "non-crying" segments. To simulate monitoring of a noisy environment, speakers were used to play the custom noisy samples while a microphone was used to record the sound, which was then immediately processed using the proposed algorithm to determine whether it was a baby crying sound, on a second-by-second basis. A Blue Snowball condenser microphone and Salpido Tron 101 Dynamic Modern Sound System were used in the online (live) testing. The distance between the microphone and the speakers was around 25 cm, and the speakers were oriented to directly face the microphone. Similar to offline testing, the accuracy of the real-time crying detection was calculated using (9), while the sensitivity and specificity were respectively calculated following (10) and (11).

E. Database

i. Clean Samples

To determine the thresholds for "crying" or "non-crying" classification, a free dataset provided by Wang [18] was used. In the dataset, there were a total of 21 baby crying samples and 6 non-crying samples containing adult speaking. For the crying samples, Samples 17.wav, 18.wav, 19.wav and 21.wav were excluded as these samples were relatively noisy, and if included, could affect the learning to set the appropriate thresholds for the classification. The remaining 17 crying samples (Samples 1.wav to 16.wav and 20.wav) used were recorded in a quiet environment (relatively noise-free) and thus were deemed suitable for learning the "pure" baby crying features. For the non-crying samples, all 6 adult speaking samples were included.

Each audio sample was segmented into one-second segments for feature extraction. From the baby crying samples (Samples 1.wav to 16.wav and 20.wav), the one-second segments of each sample were manually labelled as "crying" if the segment contained obvious shouting or crying episodes. Only these "crying" segments were selected for learning the crying features; segments with only baby inhaling or whimpering were discarded. Fig. 2 shows an example of the labelling of each one-second segment of Sample "1.wav". From the adult speaking samples (Samples 26.wav to 32.wav), all of the one-second segments from these samples were used as "non-crying" segments for learning the non-crying features.

From the dataset, a total of 301 one-second "crying" segments (from the 17 baby crying samples) and 39 one-second "non-crying" segments (from the adult speaking samples) were obtained, amounting to 340 (= 301 + 39) one-second segments in total.

ii. Noisy Samples

For offline testing, a different dataset from the learning samples was used. Videos containing children crying in noisy everyday environments such as in kindergarten school, daycare centres, and at home, were taken from YouTube. The audio of these videos were extracted as noisy crying samples. These samples contained external noises, e.g. adults speaking, television noises, music, etc. In addition, noisy non-crying samples were also extracted from YouTube, including videos of adults (both male and female) speaking in noisy everyday environments.

Similar to the previous experiment with the clean samples, the noisy samples were segmented into one-second segments for analysis. In the crying videos, there were "crying" segments as well as "non-crying" segments which were of adults or children speaking, sounds from the television, etc. As such, each one-second segment was manually labelled as "crying" or "non-crying" to determine the ground truth.

In total, 13 crying and 3 non-crying noisy videos were included in the offline testing to maintain a similar proportion of the number of "crying" and "non-crying" one-second segments in the offline testing. This was because the crying videos contained both "crying" and "non-crying" segments while the non-crying videos were all "non-crying". If the number of non-crying videos used was increased to 13 like the crying case, the number of "non-crying" one-second segments would be much more than the "crying" one-second segments.

Besides that, only the first 3 minutes of each video were used in the offline testing to prevent the results from being significantly skewed by certain clips. From all the noisy samples, 254 one-second "crying" segments and 672 one-second "non-crying" segments were obtained, amounting to 926 (= 254 + 672) one-second segments in total.

iii. Custom Noisy Samples

To evaluate the performance of the proposed algorithm for online (real-time) detection, three custom noisy samples, each consisting of "crying" and "non-crying" segments were created by randomly selecting the audio segments from the samples used in the offline testing and merged into a continuous audio clip with no repetition. The three custom noisy samples produced contained 68 one-second "crying" segments and 62 one-second "non-crying" segments in total.

Overall, Fig. 3 shows the summary of the three experiments performed and the respective dataset used for each experiment. In short, Experiment 1 trained the detection algorithm using clean samples to learn the suitable audio feature thresholds for baby crying detection, while Experiments 2 and 3 tested the designed algorithm using the noisy samples in offline and online (real-time) testing respectively. The experimental setup for the online (live) testing is shown in the third column in Fig. 3.

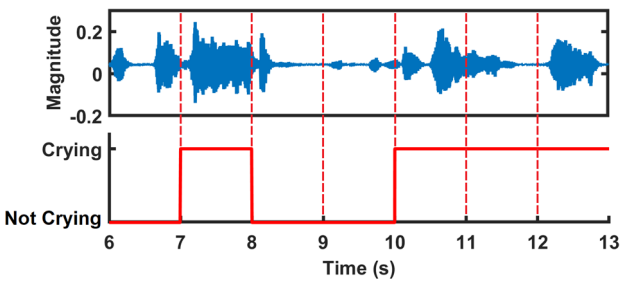

Fig. 2: Sample "1.wav" with labels indicating the "crying" one-second segments. Only a portion of the audio signal, seconds 6 – 13, are shown for illustration purposes.

III. Results

A. Clean Samples For Training

Fig. 4 shows an example of the extracted features of a "crying" segment – 2nd segment of the clean crying sample "1.wav". Using (1), the average frequency of this segment was found to be 1339 Hz. In the figure, it can also be observed that the majority of the 50 ms frames had pitch frequencies of more than 200 Hz and ZCR of more than 100. The STE acceleration was found to be maximum in the second frames at 329.6 J. The summation of the MFCCs of the frames was calculated to be -3.708.

The distribution of the extracted features of all the clean crying and non-crying samples are shown in Fig. 5. To distinguish between crying and non-crying sounds, a threshold was set for each feature based on the region of minimal overlapping in the distributions. TABLE I shows the thresholds set and these thresholds were used to implement the crying detection algorithm in all the subsequent offline and online experiments.

B. Offline Noisy Sample Crying Detection

TABLE II shows the prediction results of the crying detection algorithm on noisy samples recorded in the everyday environment. The accuracy of the predictions was calculated to be 89.20%; the sensitivity and specificity were determined to be 85.04% and 90.77% respectively.

C. Online Custom Noisy Sample Crying Detection

TABLE II also shows the crying detection algorithms of the online testing using noisy samples. The accuracy obtained was 80.77%. The sensitivity and specificity were calculated to be 78.26% and 83.61% respectively.

The proposed algorithm implemented in an Intel i7-8750H CPU (2.20 GHz) with 8 GB of RAM took around 25 ms of processing time (preprocessing, feature extraction, scoring and classification) to make a prediction.

IV. Discussion

In this study, a baby crying detection algorithm was proposed and tested with noisy crying samples in both offline

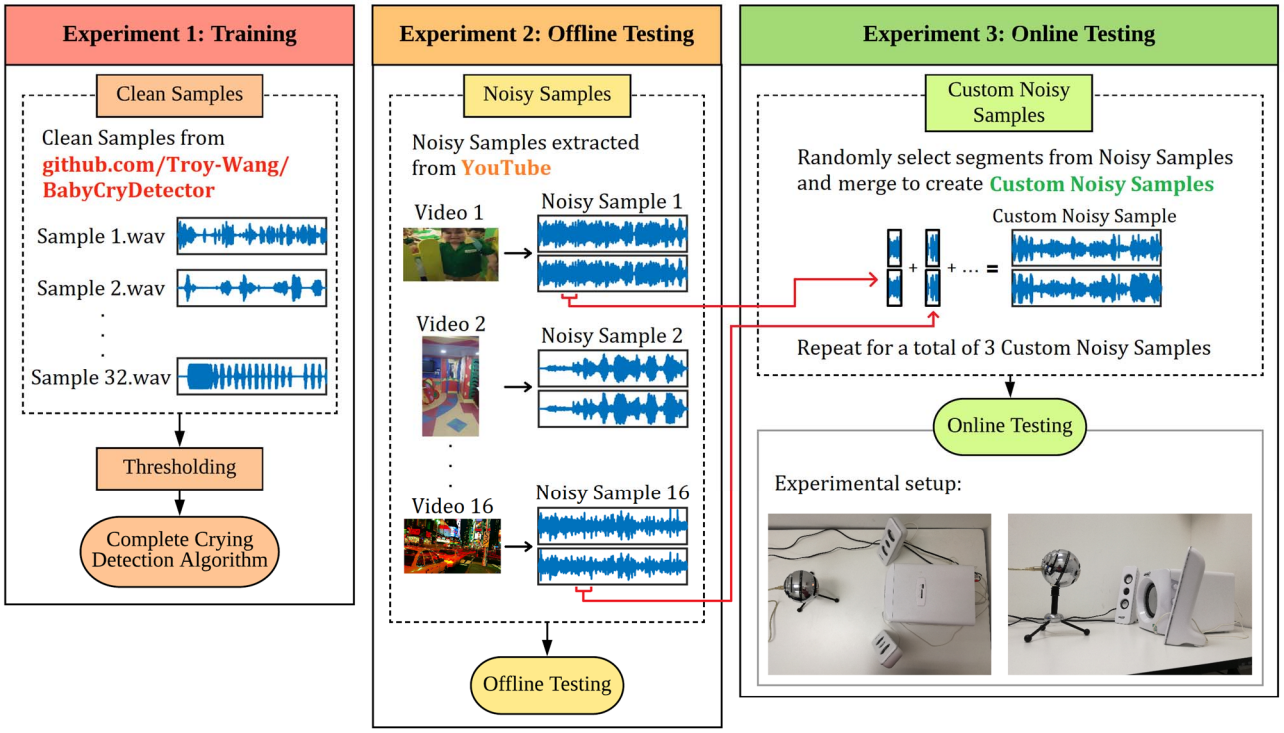

Fig. 3: Summary of experiments and dataset used. In total three experiments were carried out with three separate datasets used for each experiment.

and online (real-time) experiments. The results indicated that in the offline testing, the accuracy of the crying detection algorithm achieved an accuracy of 89.20% with high sensitivity and specificity, as shown in TABLE II. It was found that most of the crying segments falsely predicted as "non-crying" were segments in which the crying sounds were very soft in volume compared to the noisy environment, e.g. when an adult is speaking loudly over the child or when the child is very far away from the audio recorder. The crying in these segments were very soft though still distinguishable as crying for human listeners. This suggested that the crying

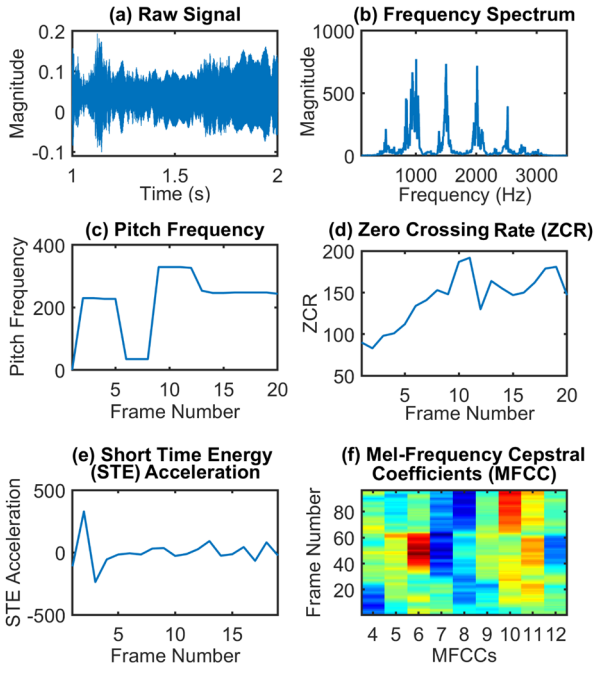

Fig. 4: Extracted features of the 2nd segment of the noise-free crying sample "1.wav".

detection algorithm was unable to identify very soft crying sounds when the signal was mainly dominated by the noisy environment. The False Positive predictions – "non-crying" segments falsely predicted as "crying", on the other hand, were found to be mostly due to segments where an adult was speaking loudly or shouting at relatively high pitch which indicated that the algorithm had difficulty in distinguishing between children and adult high pitch talking or shouting.

When applied in real-time detection using the setup in Fig. 3, there was a decrease in accuracy as well as the sensitivity and specificity, compared to offline testing. The drop should be caused by the reduced quality of the recorded signal due to the experimental setup. This is because the main difference between the offline and online testing was that the former used MATLAB to open and process the audio files directly, while in the online testing, the audio files were played using a speaker and a microphone was simultaneously used to record the sound for processing in MATLAB in real time. Therefore, the effectiveness of the algorithm was dependent on the experimental equipment used and the quality of the recorded sound.

Overall, the proposed crying detection algorithm in offline testing was comparable to published methods in the literature (ranging from 85% to 96.55%) [6], [8], [19]. In the case of real-time detection, which does not have many published works, our proposed algorithm with accuracy of 80.77% in 25 ms processing time was significantly better than one of the closely related publications which reported an accuracy of 71.6% [20]. One important thing to note in this work is that there are not many labelled noisy second-by-second crying and non-crying datasets available as it is very tedious and labor intensive to perform the manual labelling. This has made direct comparison of our results with others a non-trivial task. The labelled dataset used in this work can be

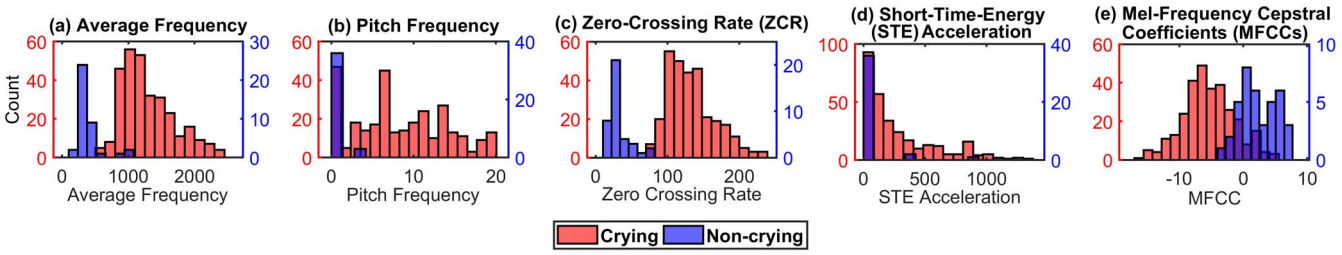

Fig. 5: Histogram of extracted features of crying and non-crying samples.

TABLE I. THRESHOLDS OF EACH FEATURE FOR CRYING CLASSIFICATION

Feature	Average Frequency	Pitch Frequency	ZCR	STE Acceleration	MFCC
Threshold	> 700 Hz	> 10	> 100	> 70	< -1

TABLE II. CRYING DETECTION ALGORITHM PREDICTION OF THE OFFLINE AND ONLINE NOISY SAMPLES

		Truth	
		Crying	***Non-crying***
Prediction (Offline)	***Crying***	216 (TP)	62 (FP)
	Non-crying	38 (FN)	610 (TN)
Prediction (Online)	***Crying***	54 (TP)	10 (FP)
	Non-crying	15 (FN)	51 (TN)

downloaded from the link: https://github.com/Jaclyn-Foo/Baby-Crying-Detection-Database.

The proposed algorithm can be further improved in a number of ways. In this study, only a total of 340 clean one-second segments were used for setting the thresholds of the extracted features due to the scarcity of the clean crying samples and tedious labelling work. The effectiveness of the feature thresholds could be further improved by using more clean samples and a larger variety of non-crying sounds. Besides that, the accuracy of real-time crying detection could be improved by using a better quality speaker and microphone such that the quality of the played and recorded signal would not be distorted significantly.

V. CONCLUSION

The proposed crying detection algorithm with five audio features was able to achieve high accuracy in both offline (accuracy = 89.20%, sensitivity = 85.04% and specificity = 90.77%) and online testing (accuracy = 80.77%, sensitivity = 78.26% and specificity = 83.61%) with negligible processing time (25 ms), paving the way for real-time second-by second automatic baby crying detection system. For future works, other audio features such as delta-delta of MFCC and machine learning techniques can be considered if processing time is less critical.

REFERENCES

[1] R. Sahak, Y. K. Lee, W. Mansor, A. I. M. Yassin, and A. Zabidi, "Optimized Support Vector Machine for classifying infant cries with asphyxia using Orthogonal Least Square," in *ICCAIE 2010 - 2010 International Conference on Computer Applications and Industrial Electronics*, 2010.

[2] A. Zabidi, L. Y. Khuan, W. Mansor, I. M. Yassin, and R. Sahak, "Classification of infant cries with asphyxia using multilayer perceptron neural network," in *2010 2nd International Conference on Computer Engineering and Applications, ICCEA 2010*, 2010.

[3] R. Hidayati, I. K. E. Purnama, and M. H. Purnomo, "The extraction of acoustic features of infant cry for emotion detection based on pitch and formants," in *International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009, ICICI-BME 2009*, 2009.

[4] S. Nagarajan, R. Rengarajan, N. Manoharan, and K. D. Baskaran, "Infant cry analysis for emotion detection by using feature extraction methods," in *Proceedings of WRFER International Conference*, 2017, pp. 66–69.

[5] L. Abou-Abbas, H. Fersaie Alaie, and C. Tadj, "Automatic detection of the expiratory and inspiratory phases in newborn cry signals," *Biomed. Signal Process. Control*, 2015.

[6] M. V. V Bhagatpatil and V. M. Sardar, "An Automatic Infants Cry Detection Using Linear Frequency Cepstrum Coefficients(LFCC)," *Int. J. Sci. Eng. Res.*, vol. 5, no. 12, pp. 1379–1383, 2014.

[7] R. Cohen and Y. Lavner, "Infant cry analysis and detection," in *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2012*, 2012.

[8] S. Dewi, A. Prasasti, and B. Irawan, "The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods," in *2019 IEEE International Conference on Signals and Systems*, 2019.

[9] Y. Lavner, R. Cohen, D. Ruinskiy, and H. Ijzerman, "Baby cry detection in domestic environment using deep learning," in *2016 IEEE International Conference on the Science of Electrical Engineering, ICSEE 2016*, 2017.

[10] A. Messaoud and C. Tadj, "A cry-based babies identification system," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010.

[11] P. R. Myakala, R. Nalumachu, S. Sharma, and V. K. Mittal, "An intelligent system for infant cry detection and information in real time," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2017, pp. 141–146.

[12] P. Ruvolo and J. Movellan, "Automatic cry detection in early childhood education settings," in *2008 IEEE 7th International Conference on Development and Learning, ICDL*, 2008.

[13] R. Torres, D. Battaglino, and L. Lepauloux, "Baby cry sound detection: A comparison of hand crafted features and deep learning approach," in *Communications in Computer and Information Science*, 2017.

[14] L. Abou-Abbas, C. Tadj, and H. A. Fersaie, "A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes," *J. Acoust. Soc. Am.*, 2017.

[15] C. S. Reddy, S. Ravi, and C. V. Giriraja, "Baby Monitoring Through MATLAB Graphical User Interface," *Int. J. Sci. Technol. Res.*, vol. 3, no. 7, pp. 174–177, 2014.

[16] K. Kuo, "Feature extraction and recognition of infant cries," in *2010 IEEE International Conference on Electro/Information Technology, EIT2010*, 2010.

[17] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1980.

[18] Z. Wang, "Baby Cry Detector," 2018. [Online]. Available: https://github.com/Troy-Wang/BabyCryDetector.

[19] M. A. Ruíz, C. A. Reyes, and L. C. Altamirano, "On the implementation of a method for automatic detection of infant cry units," in *Procedia Engineering*, 2012.

[20] M. Kia, S. Kia, N. Davoudi, and R. Biniazan, "A detection system of infant cry using fuzzy classification including dialing alarm calls function," in *2nd International Conference on Innovative Computing Technology, INTECH 2012*, 2012.