UC San Diego

Final Project Presentation ECE 251C

Ankur Sharma

Wavelet analysis of Hindustani Music Ragas

(And a wavelet based cochlea)

Problem statement

1. Assessment of wavelet based signal processing for audio analysis and developing a wavelet based silicon cochlea.
2. Verifying that wavelet based signal processing for Indian Classical music ragas yields better results than STFT
3. Research into analysis and classification of Indian Classical Music is far behind Western Classical Music.
4. The properties of Indian classical music that make it interesting for my project is the room for experimentation and improvisation during performances.
5. If wavelet based MFCCs are able to detect Indian Classical Music ragas better than STFT, it would signify that a wavelet based approach is perhaps better for silicon cochlea.

Why Mel Scale?

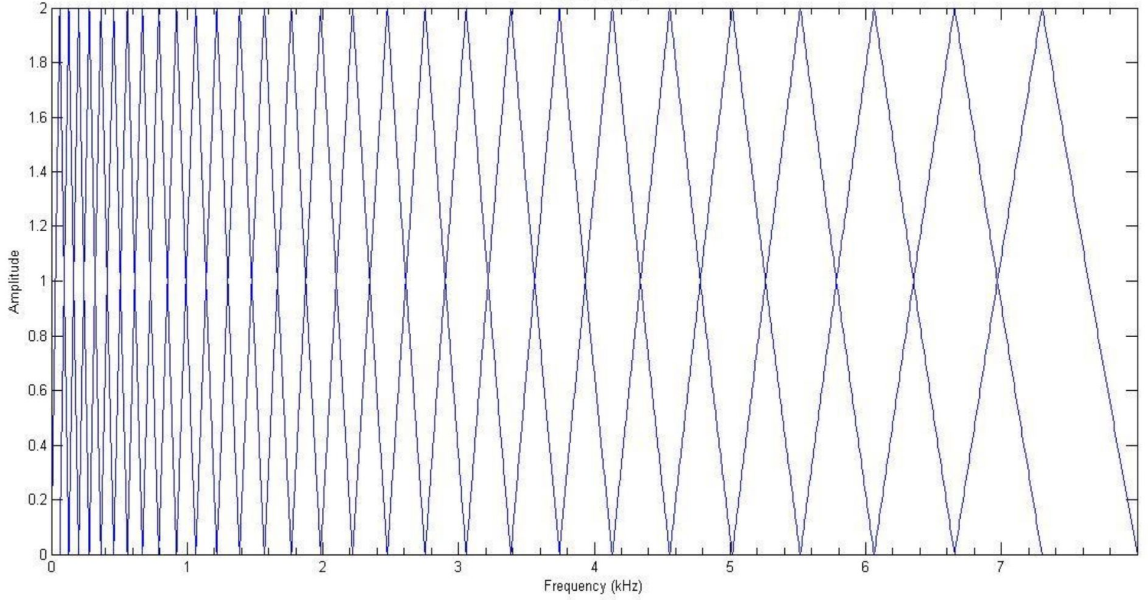


Fig. 1. Mel filter bank

- Mel is a logarithmic scale that fits the way humans perceive audio.
- Mel scale has been experimentally derived

$$F_{mel} = 2595 * log_{10}\left(1 + \frac{f_{linear}}{700}\right)$$

What is Bark Scale?

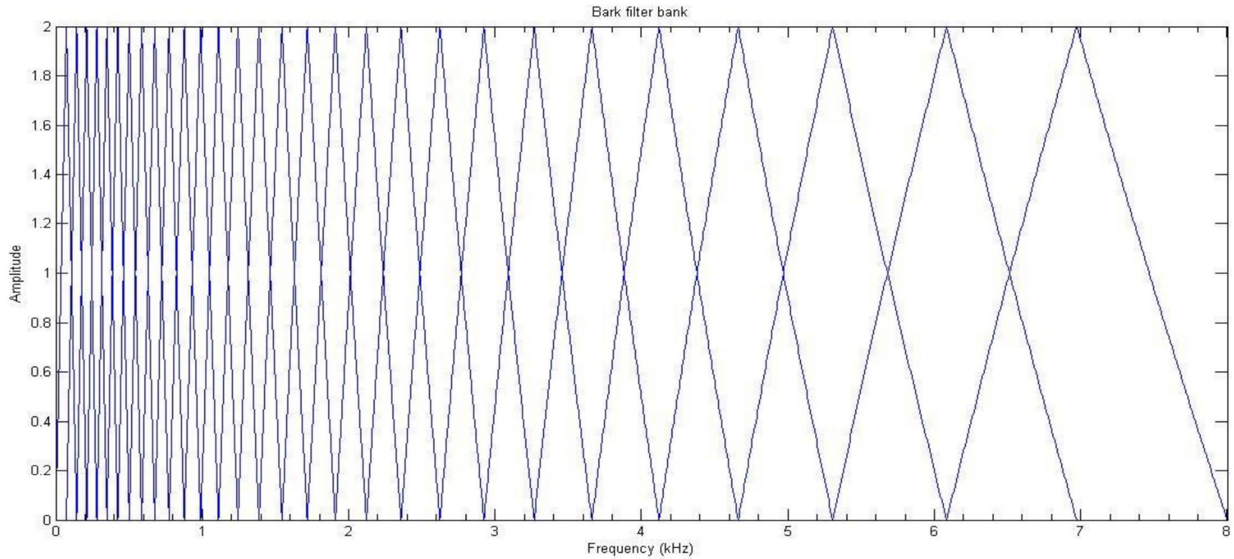


Fig. 2. Bark filter bank

$$F_{bark} = 13tan^{-1}\left(0.76\frac{f_{Hz}}{1000}\right) + 3.5tan^{-1}\left(\frac{f_{Hz}}{7500}\right)^2$$

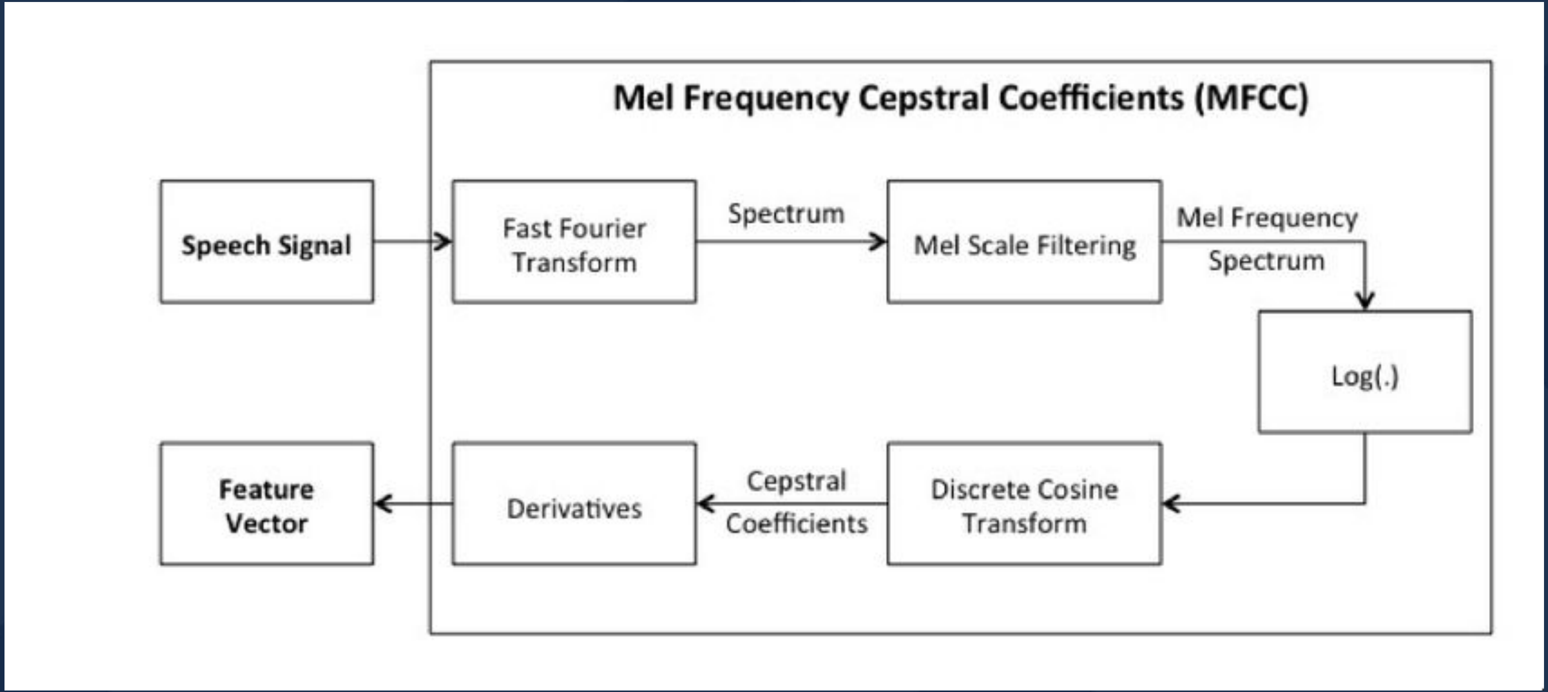
Tranmuller, 1990:

$$F_{bark} = \left[\frac{26.81f}{1960 + f}\right] - 0.53$$

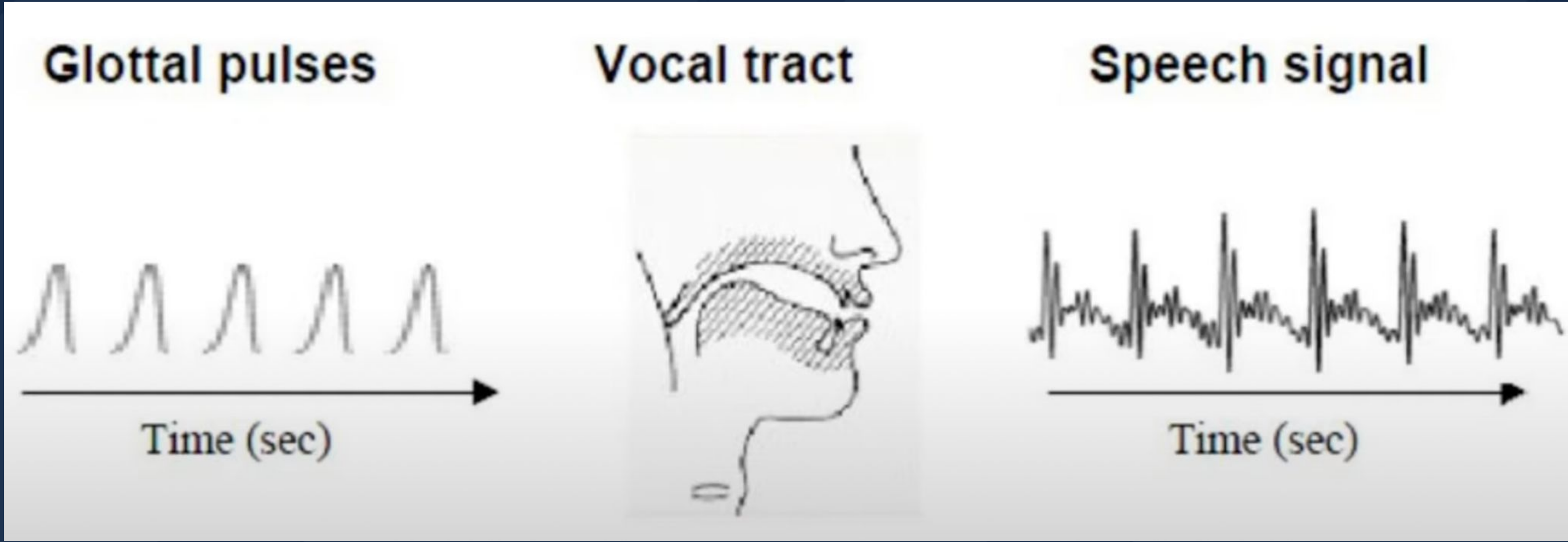
Wang, Sekey and Gersho, 1992:

$$F_{bark} = 6sinh^{-1}\left(\frac{f_{Hz}}{600}\right)$$

How to extract Mel coefficients?



Speech Generation



Speech = Convolution of vocal tract frequency response with glottal pulse

Current methods and state of the art

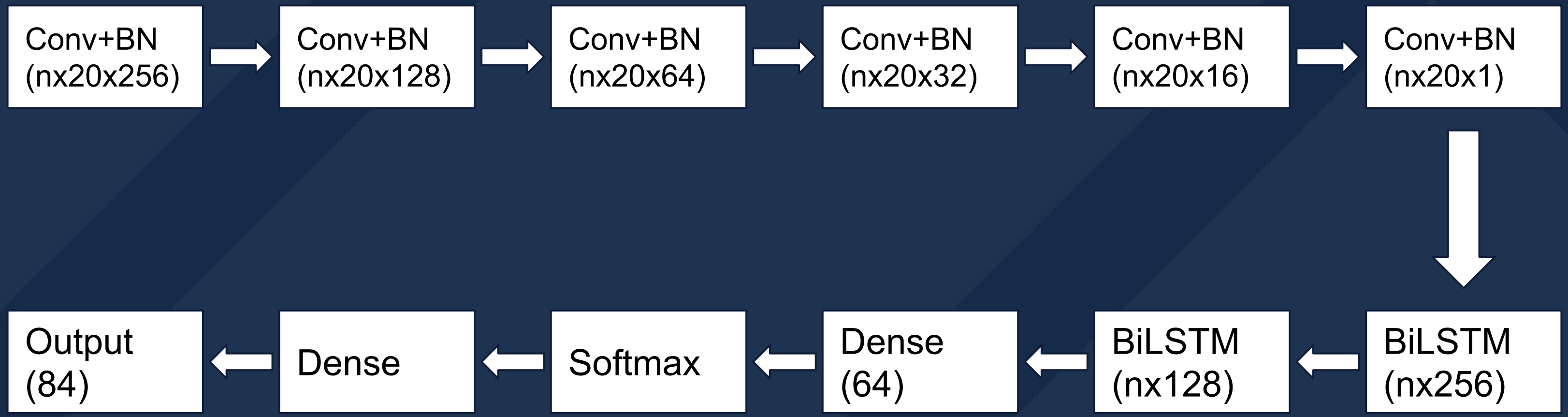
- Current methods include extraction of MFCCs using STFT and using them to characterize and identify the musical piece.
- State of the art: “Shazam”, which is a very successful application of audio fingerprinting, uses STFT and MFCC/spectral features. It then hashes the hamming distance between the spectral peaks and matches it with the input song.
- RNN-LSTM based approaches have also been used for time series matching to identify Indian Classical music.
- Wavelet processing has not been used at a commercial scale for audio fingerprinting because of the overhead of processing involved. However, the accuracy has certainly been higher for features extracted from discrete wavelet transform.
- A huge drawback of current methodologies is the fact of not treating the different music composition styles when dealing with classification problems.
- The problem with identification of Indian Ragas and fingerprinting them: The variation in each recording varies vastly and needs to be approached by characterizing the transitions between frequencies and taking into account the concept of tonality as well.

Steps in processing

1. Arrive at a working model for audio classification based on the current state of the art using FSD and MNIST database.
    a. LSTM+CNN
    b. BiDirectional LSTMs+CNN
    c. Transformer Models
    d. Siamese Models
2. Start processing the data from saraga1.5 Hindustani Music Dataset
3. Split the data into train, validation, and testing
4. List all wavelets to test initially.
5. Extract MFCCs and create a multi dimensional array to store the data for each audio segment.
6. Choose the ones that perform the best from each type and try different levels of wavelet processing
7. Analyze the result and make note of improvements in any future work.

Methodology

- Dataset:
  - Using Saraga1.5 Hindustani Music dataset
  - It has 84 different ragas
  - Each raga has audio worth 1 hour.
- Model:
  - Using a bidirectional LSTM model
  - Below is the model I used



Results

Results for different wavelets at level=5

Wavelet	Training accuracy	Testing accuracy
bior1.1	95.44%	95.17%
coif10	96.01%	97.02%
db10	96.48%	95.21%
rbio1.5	96.14%	97.26%
sym4	95.37%	96.57%
bior1.5	96.39%	97.59%
coif3	96.26%	96.23%
db12	97.37%	96.55%
rbio2.4	96.19%	97.07%
sym6	94.15%	93.91%

Results for different wavelets at level=5

Wavelet	Training accuracy	Testing accuracy
bior2.4	89.20%	93.99%
coif6	96.38%	95.94%
db4	96.57%	97.60%
rbio6.8	96.64%	96.36%
sym9	96.01%	96.39%
bior3.5	96.36%	96.18%
coif8	95.85%	96.79%
db8	97.37%	96.55%
rbio1.3	95.36%	96.57%
sym16	95.73%	94.17%

Results for the 4 best wavelets at level=3 and level=7

Level = 3

Wavelet	Training Acc.	Testing Acc.
bior1.5	94.99%	94.60%
db4	95.43%	96.35%
coif10	95.26%	94.29%
rbio2.4	96.18%	96.48%

Level = 7

Wavelet	Training Acc.	Testing Acc.
bior1.5	96.07%	96.03%
db4	96.12%	96.81%
coif10	96.78%	97.07%
rbio2.4	96.11%	95.91%

Reference Value: Accuracy for STFT based classification

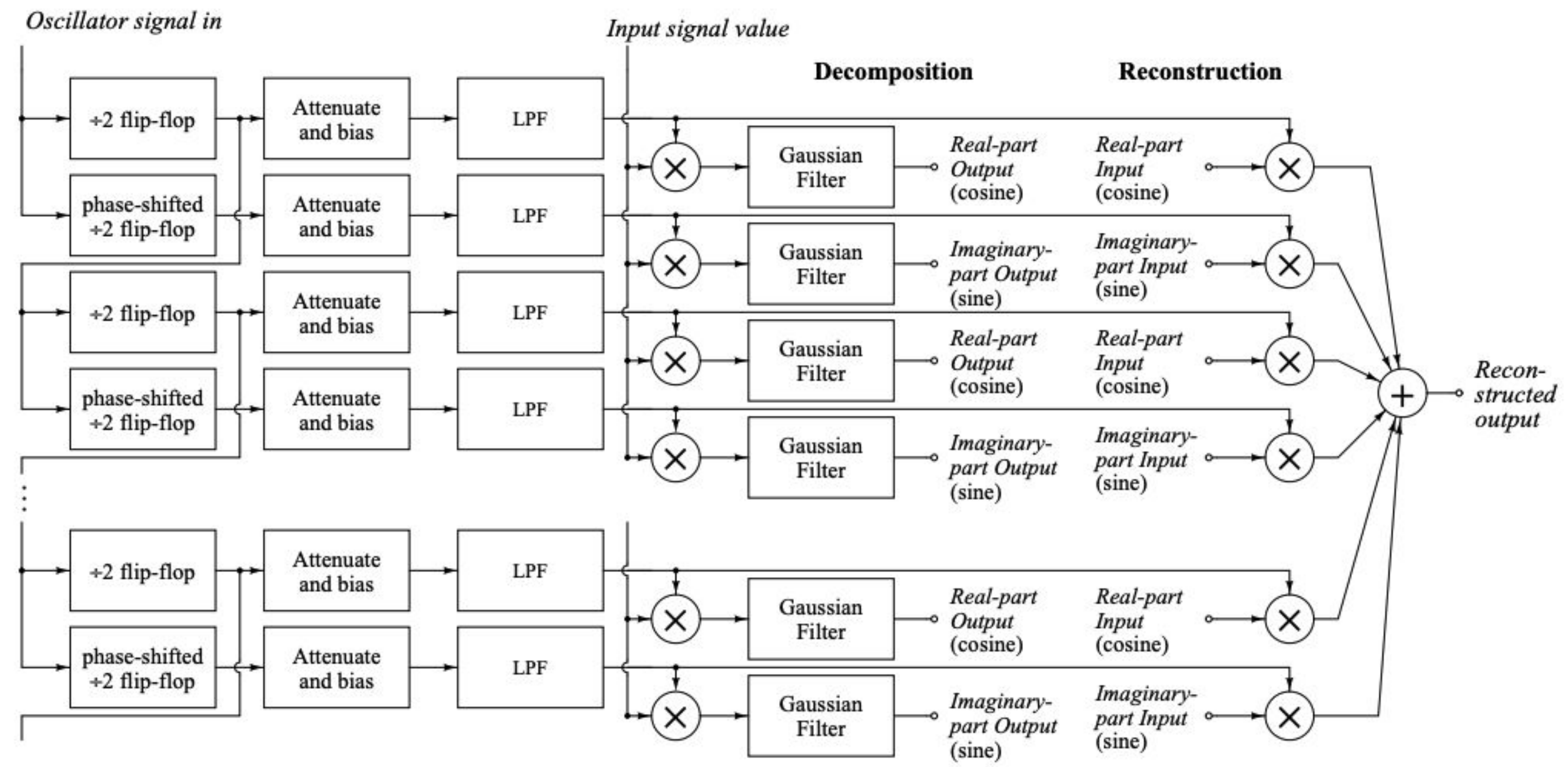
For the bidirectional LSTM model,

- Indian Music dataset:
  - Training accuracy: 36.24%
  - Testing accuracy: 38.13%
- FSD Dataset:
  - Training accuracy: 98%
  - Testing accuracy: 90%
- MNIST dataset:
  - Training accuracy: 99%
  - Testing accuracy: 97%

Takeaways

- Wavelet processing is indeed much better than the STFT based audio processing.
- The good performance on Indian classical music leads to the conclusion that wavelet processing is better for a hearing mechanism.
- The good performance of wavelet based approach can be attributed to the log spaced frequencies that are the core of psycho acoustic models like MFCCs and bark scales.

Future Plans



References

1. Edwards, R. Timothy, and Michael D. Godfrey. "An analog wavelet transform chip." IEEE International Conference on Neural Networks. IEEE, 1993.
2. M. I. Abdalla and H. S. Ali, “Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models,” vol. 1, no. 2, p. 6, 2010.
3. L.-C. Yang and A. Lerch, “On the evaluation of generative models in music,” Neural Comput & Applic, vol. 32, no. 9, pp. 4773–4784, May 2020, doi: 10.1007/s00521-018-3849-7.
4. R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, “History and future of auditory filter models,” in Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, France, May 2010, pp. 3809–3812. doi: 10.1109/ISCAS.2010.5537724.
5. A. K. Sharma et al., “Classification of Indian Classical Music With Time-Series Matching Deep Learning Approach,” IEEE Access, vol. 9, pp. 102041–102052, 2021, doi:10.1109/ACCESS.2021.3093911.