Features Motivated From Uncertainty Principle for Classification of Normal *vs.* Pathological Infant Cry

Aastha Kachhi, Priyanka Gupta, Hemant A. Patil
Speech Research Lab, DA-IICT Gandhinagar, Gujarat, India
Email: {aastha_kachhi, priyanka_gupta, hemant_patil}@daiict.ac.in

***Abstract*—This study investigates the significance of the proposed uncertainty vector (u-vector) derived from Heisenberg's uncertainty principle in signal processing framework, and related t-vector and ω-vector for classification of normal *vs.* pathological infant cries. Fundamental frequency (F_0) contours in an infant's cries are rhythmic and have patterns w.r.t. the variances in time and frequency-domains. On the other hand, pathological cries have more smeared variances. To that effect, we propose u-vector as a combination of two features, namely, t-vector, and ω-vector, which capture variances in time and frequency domains, respectively. Furthermore, we also find that non-cepstral features are better suited for pathology detection as compared to the cepstral features. Given that early diagnosis of pathology in infants is important, the detection procedure should be efficient in terms of the time taken. To that effect, we also present latency analysis of u-vector, t-vector, and ω-vector w.r.t. Constant Q Transform (CQT) as baseline. We observe that ω-vector achieves remarkable latency w.r.t. the remaining features, indicating the potential of the proposed feature for practical system deployment. The ω-vector also shows the best performance of 98.50% as overall accuracy, and achieves an absolute improvement of 1.5% in accuracy.**

***Index Terms*—Infant cry classification, Heisenberg's uncertainty principle, Time-Bandwidth Product, *u*-vector, latency.**

I. Introduction

An infant's only means of communication is through crying. Various diseases, malnutrition, and vaccine-preventable diseases claim the lives of millions of infants within a few months of birth. Fingerprint and cry-based identification methods for neonates are being developed for this purpose [1], [2]. Asphyxia, asthma, and Sudden Infant Death Syndrome (SIDS) are among the most common causes of infant deaths [3]. To clinically identify these illnesses, several measurements, such as images from Head Ultrasound (HUS), Computed Tomography (CT) scans, and Magnetic Resonance Imaging (MRI) are used which indicate damaged areas of the brain. However, in many developing countries, detecting pathology takes a long time and is costly which can impact the health of the infant because not every infant has the comfort of having access to healthcare, and support from paediatricians.

For example, one of the pathologies, asphyxia, is recognized through visual symptoms, such as pale and bluish limbs. However, by that time, significant neurological damage to the infant would have already been occurred [3], [4]. In the same manner, the acoustical and perceptual characteristics of deaf infants are influenced by the severity of the hearing loss, the type and duration of rehabilitation, and the age at which the pathology was diagnosed. Thus, need of developing diagnostic assistive tools using infants' cries is increasing to help paediatricians in detecting the early signs of such pathologies [5]. The physiological, neurological, paediatrics, engineering development linguists, and psychology disciplines are all involved in the infant cry analysis and classification. In this context, this paper proposes a signal-processing based method for binary classification of normal *vs.* pathological cries.

The initial analysis of normal *vs.* pathological cries started in early 1960s, where spectrogram was used [6]. In both the time and frequency-domains, a number of various characteristics of cry sounds (also called as 'cry modes' or 'cry phonemes'), such as vibration, dysphonation, inhalation and hyperphonation were explored [6]. This study was extended to the pathological cries in [7], where these cry modes were found to be correlated. Because of quasi-periodic sampling of the vocal spectrum via high pitch-source harmonics, the formant structures beyond F_0 are barely visible in the cry mode investigation. The spectral variations in a cry unit, have the ability to fetch the abundance of information from the cry signals [7].

To that effect, we propose a new method using variations in both time and frequency-domains simultaneously obtained using Time-Frequency Distribution (TFD). TFD indicates the energy spectral density of a signal in both time and frequency domains, which represents information in the form of Heisenberg boxes. The area of the Heisenberg's box is dictated by the Heisenberg's principle in signal processing framework [8]. Mathematically, the area of the Heisenberg's box is given by Time-Bandwidth Product (TBP). The value of TBP is proportional to the number of sample points needed to generate the distribution of the stochastic process, hence, the higher the value of TBP, the higher is the information content in the signal under consideration [9]. Furthermore, the randomness in a signal is captured by the uncertainty information. The uncertainty is therefore caused by the randomness of the process that generates the signal [9], [10].

This work is motivated by the proposition that melody and rhythm (prosody) understanding and memory begin around the third trimester of pregnancy, and infants have a remarkable musical aptitude, where melody contours of F_0 are prominent [11]. Hence, TBP (i.e., the uncertainty in information) can be helpful to distinguish normal *vs.* pathological cries based on the information from the F_0 contours. To that effect,

we propose uncertainty-based feature set called as *u-vector* for the classification of normal *vs.* pathological cries. This proposed feature set is based on the Heisenberg's uncertainty principle, to capture the uncertainty in the cry signal in signal processing framework. The u-vector is constructed with the help of two other feature vectors, namely t-vector and ω-vector, which are also introduced through this work for infant cry analysis. These vectors represent the variance in time and frequency domains, respectively. Unlike other state-of-the-art and handcrafted features, the u-vector, t-vector and ω-vector features are easy and fast to compute. This is validated by latency period analysis shown in this work. The significance of this study is further strengthened by information-theoretic measure, namely Kullback-Leibler Divergence (KLD) and Jenson-Shannon Divergence (JSD) based discriminative measures.

II. Time Bandwidth Product

Let $s(t)$ be a practical non-stationary signal having Fourier transform $S(\omega) = \mathcal{F}\{s(t)\}$. If $s(t)$ has *regular* time variations, then $S(\omega)$ decays fast in high frequency region. This leads to longer spread of energy of $s(t)$ in the time-domain [8]. In this context, from Mallat's proposition in [8] (chapter 2, proposition 2.1), a function $s(t)$ is bounded and i times continuously differentiable with bounded derivatives if [8]

$$\int_{-\infty}^{\infty} |S(\omega)|(1+|\omega|^i)d\omega < +\infty. \tag{1}$$

Here, $S(\omega) = F\{s(t)\} \in L^1(R)$. However, time spread can be restricted by doing the following operation given by:

$$s_\alpha(t) = s\left(\frac{t}{\alpha}\right), \tag{2}$$

where the scaling factor is $\alpha < 1$. Using the time-scaling property [12], the Fourier transform of the signal $s(t)$ is

$$S_\alpha(\omega) = |\alpha| S(\alpha\,\omega). \tag{3}$$

Since the scaling factor $\alpha < 1$, eq. (3) the Fourier transform expands by a factor of $\frac{1}{\alpha}$. Thus, it shows that gain in time localization counter-affects the gain in localization in the frequency-domain and vice-versa [12].

From Heisenberg's uncertainty principle in quantum mechanics, it is impossible to find the precise location and momentum of any particle simultaneously [13]. Similarly, in signal processing framework, the energy spread in time and frequency-domain is restricted by Heisenberg's uncertainty principle [10]. Hence, the average location of the signal $s(t) \in L^2(R)$ is given as:

$$\bar{t} = \int_{-\infty}^{\infty} \frac{1}{||s||^2} t|s(t)^2|dt, \tag{4}$$

and the average momentum is given by

$$\bar{\omega} = \int_{-\infty}^{\infty} \frac{1}{2\pi||s||^2} \omega|S(\omega)^2|dt. \tag{5}$$

The obtained $\bar{\omega}$ from eq. (5) is also called as *effective bandwidth* by Gabor [14]. The variance, i.e., σ_t^2 and σ_ω^2 around these averages represents the uncertainty in determining the particle's position and momentum, respectively [14]. The average time variance can be calculated as

$$\sigma_t^2 = \int_{-\infty}^{\infty} \frac{1}{||s||^2} (t-\bar{t})^2|s(t)^2|dt, \tag{6}$$

and the average momentum is given by

$$\sigma_\omega^2 = \int_{-\infty}^{\infty} \frac{1}{2\pi||s||^2} (\omega-\bar{\omega})^2|S(\omega)^2|dt. \tag{7}$$

It can be seen from eq. (6) and eq. (7), that signal expansion in one domain results in signal contraction in the other domain. As a result, the signal spread in either domain has an inverse relationship. Hence, the Time Bandwidth Product (TBP) given by $\sigma_t^2\sigma_\omega^2$ is constant and represents the area of the Heisenberg's box. This product gives the "richness" of information from the infant cry segment under consideration [9], [15]–[17].

Given that melody contours of F_0 are prominent in children and infant cries [11], and the F_0 contours are not as *rhythmic*, and are smeared in pathological cries, in this study we extract discriminative features for infant cry classification using σ_t^2, σ_ω^2 and the product $\sigma_t^2\sigma_\omega^2$. To that effect, we propose t-vector, ω-vector, and u-vector features for the detection of pathological cries.

III. Feature Vector Extraction Procedure

The proposed feature extraction for infant cry classification is based on the fact that the spectral energy density patterns are different for healthy *vs.* pathological cries. This is also shown by the spectrographic analysis in Figure 2 which shows that pathological cries have high frequency of inhalation, indicating problem while breathing. Hence, spectral smearing is found in the entire frequency range. It can also be observed that there is a sudden rise in the pitch harmonics and spreading in some regions. Therefore, the frequency variance helps to capture the regions of spectral smearing.

The feature extraction procedure in this work begins by passing the cry signal $s(t)$ through a Gabor filterbank of 40 filters. This results in 40 subband signals $s_i(t)$, where $i \in$ [1,40]. Since the cry signal is multi-component, the subband signals help in capturing frequency variances effectively [18]. Here, 40 linearly-spaced Gabor filterbank is used because of its optimal time and frequency resolution [19], [20]. Each of the subband output signals is frame blocked with a window size of 30 ms and window shift duration of 15 ms (experimentally optimized). For each of these frames, both σ_t^2 and σ_ω^2 is computed using the eq. (6) and eq. (7), respectively and hence, three different vector representation of the input speech signal are obtained as shown in Algorithm 1. Next, logarithmic operation is then performed on σ_t^2 and σ_ω^2 to give *t-vector* and ω *-vector* of the cry signal. Similarly, logarithm on the product $\sigma_t^2\ \sigma_\omega^2$ gives the *u-vector* or the uncertainty vector of the cry signal, as indicated by eq. (8) and eq. (9).

$$log(\sigma_t^2\sigma_\omega^2) = log(\sigma_t^2) + log(\sigma_\omega^2), \tag{8}$$

$$\textit{u-vector} = \textit{t-vector} + \omega\textit{-vector}. \tag{9}$$

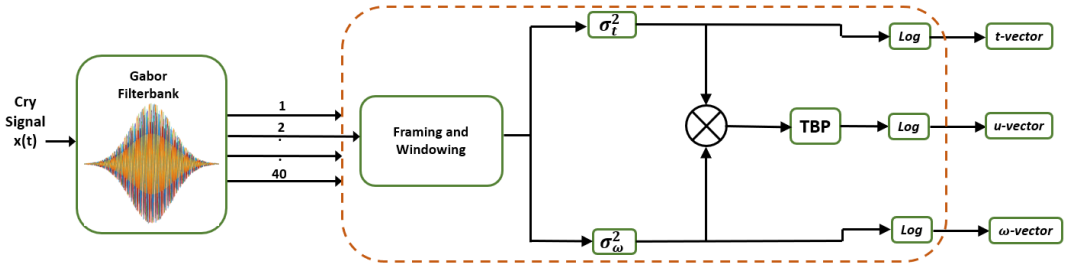
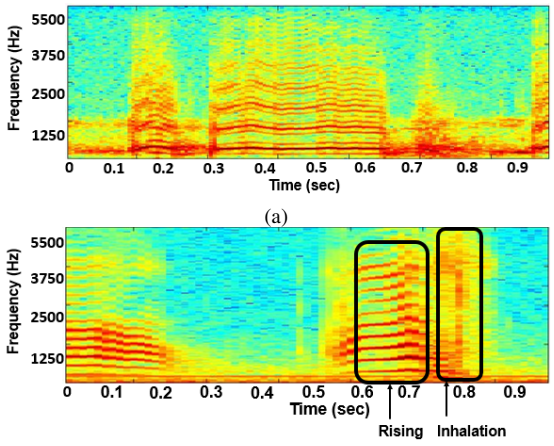


Fig. 1: Functional block diagram of u-vector, t-vector, and ω-vector feature extraction.

Algorithm 1: TBP Computation for Infant cry

Input: Input: cry signal x
Output: Output: u-vector
$T \leftarrow$ Gabor filterbank (x)
Window length = 30 ms, window overlap = 15 ms
For j=1:number of frames **do**
$Var_t \leftarrow$ Variance(T(j, :), mean) $\leftarrow$**{t-vector}**
$mean_f \leftarrow$ mean(FFT($T(j,:)$), $freq$)/$(2 * pi)$
$Var_f \leftarrow$ Variance(A, $mean_f$, freq) $\leftarrow$**{ω-vector}**
$tbp_{gen} \leftarrow var_t * var_f \leftarrow${$u$-**vector**}
end for
return tbp_{gen}



(a)

(b)

Fig. 2: Spectrograms of (a) healthy *vs.* (b) pathological cries.

TABLE I: Statistics of the Chillanto dataset used. After [21].

Class	Category	# Utterances
Healthy	Normal	507
	Hungry	350
	Pain	192
Pathology	Asphyxia	340
	Deaf	879

IV. Experimental Setup

A. Dataset Used

In this study, we use Baby Chillanto database, which was designed using recordings made by doctors and is the property of Mexico's NIAOE-CONACYT [21]. Each recording was segmented to make infant cry signals of 1 second duration each. Since the sampling rate of the cry signals provided in the dataset is not uniform, we resampled all the utterances at a sampling rate of 11.025 kHz. The statistical details of the dataset are shown in Table I.

B. Classifier Used

The experiments were performed using Gaussian Mixture Model (GMM) classifier, which is commonly used for infant cry classification [22], [23]. In this study, 512 mixture components are used to train the model. Further, Log-Likelihood Ratio (LLR) score is used to evaluate the test cry signal.

C. Baseline

We consider CQT feature as the baseline for this work [24], where for low frequency regions, CQT gives better frequency resolution. For this baseline, we performed experiments with 96 number of bins per octave, keeping f_{min} = 100 Hz.

V. Experimental Results

This work is performed using *10-fold* cross-validation on Baby Chillanto dataset. We performed the experiments by fine-tuning feature parameter such as window overlap and number of subband filters. To that effect, we first varied the window overlap with values as 10, 15 and 20 ms. Number of filters were kept constant. The obtained experimental results are presented in Table II. From Table II, it can be observed that the highest performance is achieved as 93.83% accuracy, obtained when window length, window overlap and number of subband filters are of 30 ms, 15 ms and 40 respectively. Further, the next set of experiments were performed by varying the subband filters and keeping window overlap constant. These fine-tuning were performed considering the two cases of non-cepstral, and cepstral u-vector. It should be noted that the non-cepstral u-vector (with 93.83% accuracy) performs better than its cepstral version (with 93.48% accuracy). It can also be observed from Table II that as the number of subband filters increases, the % classification accuracy decreases.

TABLE II: % Accuracy for non-cepstral and cepstral u-vector

Window Length	Window Overlap	# Filters	% Accuracy (non-cepstral)	% Accuracy (cepstral)
30	15	40	**93.83**	93.04
30	15	60	93.08	**93.48**
30	15	80	87.71	87.00
30	20	40	93.35	92.42

Next, we compare the performance of u-vector, t-vector and ω-vector with the CQT baseline in Table III. The comparison is done for both the cases of cepstral and non-cepstral features.

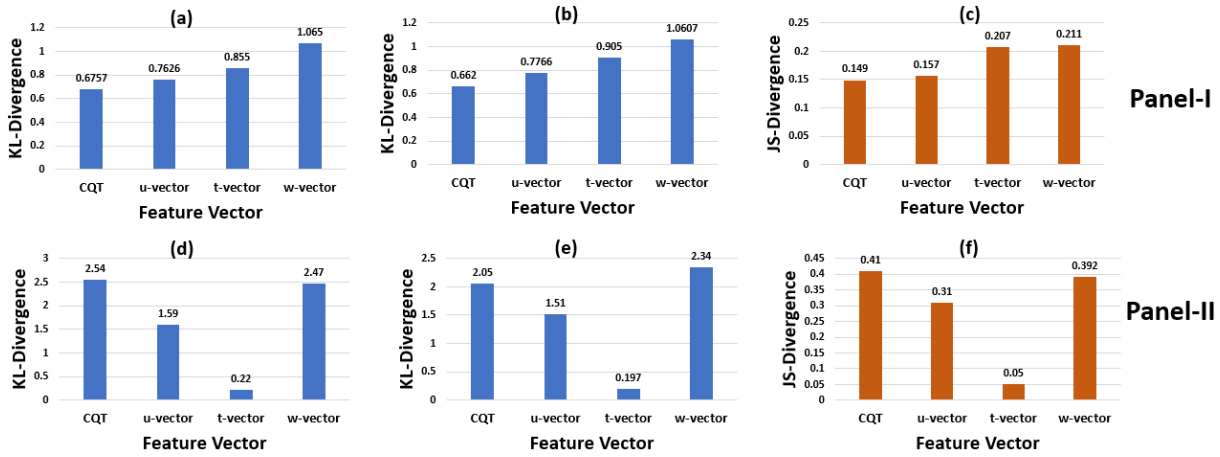


Fig. 3: KLD and JSD of the proposed feature sets. Panel-I and Panel-II denote the cases of non-cepstral features and cepstral features, respectively. KLD (healthy || pathology) is shown in (a) and (d). KLD (pathology || healthy) is shown (b) and (e). JSD (healthy || pathology) is presented in (c) and (f).

It can be observed that, the non cepstral ω-vector performs the best with % classification accuracy of 98.5% with overall increase of about 1.5% than the baseline CQT feature. Hence, it can be observed that the frequency distribution patterns of the different cry modes smeared over the entire frequency band, are captured by the ω-vector as discussed in [6]. Further, it can be observed that out of all the features shown in Table III, the best performance is achieved by ω-vector in the non-cepstral case with an accuracy of 98.50%. Furthermore, it should also be noted that the average overall accuracy of non-cepstral feature is higher than the cepstral features. In particular, the non-cepstral features achieve average higher accuracy (95.14%) as compared cepstral features. This indicates that non-cepstral features are better suited for pathology detection.

Given that ω-vector achieves the best performance in the non-cepstral domain, we performed the next set of experiments to observe the effect of number of subband filters in the ω-vector. Table IV presents the corresponding results, and it can be observed that the best result 98.50% is achieved with 40 number of subband filters. From Table IV, we can say that when the entire frequency band is divided into 40 subbands, the frequency variance captured in each subband is optimum for our binary classification task.

TABLE III: % Classification accuracy for various cepstral and non cepstral feature set

Non Cepstral Feature		**Cepstral Feature**	
Feature	**% Accuracy**	**Feature**	**% Accuracy**
u-vector	93.83	u-vector	93.48
t-vector	91.23	t-vector	89.38
ω-vector	**98.50**	ω-vector	96.74
CQT	97.00	CQT	**98.55**
Average	95.14	Average	94.53

TABLE IV: % classification accuracy of ω-vector with various number of subband filters

Subband Filters	40	60	80	100
% Accuracy	**98.50**	91.37	92.20	96.78

A. *Model-level measure of discriminative ability*

To estimate the model-level measure of discriminative ability w.r.t. various feature sets, we use KLD and JSD. If f and g are two PDFs, KLD is calculated as [25]

$$KLD(f||g) = -\int f(x) \ln \left\{ \frac{g(x)}{f(x)} \right\} dx. \qquad (10)$$

It is an *asymmetric* measure, i.e., $KLD(f||g) \neq KLD(g||f)$, since it does not satisfy the triangle inequality. To eliminate the asymmetry between two PDFs, we estimate another model-level measure called Jensen-Shannon Divergence (JSD) as:

$$JSD(f||g) = \frac{1}{2} KLD(f||m) + \frac{1}{2} KLD(g||m), \qquad (11)$$

where m is estimated as $\frac{1}{2}(f+g)$ [25]. Higher value of KLD and JSD signifies better discriminative ability of the model. Figure 3 shows KLD and JSD of non-cepstral and cepstral features in Panel-I and Panel-II, respectively. It can be observed that for the case of non-cepstral features as shown in Panel-I, ω-vector outperforms all the remaining features sets in terms of KLD as well as JSD. This shows that the better discriminative ability of our model is achieved when ω-vector is used. This is also reflected in the % accuracy results achieved, shown in Table III, where ω-vector outperforms all the remaining features in the case of non-cepstral features. Furthermore, for the case of cepstral features as shown in Panel-II, it can be observed that CQT shows better discriminative ability for the case of KLD (healthy|| pathology) and JSD (healthy|| pathology). However, it should be noted that ω-vector shows better performance for the case of KLD (pathology || healthy).

B. *Analysis of Latency Period*

In this study, we also investigate the latency period for t-vector, ω-vector, and u-vector w.r.t the baseline CQT feature set considered in this study. The latency is estimated by the performance evaluation in terms of % accuracy w.r.t. varying durations of speech segment in an utterance. The duration of the utterance ranges from 20 ms to 600 ms, with an interval of 150 ms. In this work, we have estimated KLD and JSD

on GMMs of 512 mixtures, for each of the feature vectors. Figure 4 shows comparison between non-cepstral features of CQT, u-vector, t-vector, and ω-vector. It can be observed that the ω-vector outperforms u-vector and t-vector, and shows remarkable latency as compared to the CQT. Moreover, it can be observed that all the three features, i.e., u-vector, t-vector, and ω-vector gave increased % accuracy in a short duration of speech utterance of $<$ 200 ms. On the other hand, CQT showed no improvement in accuracy even for a long duration of 600 ms of speech utterance. Additionally, the feature performance is better if for a low latency period the accuracy is high, which indicates the faster classification by the model and thus, indicating suitability for practical infant cry classification system deployment.

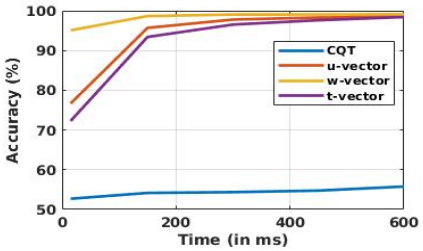


Fig. 4: Latency period *vs*. % accuracy between the various non-cepstral features for CQT, u-vector, t-vector, and ω-vector.

VI. Summary and Conclusions

In this study, u-vector (which is a combination of t-vector and ω-vector) is used to detect pathology from infant cries. These features are motivated from the Heisenberg's uncertainty principle in signal processing framework for classification of normal and pathological crying. To categorize the cries, the u-vector uses variances in time and frequency-domains, which correspond to t-vector and ω-vector. It is observed that ω-vector outperforms the remaining feature sets. This justifies the proposition that as compared to healthy cries, pathological cries have an irregular frequency dispersion across the entire frequency band. Our experiments also show that non-cepstral features are better suited for detection of pathological cries. Further, we have also demonstrated the discriminative capability of our models using information-theoretic measures, such as KL divergence and JS divergence. Given the early detection of pathology in infants is also associated with faster detection, we also analyzed the latency period performance of the proposed features and achieved relatively the best performance in latency.

VII. Acknowledgements

The authors would like to thank the Ministry of Electronics and Information Technology (MeitY), New Delhi, Govt. of India, for sponsoring consortium project titled 'Speech Technologies in Indian Languages' under 'National Language Translation Mission (NLTM): BHASHINI', subtitled 'Building Assistive Speech Technologies for the Challenged' (Grant ID: 11(1)2022-HCC (TDIL)). We also thank the consortium leaders Prof. Hema A. Murthy, Prof. S. Umesh, and the authorities of DA-IICT Gandhinagar, India for their support and cooperation to carry out this research work.

References

[1] J. J. Engelsma, D. Deb, K. Cao, A. Bhatnagar, P. S. Sudhish, and A. K. Jain, "Infant-ID: Fingerprints for global good," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[2] H. A. Patil, "Infant identification from their cry," in *2009 Seventh International Conference on Advances in Pattern Recognition, Kolkata, India*. IEEE, 2009, pp. 107–110.

[3] C. C. Onu, I. Udeogu, E. Ndiomu, U. Kengni, D. Precup, G. M. Sant'Anna, E. Alikor, and P. Opara, "Ubenwa: Cry-based diagnosis of birth asphyxia," *arXiv preprint arXiv:1711.06405*, 2017 {Last Accessed: 01-Feb-2022}.

[4] C. C. Onu, J. Lebensold, W. L. Hamilton, and D. Precup, "Neural transfer learning for cry-based diagnosis of perinatal asphyxia," *ICLR Workshop, Graz, Austria*, 2019.

[5] K. Manickam and H. Li, "Complexity analysis of normal and deaf infant cry acoustic waves," in *Fourth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, Italy*, 2005.

[6] Q. Xie, R. K. Ward, and C. A. Laszlo, "Determining normal infants' level-of-distress from cry sounds," in *Proceedings of Canadian Conference on Electrical and Computer Engineering, Vancouver, BC, Canada*, 1993, pp. 1094–1096.

[7] H. A. Patil, ""cry baby": Using spectrographic analysis to assess neonatal health status from an infant's cry," in *Advances in Speech Recognition*. Springer, 2010, pp. 323–348.

[8] S. G. Mallat, *A Wavelet Tour of Signal Processing*. Elsevier, 2^{nd} Edition, 1999.

[9] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. *I.* fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 520–538, 1992.

[10] Boashash, Boualem, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*. Academic Press, Second Edition, 2015.

[11] L. Armbrüster, W. Mende, G. Gelbrich, P. Wermke, R. Götz, and K. Wermke, "Musical intervals in infants' spontaneous crying over the first 4 months of life," *Folia Phoniatrica et Logopaedica*, vol. 73, no. 5, pp. 401–412, 2021.

[12] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*. Prentice Hall, 2^{nd} Edition, 1997.

[13] P. Busch, T. Heinonen, and P. Lahti, "Heisenberg's uncertainty principle," *Physics Reports*, vol. 452, no. 6, pp. 155–176, 2007.

[14] D. Gabor, "Theory of communication-part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[15] M. D. Beecher, "Spectrographic analysis of animal vocalizations: Implications of the "uncertainty principle," *Bioacoustics*, vol. 1, no. 2-3, pp. 187–208, 1988.

[16] L. Cohen, *Time-Frequency Analysis*. 1^{st} Edition, Prentice-Hall, 1995, vol. 778.

[17] B. Boashash, "Time-frequency and instantaneous frequency concepts (chapter 1)," in *Time-Frequency Signal Analysis and Processing (Second Edition)*. Oxford: Academic Press, 2016, pp. 31 – 63.

[18] M. R. Schädler and B. Kollmeier, "Separable spectro-temporal gabor filter bank features: Reducing the complexity of robust features for automatic speech recognition," *The Journal of the Acoustical Society of America (JASA)*, vol. 137, no. 4, pp. 2047–2059, 2015.

[19] S. G. Mallat, "Time Meets Frequency," in *A Wavelet Tour of Signal Processing*, 3^{rd} ed. Boston: Academic Press, 2009.

[20] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Pearson Education, 3^{rd} edition, India, 2006.

[21] A. Rosales-Pérez, C. A. Reyes-García, J. A. Gonzalez, O. F. Reyes-Galaviz, H. J. Escalante, and S. Orlandi, "Classifying infant cry patterns by the genetic selection of a fuzzy model," *Biomedical Signal Processing and Control*, vol. 17, pp. 38–46, 2015.

[22] H. F. Alaie, L. Abou-Abbas, and C. Tadj, "Cry-based infant pathology classification using gmms," *Speech Communication*, vol. 77, pp. 28–52, 2016.

[23] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The det curve in assessment of detection task performance," National Inst of Standards and Technology Gaithersburg MD, Tech. Rep., 1997.

[24] J. C. Brown, "Calculation of a constant Q spectral transform," *The Journal of the Acoustical Society of America (JASA)*, vol. 89, no. 1, pp. 425–434, 1991.

[25] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans on Info theory*, vol. 37, no. 1, pp. 145–151, 1991.