Constant Q Cepstral Coefficients for Classification of Normal *vs.* Pathological Infant Cry

Authors: Hemant A. Patil, Ankur T. Patil, Aastha Kachhi

@ ICASSP-2022

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar.

April 20, 2022

Outline

- Significance of Infant Cry Analysis
- Previous Studies on Infant Cry
- STFT *vs.* Constant-Q Transform (CQT)
- Proposed CQCC Feature Set
- Form Invariance Property of CQT
- Experimental Setup
- Spectrographic Analysis
- Experimental Results
- Summary and Conclusion

Significance of Infant Cry Analysis

- Challenges
  - Highly interdisciplinary in nature.
  - Death rate of infants.
  - Sudden Infant Death Syndrome (SIDS) and asphyxia.
  - Visual symptoms delays treatment.
  - High cost for clinical diagnosis of asphyxia.
- Infant Cry Analysis: Cost effective and non-invasive tool.

Previous Studies on Infant Cry

- Identification of pain, hunger, birth, and pleasure based on infant cry [1].
- Identification of the 10 distinct *cry modes* based on the spectrographic analysis [2].
- Identification of dysphonation and hyperphonation based on the spectrographic analysis [3].
- MFCC-GMM system for pathological infant cry classification [4,5].

Source:

1. O Wasz-Hockert, T.J. Partanen, V. Vuorenkoski, K Michelsson, and E Valanne, "The identification of some specific meanings in infant vocalization," Experientia, vol. 20, no. 3, pp. 154–154, 1964.
2. Qiaobing Xie, Rabab K. Ward, and Charles A Laszlo, "Automatic assessment of infants' levels-of-distress from the cry signals," IEEE Transactions on Speech and Audio Processing, vol. 4, no. 4, pp. 253, 1996.
3. Hemant A. Patil, "Cry baby": Using spectrographic analysis to assess neonatal health status from an infant's cry," in A. Newtein (Ed.) Advances in Speech Recognition, Springer, pp. 323–348. 2010.
4. Hesam Farsaie Alaie, Lina Abou-Abbas, and Chakib Tadj, "Cry-based infant pathology classification using GMMs," Speech Communication, vol. 77, pp. 28–52, 2016.
5. Chunyan Ji, Thosini Bamunu Mudiyanselage, Yutong Gao, and Yi Pan, "A review of infant cry analysis and classification," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2021, no. 1, pp. 1–17, 2021.

STFT *vs.* Constant-Q transform (CQT)

- Linearly- *vs.* geometrically-spaced frequency bins [1].
- Fixed *vs.* varying time-frequency resolution.
- CQT possesses form-invariance property *w.r.t.* linear time scaling.
- STFT is undersampled at low frequency regions.
- Parameters of the CQT can be fine-tuned to focus upon desired low frequency regions.

Source:

1. Judith 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.

Proposed CQCC feature set

- For discrete-time input speech signal $x(n)$, STFT is expressed as:

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n) \cdot \mathrm{w}(n, \tau) \cdot e^{-j\omega n}, \tag{1}$$

  where $\mathrm{w}(n, \tau)$ represents the analysis window, centered at time τ.
- For a frame of the speech signal $y(n)$, DFT $Y(k)$ represented as:

$$Y(k) = \sum_{n=0}^{N-1} y(n) \cdot e^{-j(\frac{2\pi}{N})kn}, \tag{2}$$

  where k is the frequency bin index, and $\omega_{DFT} = (2\pi k)/N$.

Proposed CQCC feature set

- The CQT of a signal $y(n)$ is represented as [5]:

$$Y^{CQT}(k) = \frac{1}{N(k)} \sum_{k=0}^{N(k)-1} y(n)\mathrm{w}(n,k)e^{-j\left(\frac{2\pi}{N(k)}Qn\right)}, \tag{3}$$

where $\omega_{CQT} = (2\pi Qn)/N(k)$, and Q is quality factor.

- The quality factor (Q) is ratio of center frequency (f_k) to the bandwidth (Δf_k):

$$Q = \frac{f_k}{\Delta f_k} = \frac{f_k}{f_{k+1} - f_k} = \frac{1}{2^{1/B} - 1}, \tag{4}$$

where B represents the number of bins per octave.

Proposed CQCC feature set

- Furthermore,

$$f_k = (2^{(k-1)/B}) f_{min}, \tag{5}$$

where f_{min} is the minimum frequency of the signal.
Thus, CQT produces geometrically-spaced frequency bins as shown in Table 1.

Table 1: Window length in samples as a function of analysis frequency (f_k). After [1].

k	Frequency (Hz)	# Samples	Duration (in ms)
1	100	29547	1340
100	204.37	14457	655.64
200	420	7022	318.48
400	1783	1657	75.15
600	7556	391	17.73

Source:

1. Judith 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.

Proposed CQCC feature set

- Length of w(n, k) varies in CQT w.r.t. frequency as:

$$N(k) = Q(F_s/f_k) \tag{6}$$

where Q is quality factor.

- Resampling from geometrically-spaced to linearly-spaced frequency scale [1].
- Discrete Cosine Transform (DCT) on resampled CQT-gram produces CQCC.

Source:

1. Massimiliano Todisco, H ´ector Delgado, and Nicholas Evans, "Constant-Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," Computer Speech & Language,2017 Bilbao, Spain, vol. 45, pp. 516–535, Bilbao, Spain, June 21-24, 2017.

Form Invariance Property of CQT

- Let STFT of the signal $f(t)$ is given as:

$$F(t,\omega) = \int_{-\infty}^{t} f(\tau)\mathrm{w}(t-\tau)e^{-j\omega\tau}d\tau. \tag{7}$$

- For form-invariance of STFT, we must have

$$F(t,\omega) = \gamma F(\alpha t, \beta\omega), \tag{8}$$

where α and β are scaling factor for time and frequency, respectively.
- However, it achieves the unstable filter characteristics.
- For CQT,

$$F(t,\omega) = \int_{-\infty}^{t} f(\tau)\mathrm{w}(t-\tau,\omega)e^{-j\omega\tau}d\tau, \tag{9}$$

- Due to variable length window, stability condition for LTI filter is achieved.

Experimental Setup

- Dataset Used: Baby Chillanto database.

Table 2: Statistics of the Baby Chilanto dataset.

Class	Category	# Samples
Healthy	Normal	507
	Hunger	350
	Pain	192
Pathology	Asphyxia	340
	Deaf	879

- Feature Sets: Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC), Cepstrals derived from STFT, and proposed CQCC.
- Classifier:
  - Gaussian Mixture Models (GMMs).
  - Support Vector Machine (SVM).
- Evaluation Strategy: *10*-fold cross-validation.

Spectrographic Analysis

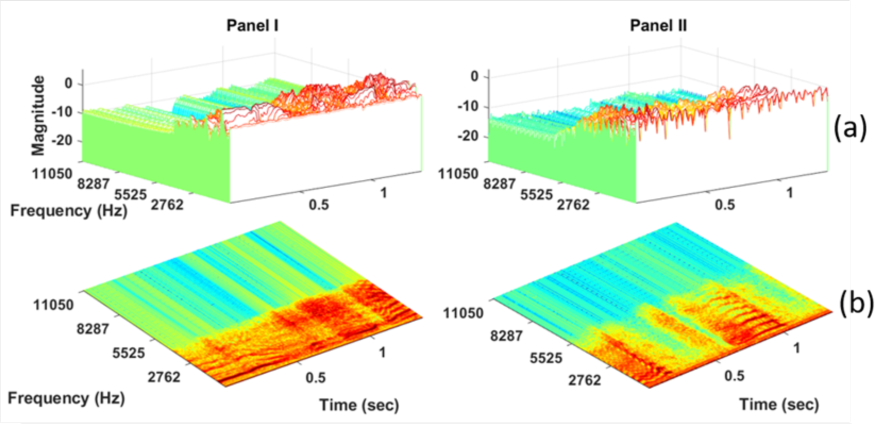

Figure 1: Panel-I and Panel-II depicts the spectrographic analysis for healthy (normal) and pathology (asphyxia) infant cry signal: (a) the waterfall plot for STFT, (b) the top view of the STFT waterfall plot.

Spectrographic Analysis

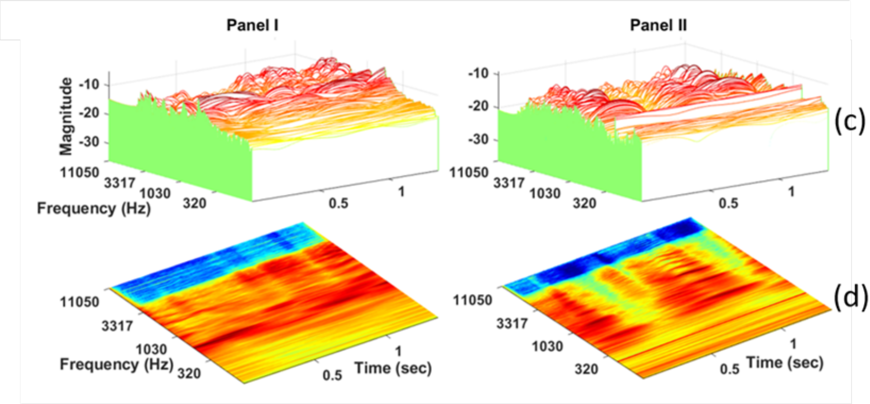

Figure 2: Panel-I and Panel-II depicts the spectrographic analysis for healthy (normal) and pathology (asphyxia) infant cry signal: (c) waterfall plot for CQT, and (d) the top view of the CQT waterfall plot.

Results

- Results with varying f_{min}:

Table 3: Results in % classification accuracy (Acc) for various f_{min} (Hz) of using GMM.

f_{min}	Acc.	f_{min}	Acc.	f_{min}	Acc.	f_{min}	Acc.
5	98.7	10	99.4	20	98.2	50	99.1
100	**99.8**	150	98.8	200	98.6	250	98.9

- Results for various window functions:

Table 4: Results (in % Classification Accuracy) for various window functions using GMM.

Window	Acc.	Window	Acc.
Hanning	**99.82**	Hamming	99.60
Gaussian	98.81	Rectangular	97.75

Results

- Results for varying number of Gaussian mixtures:

Table 5: Results (in % classification accuracy) w.r.t. number of mixtures.

Mixtures	64	128	256	512	1024
Accuracy	97.53	99.43	98.94	**99.82**	98.67

- Results for various feature sets:

Table 6: Results in (% classification accuracy and % EER) for various feature sets using GMM as a classifier.

		MFCC	LFCC	Cepstrals	CQCC
GMM	Acc.	98.55	98.28	98.68	**99.82**
	EER	1.23	0.50	0.47	**0.44**
SVM	Acc.	88.11	80.18	80.62	**91.19**
	EER	12.72	18.78	17.73	**6.38**

Summary and Conclusions

- Form invariance property is the key for better performance.
- Parameters of the CQT can be tuned to emphasize desired low frequency region.
- CQCC performs better over MFCC, LFCC, and Cepstrals feature sets.

Acknowledgments

- Authorities of DA-IICT .
- Organizers of Baby Chillanto datbase for providing the database.

Selected References

1. O Wasz-Hockert, T.J. Partanen, V. Vuorenkoski, K Michelsson, and E Valanne, "The identification of some specific meanings in infant vocalization," Experientia, vol. 20, no. 3, pp. 154–154, 1964.
2. Qiaobing Xie, Rabab K. Ward, and Charles A Laszlo, "Automatic assessment of infants' levels-of-distress from the cry signals," IEEE Transactions on Speech and Audio Processing, vol. 4, no. 4, pp. 253, 1996.
3. Hemant A. Patil, "Cry baby": Using spectrographic analysis to assess neonatal health status from an infant's cry," in A. Newtein (Ed.) Advances in Speech Recognition, Springer, pp. 323–348. 2010.
4. Hesam Farsaie Alaie, Lina Abou-Abbas, and Chakib Tadj, "Cry-based infant pathology classification using GMMs," Speech Communication, vol. 77, pp. 28–52, 2016.
5. Judith 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.
6. Massimiliano Todisco, H´ector Delgado, and Nicholas Evans,

Thank You