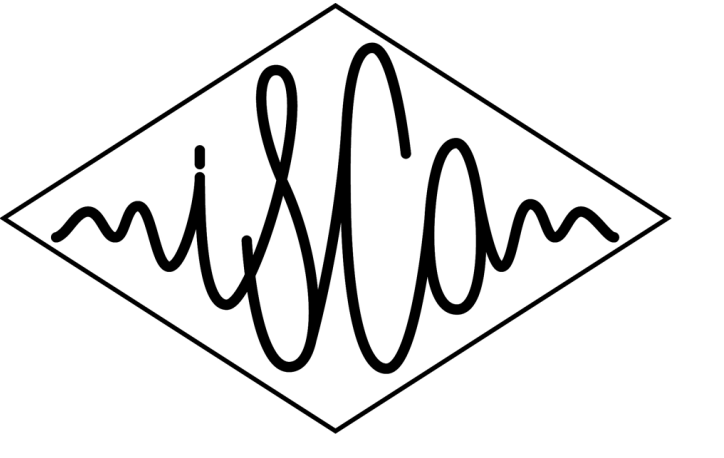




Music Tempo Estimation Using Sub-band Synchrony



Shreyan Chowdhury, Tanaya Guha, Rajesh Hegde
Indian Institute of Technology Kanpur, India

What makes us tap our feet to music?



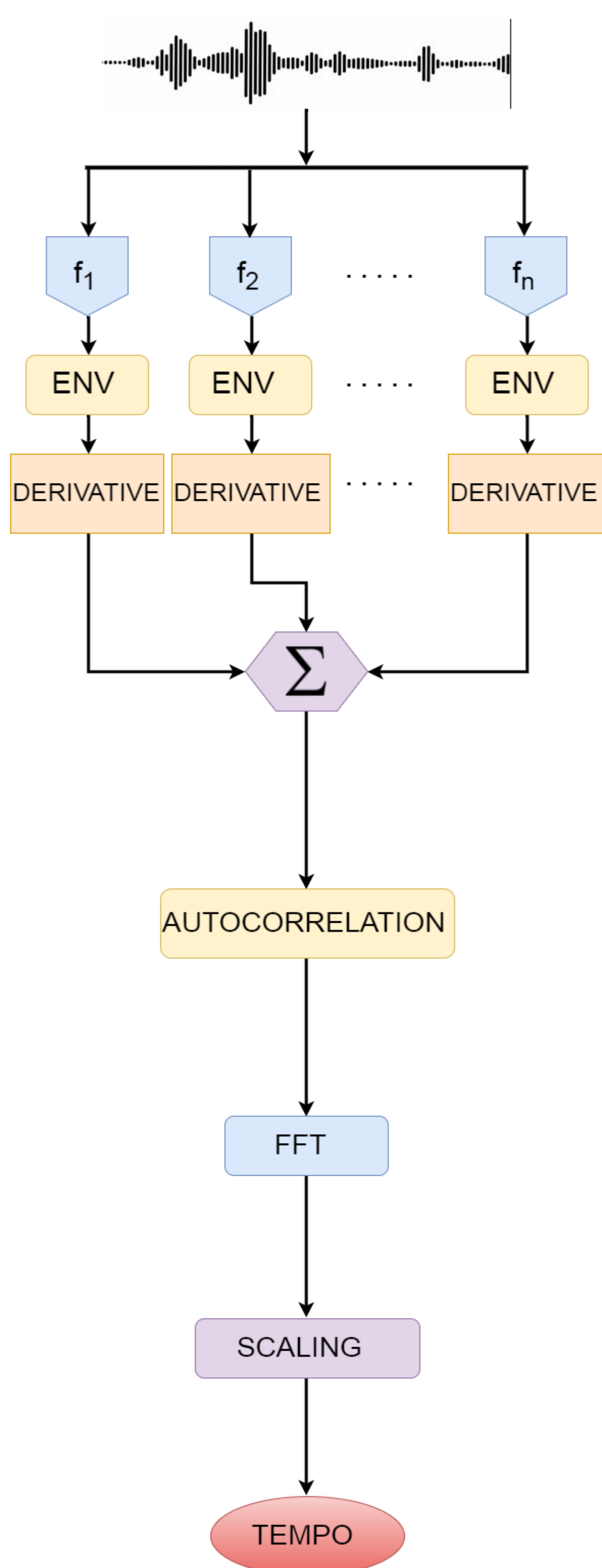
Humans can perform the seemingly complex task of comprehending rhythm and tempo from auditory information quite effortlessly.

- Approximate range: 40 – 300 BPM
- Periodicity in percussive events or harmonic changes
- Temporal salience as a feature to detect onsets

Objective: Develop a method for onset detection and tempo estimation that reflects the above stated observations.

Previous work

- Spectral flux and energy based onset detectors have been proposed in the past. Mel spectrogram and filter bank based tempo estimators also exist.
- Comb resonators have also been used in the past to estimate tempo.
- Recent methods have used neural networks to determine beat onset curve.
- Some of these works have been cited in the references at the bottom.



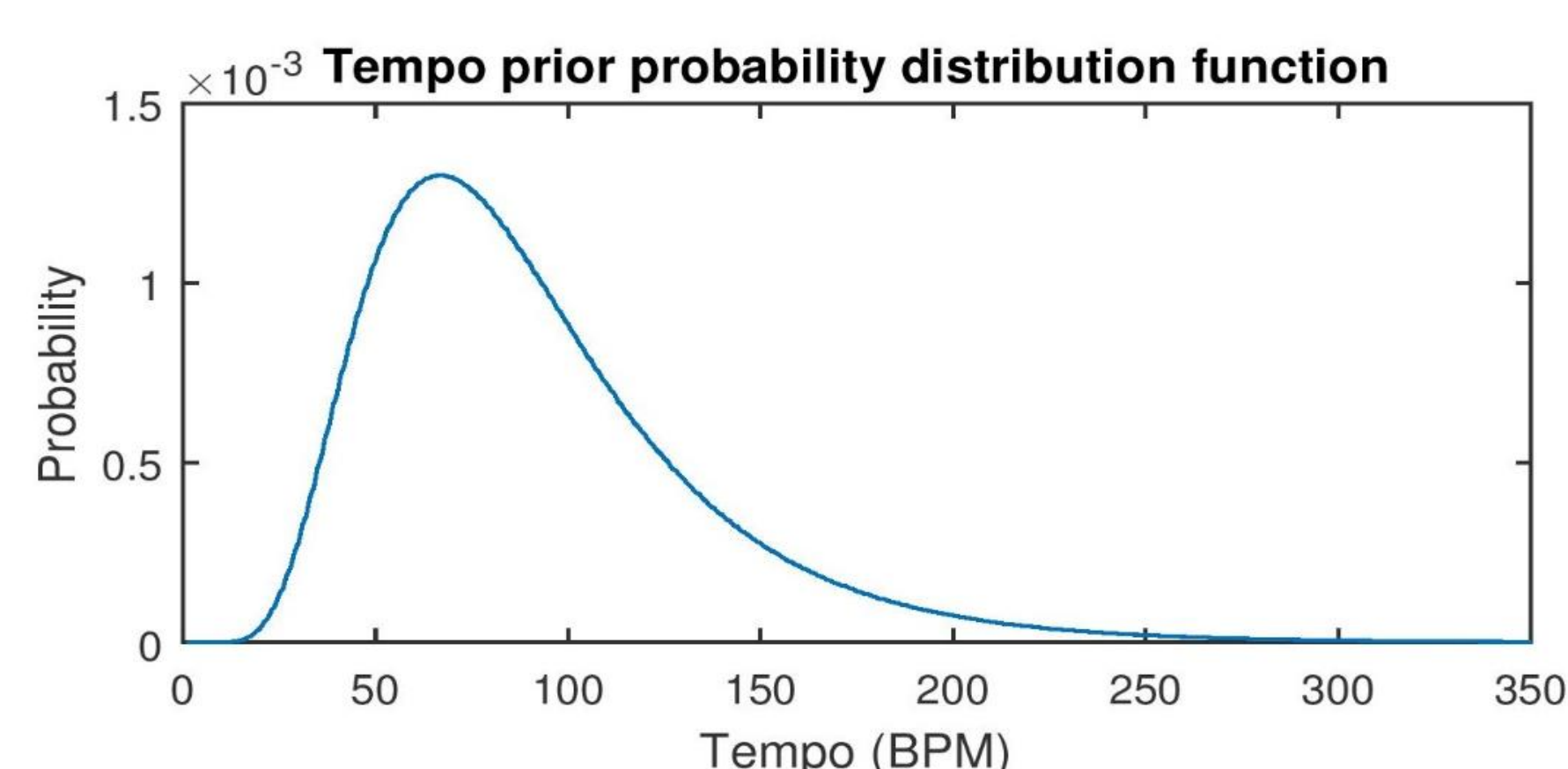
Sub-band synchrony for onset detection

Motivated by how humans perceive and identify onsets.

- **Sub-band synchrony:** A measure of temporal coherence of energy between frequency bands.
- Gammatone filter bank is used (to approximate human auditory filtering) with around 40 bands.
- **Changes in energy envelope:** Peaks in the derivative of energy envelope across bands = onset times.

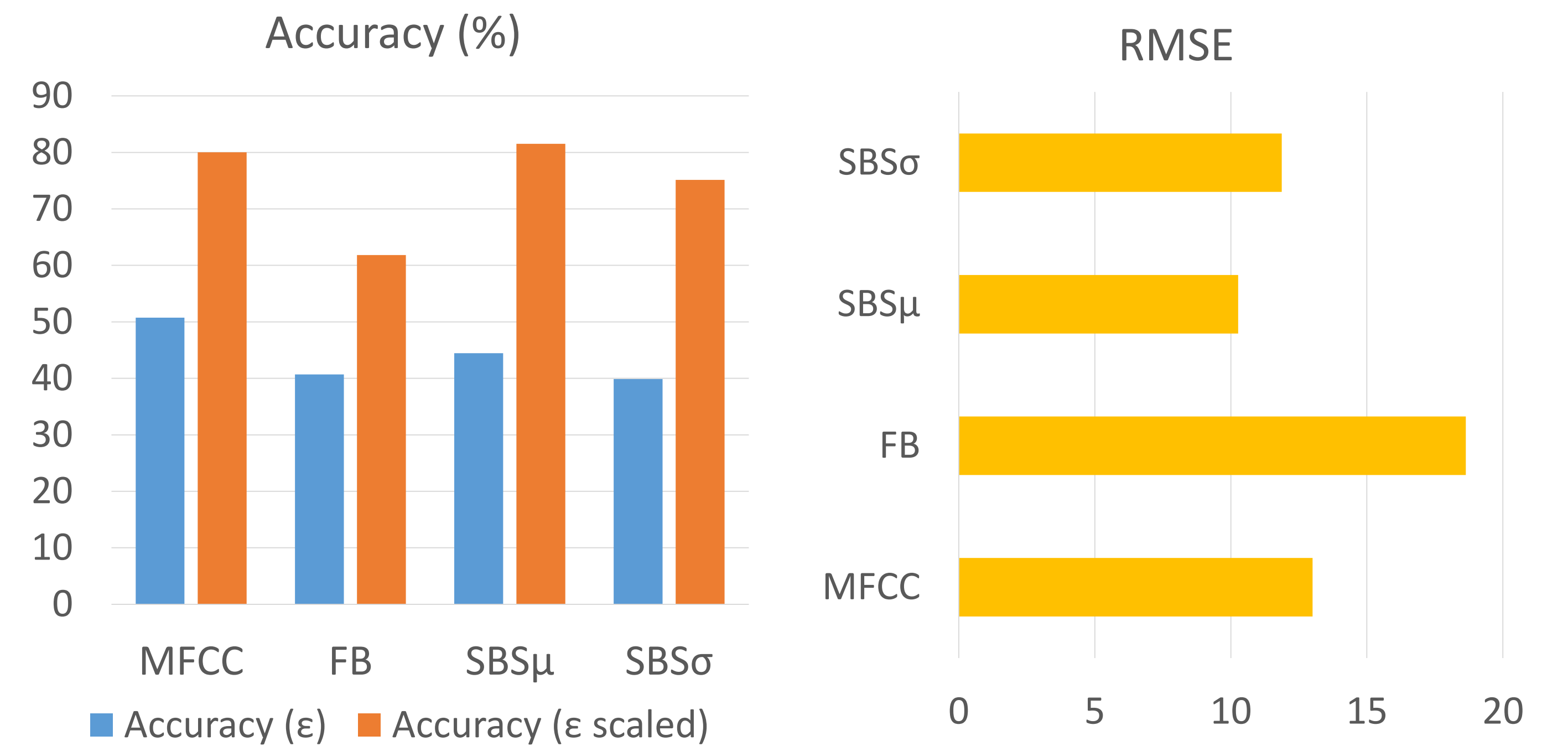
Periodicity and tempo induction

- Periodicity is extracted by calculating a frequency spectrum (FFT) of the autocorrelation of the onset curve.
- **Scaling:** to reflect the range of expected tempo for a human.
- Conversion to BPM; peak picking.



Scaling function: derived from a model of beat periods of actual songs.

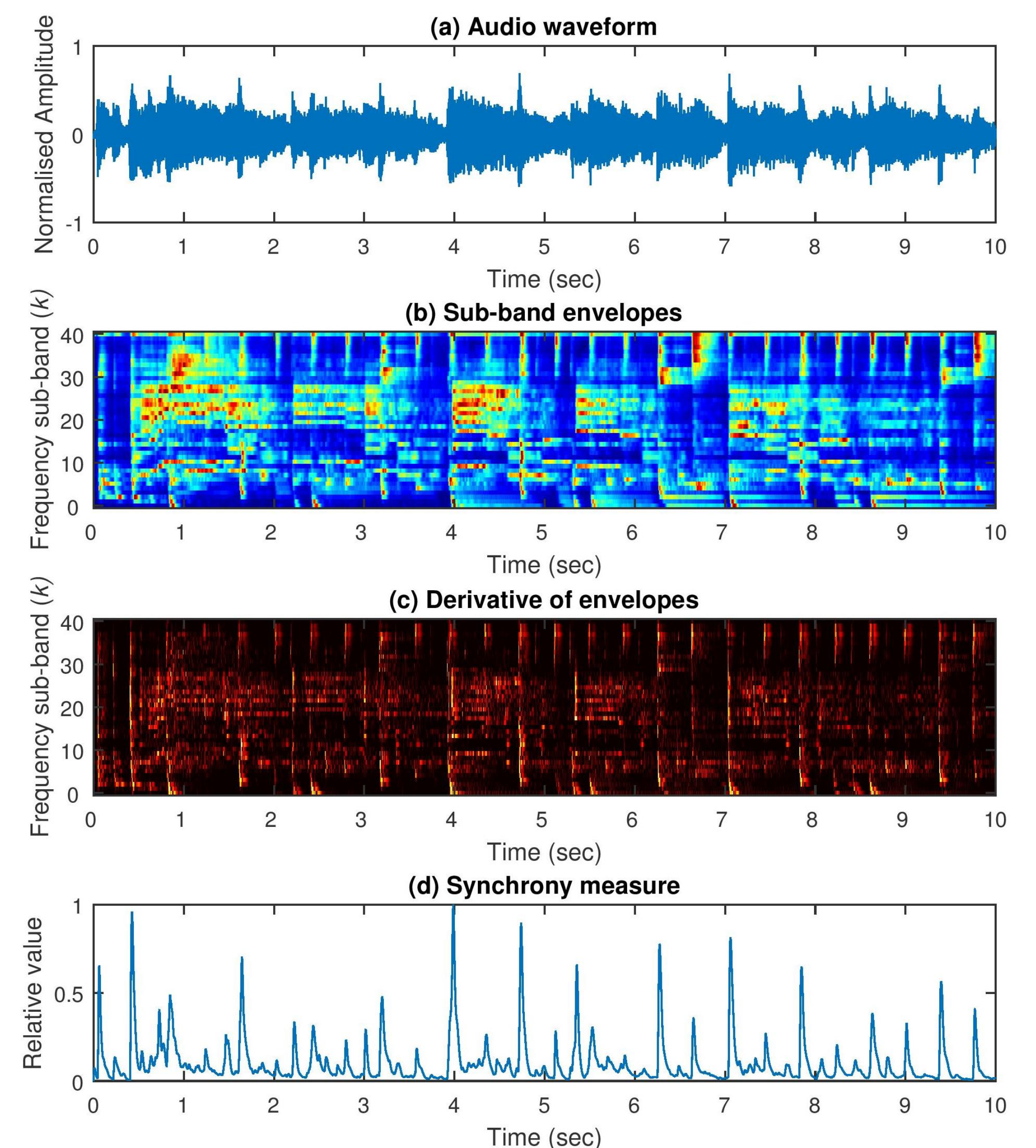
Results



- Three metrics used for evaluation:
- ϵ : percentage of estimates in a 4% band around the ground-truth.
- ϵ_{scaled} : takes into account octave deviations in estimates.
- RMSE: Root mean squared error from ground truth tempo.

Ground truth values were determined using the tapping experiment.

We find sub-band synchrony performs better than conventional spectral features that model human auditory perception of tempo.



Conclusion and future scope

- Sub-band synchrony is found to be a reasonable feature to compute onsets which give good tempo estimates for a wide variety of music.
- Both harmonic and percussive onsets are effectively detected.
- Future work will be focused on real-time tempo tracking and improving accuracy by implementing higher order statistical measures.

References

- R. Parncutt, "A perceptual model of pulse salience and metrical accent in musical rhythms," Music perception, pp. 409–464, 1994.
- S. Dixon, "Onset detection revisited," in Int. Conf. on Digital Audio Effects (DAFx-06), 2006, pp. 133–137.
- D. P. Ellis, "Beat tracking by dynamic programming," J. of New Music Research, vol. 36, no. 1, pp. 51–60, 2007.
- G. Percival and G. Tzanetakis, "Streamlined tempo estimation based on autocorrelation and cross-correlation with pulses," IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), vol. 22, no. 12, pp. 1765–1776, 2014.
- S. Bock, F. Krebs, and G. Widmer, "Accurate tempo estimation based on recurrent neural networks and resonating comb filters," in ISMIR, 2015, pp. 625–631.

Acknowledgments

I would like to thank Akademi Konferens and ISCA for inviting me to Interspeech 2017, and the reviewers for providing useful insight.

Contact information

Shreyan Chowdhury: shreyan0311@gmail.com