Musical Onset Detection on Carnatic Percussion Instruments

P A Manoj Kumar, Jilt Sebastian and Hema A Murthy
Indian Institute of Technology Madras
Chennai, India

***Abstract*—In this work, we explore the task of musical onset detection in Carnatic music by choosing five major percussion instruments : the mridangam, ghatam, kanjira, morsing and thavil. We explore the musical characteristics of the strokes for each of the above instruments, motivating the challenge in designing an onset detection algorithm. We propose a non-model based algorithm using the minimum-phase group delay for this task. The music signal is treated as an Amplitude-Frequency modulated (AM-FM) waveform, and its envelope is extracted using the Hilbert transform. Minimum phase group delay processing is then applied to accurately determine the onset locations. The algorithm is tested on a large dataset with both controlled and concert recordings (*tani avarthanams*). The performance is observed to be the comparable with that of the state-of-the-art technique employing machine learning algorithms.**

Keywords*—Onset detection, group delay functions, envelope detection

I. Introduction

Percussion instruments play an important role in many genres of world music. In addition to keeping track of rhythm, they have been developed to produce individual performances rich in artistic quality. The major percussive accompaniments to Carnatic (South Indian classical) music include the mridangam, ghatam, kanjira, morsing and the thavil. The choice between these instruments is made based on the rapport between the lead and accompanying artists and nature of the performance itself. The thavil, for example, is widely found in musical performances associated with traditional festivals and ceremonies as well as in professional musical concerts. Each of the instruments may be played along with the lead artist (usually a vocalist) or individually (*tani avarthanam*). A study of these instruments on the lines of Music Information Retrieval (MIR) will be aimed at first understanding the patterns involved in the performance - beat tracking, stroke classification, phrase/syllable classification *etc*. The information obtained can be further extended to study higher level problems such as tala classification, sama detection, artist and song recognition and so on.

We present our work in audio onset detection characterized as detecting relevant musical events in general, and as identifying the stroke instants in the case of percussive instruments. For a single note, Bello [1] defines an onset as the instant chosen to mark the transient, although in most cases it coincides with the start of the transient. Onset detection can be extended to polyphonic and ensemble performances too. Numerous basic algorithms and advanced improvisations including knowledge based post-processing have been proposed, but almost all onset detection algorithms consist of two tasks - extraction of the *detection function*, and a peak-picking algorithm. Approaches based on energy [2] and magnitude [3] report onsets as either instants of high energy or change in energy [4][5] in the *detection function*. Phase based approaches [6] were introduced to detect soft onsets followed by a combination of both energy and phase in [7]. Multi resolution analyses with varying window sizes have been tried to study both higher and lower frequency components[8]. Linear prediction error has also been used as a detection function in onset detection [9][10][11]. Recently, the state-of-the-art techniques employ recurrent [12] and convolutional [13] neural networks for training. Bello [1], Dixon [14] and Böck [15] analyze the task in detail and evaluate the major onset detection approaches over the years.

978-1-4799-6619-6/15/$31.00

The negative derivative of the phase spectrum, known as the group delay function, has been successfully used to extract formants, pitch, and features for speaker and speech recognition and spectrum estimation. Recently, segmentation of speech into syllable-like units using group delay functions has been studied for speech synthesis [16][17].

In this work, we extend the segmentation task to onset detection. The music waveform is treated as an amplitude and frequency modulated signal (AM-FM). First, the derivative of the signal is taken with respect to time which emphasizes the frequency content. Next, an envelope tracker algorithm using the Hilbert transform is implemented, which tracks the amplitude variations. Treating a smoothened version of this improvised envelope as a time domain signal, minimum phase group delay processing is applied and a global threshold is used to report the onsets. It must be mentioned that while group delay based onset detection algorithms have been attempted in the past [18][19][20], we have employed the high spectral resolution property of group delay functions for accurate location of onsets after emphasizing the amplitude and frequency characteristics. The algorithm is then evaluated on a large dataset of $\approx 17,000$ strokes consisting of *tani avarthanams* as well as ensemble (multiple instruments in a concert) recordings. Results are reported instrument wise and compared with a state-of-the-art onset detection technique [13].

The rest of the paper is organized as follows - Section II provides a brief overview of the instruments considered for evaluation. Section III describes the envelope detection and minimum phase group delay processing techniques used in the proposed algorithm. Sections IV and V discuss the experiments

and future work respectively.

II. Carnatic percussion instruments - A brief overview

A. Mridangam

The mridangam is the most widely used percussion accompaniment in Carnatic music. It resembles a two sided drum with tightly stretched membranes on either side, the two sides being unequal in size. The instrument is tuned by loading the smaller side with a black paste of adhesive mixed with finely ground special stone. In this context, the tonic is defined as the base pitch, which is used as a reference for all other higher harmonics. The strokes can be categorized based on the side of the mridangam being played and the position, manner and force with which the membranes are struck. However, the exact number of unique strokes varies across different schools of mridangam. The two sides allow composite strokes (individual strokes from the left and right side at the same instant) to be created, which from an MIR perspective ought to be treated as a single stroke, although they sometimes appear as separate strokes while performing the onset detection analysis. The first study on the mridangam carried out by the renowned scientist Raman [21] and later by Siddharthan [22] analyzed the harmonics of the strokes. More recently, Akshay [23] employed non-negative matrix factorization to classify the strokes.

B. Ghatam

The ghatam is a hollow pot that is placed on the lap of the artist and struck with the palm and fingers. The instrument is made of specially burnt clay with metallic powder for strength and care is taken that the walls of ghatam are of equal thickness. Distinct ghatam strokes count lesser in number than mridangam. Tuning of the pitch is possible to a limited extent by application of *play-doh*, but mostly multiples ghatams are chosen to achieve significant variations. Ghatam strokes produce a characteristic sound when struck on the neck of the pot. The artist modulates the sound by altering the size of the mouth during the performance, by partly or fully closing the area of the mouth with palms.

C. Morsing

Known as *Jew's Harp*, the morsing is a wind percussion instrument. It resembles a metallic clamp with a thin film (the *tongue*) in between. The instrument is caught by the hand and placed in the mouth of the artist, the teeth firmly holding it in place. Sound is produced inside the mouth of the artist by triggering the tongue of the instrument with the index finger. The artist's tongue is also used to produce morsing notes. Pitch of the intrument cannot be varied significantly and the artists prefer to carry morsings of different dimensions for the purpose of fine tuning.

D. Thavil

The thavil is similar to the mridangam in the sense that it is a two sided barrel, with both sides participating in sound production. The left side is struck with a stick while the artist plays the right side with fingertips covered with *thumb caps*. The *thumb caps* are mostly made of hardened rice flour and give rise to sharp, cracking sounds. Variations in pitch are achieved by tightening the left side of the instrument. Distinct strokes exist, based on the side of the instrument struck and the number of fingers involved in production (for the right side). For instance, *Ta* and *Di* involve four fingers, but are still treated as a single stroke by musicians.

E. Kanjira

The kanjira is a one-sided percussion instrument and is small enough to be held with one hand. The instrument is made of belly skin of monitor lizard stretched across a circular wooden frame made from the jackfruit tree. A high pitched sound is produced by striking the circular face with the palm and fingers of the free hand. Unlike the mridangam, the face of the kanjira is not loaded with any paste. The pitch can be varied to an extent by applying pressure on the face using the hand holding the kanjira or by sprinkling water on the kanjira skin from behind.

III. Proposed Method

We discuss the major concepts involved in the proposed algorithm -

A. Amplitude Frequency Modulation

In the context of communications, a message signal $m(t)$ is encoded with a high frequency carrier signal $s(t)$ before transmission. Modulation is performed to reduce transmitter and receiver antennae sizes, and to multiplex different message signals. Various modulation schemes exist, and are characterized by the influence of $m(t)$ on $s(t)$. In the case of amplitude-frequency modulation (AM-FM), both the amplitude and frequency of the carrier signal are influenced by message signals $m_1(t)$ and $m_2(t)$. Figure 1(*a*) presents an example using sinusoids in place of $m_1(t)$ and $m_2(t)$.

It is observed that most percussive strokes in Carnatic music can be modeled by an AM-FM signal, based on the variations in amplitude and frequency in the vicinity of an onset. Figure 1 illustrates this resemblance by comparing a carrier signal modulated using sinusoidal message signals with individual strokes of mridangam, ghatam, kanjira, morsing and thavil. In this context, we propose that the messages $m_1(t)$ and $m_2(t)$ contain information necessary to pinpoint the location of onsets. A demodulation technique is necessary to extract this information before proceeding to locate the onsets.

B. Demodulation and Envelope Detection

Consider a signal $x(t)$ that is amplitude-frequency modulated. The basic representation is given as:

$$x(t) = m_1(t)cos(\omega_c t + k_f \int m_2(t)dt) \tag{1}$$

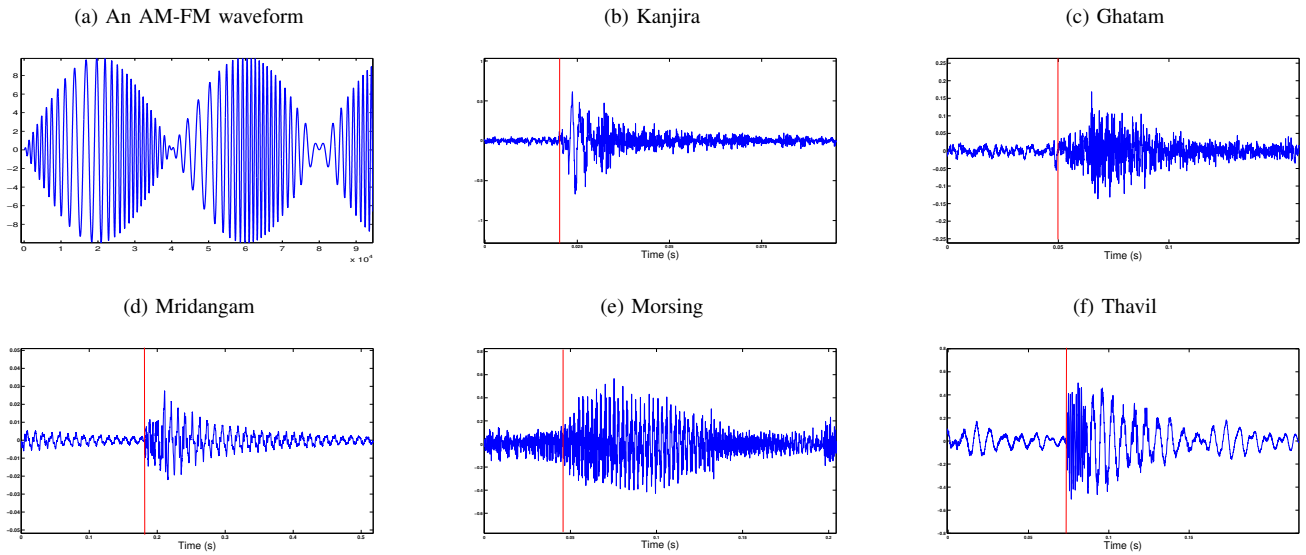

Fig. 1: Resemblance of Carnatic percussion strokes (*with ground truth marked*) to an Amplitude-Frequency modulated waveform

$m_1(t), m_2(t)$ represent the message signals, k_f is the frequency modulation constant and ω_c is the carrier frequency. Differentiating $x(t)$ with respect to time,

$$x'(t) \approx -e(t)sin(\omega_c t + k_f \int m_2(t)dt) \tag{2}$$

where

$$e(t) = m_1(t)(\omega_c + k_f m_2(t)) \tag{3}$$

The term $m_1'(t)cos(\omega_c t + k_f \int m_2(t)dt)$ has been ignored in (2) since ω_c can be assumed large. Both the message signals are now part of the amplitude in (2). We postulate that all the information about an onset is contained within the envelope function $e(t)$. $e(t)$ is extracted from $x'(t)$ using the Hilbert transform as follows:

Any real-valued signal $S(t)$ with Fourier transform $S(w)$ can be represented by its analytic version (as introduced by Gabor in 1946 [24]) and is given by

$$S_a(t) = 2\int_0^\infty S(w)exp(j2\pi wt)dw \tag{4}$$

As seen, it is the inverse Fourier transform of the positive frequency part alone. In terms of input signal $S(t)$,

$$S_a(t) = S(t) + iS_H(t) \tag{5}$$

where, $S_H(t)$ is the Hilbert Transform of $S(t)$. The real part of this analytic signal represents the actual signal and imaginary part is its Hilbert Transform. The magnitude of the analytic signal gives an estimate of the envelope.

C. Minimum phase group delay processing

Let $x[n]$ be a discrete-time signal, whose continuous phase spectrum is given by $\theta(\omega)$. The group delay function $\tau(\omega)$ is defined as

$$\tau(\omega) = -\frac{d(\theta(\omega))}{d\omega} \tag{6}$$

In general, the phase is wrapped at multiples of 2π, and hence it becomes difficult to infer meaningful information directly from the phase spectrum. An alternate form for computation from the magnitude spectrum exists:

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2} \tag{7}$$

where the subscripts R and I represent real and imaginary parts of the Fourier spectrum respectively. $X(\omega)$ and $Y(\omega)$ denote the discrete time Fourier transforms of $x[n]$ and $nx[n]$ respectively.

For a cascade of resonators, the group delay function exhibits high spectral resolution due to the additive property of phase - Figure 2 illustrates this phenomenon by considering two minimum phase systems, complex conjugate poles at (i) $(0.8e^{j0.1\pi}, 0.8e^{j0.3\pi})$ and (ii) $(0.8e^{j0.1\pi}, 0.8e^{j0.5\pi})$. The peaks are not resolved in the case of the magnitude spectrum for system (i). Moreover, the peak locations do not coincide exactly with the poles in both the systems, while the ability of group delay function to clearly differentiate between the poles is visible.

It was also shown that for minimum phase signals, the group delay function and the magnitude spectrum resemble each other [25]. This property, alongwith the high spectral resolution has since been used in group delay based feature extraction [26], [27], [28]

Zeroes outside the unit circle appear as peaks in the group delay domain. This makes it difficult to differentiate from poles inside the unit circle, which exhibit the same phenomenon in the group delay domain. This results in a drawback of group delay functions for representing non-minimum phase signals. As practical signals are rarely minimum phase, and zeroes in the vicinity of the unit circle are common, the group delay function as it is cannot be applied for estimation or segmentation analysis. Nagarajan [29] showed that it is possible to derive the minimum-phase equivalent from a non-minimum phase signal using the root cepstrum method. The causal portion of the inverse Fourier transform of the squared magnitude spectrum (in fact, to any power γ) was

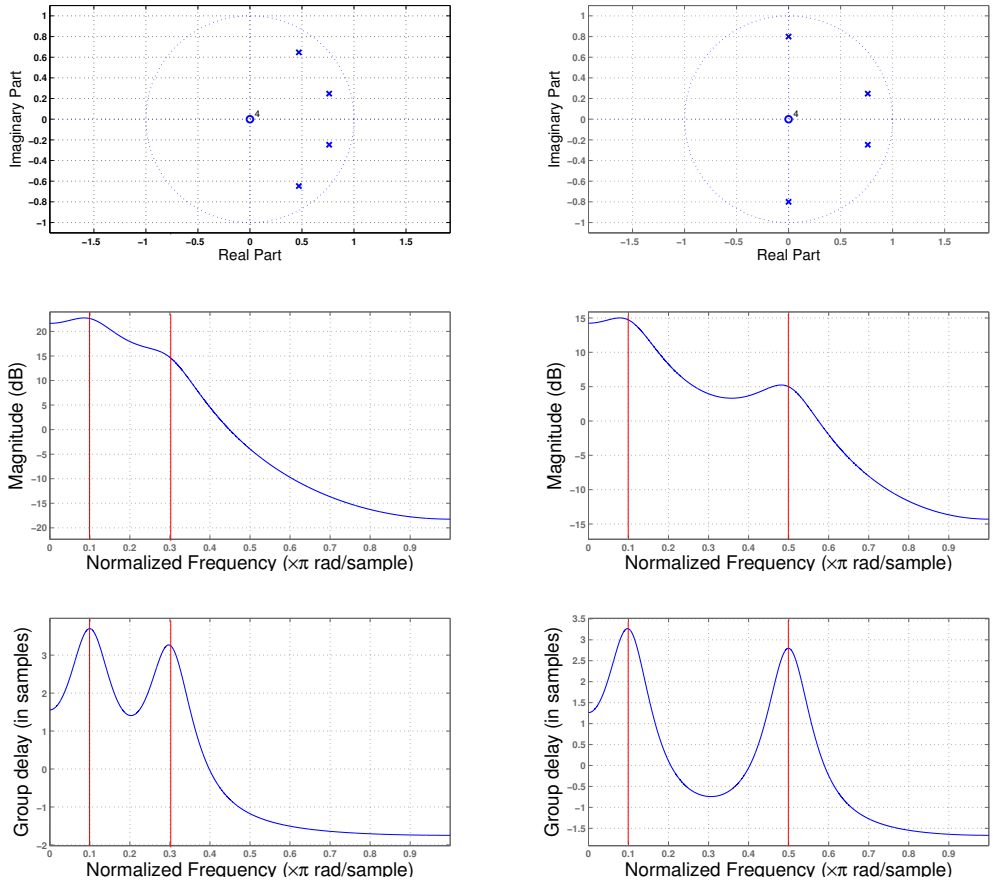

Fig. 2: Resolving power of the group delay function *(Top)* Pole-Zero plots for two minimum phase systems. *(Center)* Corresponding magnitude spectra with the resonant frequencies marked *(Bottom)* Group delay spectra.

proved to be minimum-phase. This property has since been exploited for segmentation of speech into syllables [30], [31], [32].

A brief overview of the algorithm is presented :

Onset Detection Algorithm

1: Differentiate the music signal $s(t)$ to obtain $s'(t)$.
2: Compute the analytic signal $s'_a(t) = s'(t) + s'_H(t)$.
3: Obtain the envelope $e(t) = |s'_a(t)|$.
4: Compute $x(t) = \mathcal{F}^{-1}(e(t) + e(-t))$. $x(t)$ is necessarily minimum phase.
5: Compute detection function $d(t)$ = Group delay of $x(t)$ using (6).

Figure 3 illustrates the various stages of the algorithm using a mridangam clip with both loud and silent strokes. Differentiating the music signal emphasizes the location of all onsets considerably as shown in Fig. 3(b) when compared to the original music signal as shown in Fig. 3(a). The envelope function is estimated on 3(b) using Hilbert transform and down-sampled. Treating 3(c) as the positive half of the magnitude spectrum of a hypothetical signal, the minimum-phase group delay equivalent is computed in 3(d). It is interesting to note that in the final step, the group delay function aprroximates all strokes to an equal amplitude irrespective of their original amplitude values. Onsets are reported on the group delay function as instants of significant rise using a threshold[1].

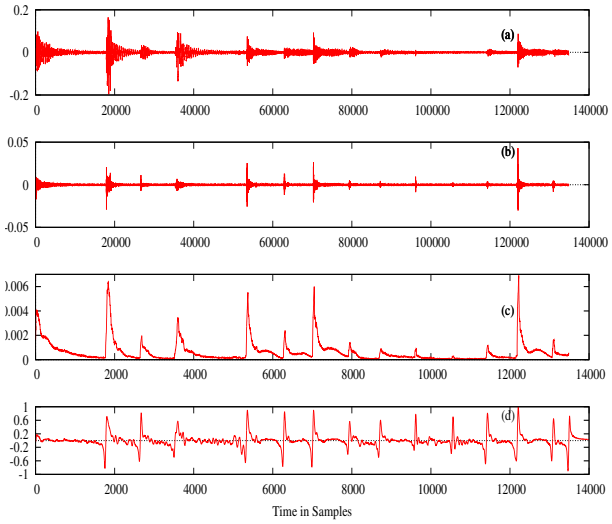

Fig. 3: Working of proposed algorithm - (a) Music signal (b) Derivative of music signal c) Envelope estimated using Hilbert transform (d) Minimum phase group delay computed on the envelope.

[1]More examples can be found at www.iitm.ac.in/donlab/music/mridangam

IV. EXPERIMENTATION

Since there exist no annotated datasets for the instruments considered, we have created one in the course of this work. The experiments have been performed on this dataset of annotated onsets, consisting of $17,592$ strokes from mridangam, ghatam, kanjira, morsing, thavil and an ensemble of all instruments except thavil. The recordings were taken from *tani avarthanams* (solo) of Carnatic concerts. The mridangam recordings are split into musically relevant 'phrases' by professional musicians while all other instruments are split into segments of $20s$ each. The entire dataset is sampled at $44.1KHz$ and the combined duration of the 277 clips is ≈ 42 minutes. Instrument-wise details of the dataset are provided in Table I.

Instrument	Total length (min:sec)	Strokes
Mridangam	18:41	5982
Ghatam	4:14	2616
Kanjira	3:11	1377
Morsing	6:35	2184
Thavil	4:39	2904
Ensemble	5:00	2529

TABLE I: Instrument wise details

An onset is treated as correct (*True Positive*) if it is reported within a threshold ($\pm 50ms$) of the ground truth. The tolerance is introduced to account for errors in manual annotation. A *False Positive* does not fall within the threshold of any of the time instants in the ground truth whereas a *False Negative* is a missed onset. The following metrics are used to evaluate the algorithm :

$$Precision(P) = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{8}$$

$$Recall(R) = \frac{N_{FN}}{N_{TP} + N_{FP}} \tag{9}$$

$$F\text{-}measure = \frac{PR}{P + R} \tag{10}$$

where N_{TP}, N_{FP} and N_{FN} represent the number of true positives, false positives and false negatives respectively. Previous evaluations of onset detection algorithms [33][15] have merged closely spaced onsets (most commonly, within $30ms$) and treated them as one, based on pyscho-acoustical studies of human perception of onsets[34]. In such cases, the arithmetic mean of consecutive onsets was taken and replaced the multiple onsets. We do not perform the above step since it becomes impossible to differentiate between simple and composite [2] strokes, the latter being quite common in mridangam, kanjira and thavil. Further, we have not considered the case of multiple onset outputs within the threshold of a target and a single onset output within the threshold of multiple targets. They are treated as false positives and false negatives respectively.

Comparison is made with a state-of-the-art algorithm [13] based on convolutional neural networks (CNNs). The network is trained with 80-band Mel filter banks scaled logarithmically in magnitude from spectrograms of multiple resolutions. 102 minutes of monophonic and polyphonic instrumental recordings are used for training the network. The output activation function is smoothened and a local threshold is used to detect onsets.

By varying the thresholds in the proposed algorithm as well as in [13], we report the optimum results.

A. Results

Both onset detection algorithms report fairly good F-measures on all instruments, as expected of percussion instruments. It is interesting to note from Table II that the proposed algorithm stands out in performance for the mridangam dataset, in spite of considerable variations in tempo and loudness. The lower recall for the proposed algorithm on all other instruments can be attributed to the faster tempo in comparison to the mridangam, suggesting even better temporal resolution might be necessary. Both algorithms report significantly lower recall on kanjira and thavil, which have a relatively high occurrence of composite strokes. Finally, morsing strokes lack a sharp onset, and this is directly reflected in the decrease in F-measures.

V. CONCLUSION

In this paper, we begin our investigation of five major Carnatic percussion accompaniments with the task of onset detection. We propose a purely signal processing based approach using the segmentation property of group delay functions, that is comparable in performance with the state-of-the-art technique. We have also created a fairly large annotated dataset for onset detection consisting of only Carnatic instruments in the process of this work. In the future, we plan to improve the proposed algorithm by replacing the envelope function with existing *detection functions*. We also plan to take up the task of stroke and pattern classification, which are well defined for all the instruments considered except morsing. We intend to use the onset detector as a pre-processing tool to segment concert recordings, in order to train isolated stroke models.

ACKNOWLEDGMENT

This research was partly funded by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). The authors are grateful to renowned mridangam artist Padma Vibhushan Sri Umayalpuram K Sivaraman and his disciples, for assisting in preparation of the mridangam dataset. They would also like to thank Jan Schlüter for sharing the CNN implementation.

[2] Composite refers to both left-right strokes occurring together in mridangam, and strokes which involve multiple fingers in case of kanjira and morsing

REFERENCES

[1] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 5, pp. 1035–1047, 2005.

[2] M. Goto and Y. Muraoka, "Beat tracking based on multiple-agent architecture a real-time beat tracking system for audio signals," in *Proc. Second International Conference on Multiagent Systems*, 1996, pp. 103–110.

	Group delay			CNN		
Instrument	**Precision**	**Recall**	**F measure**	**Precision**	**Recall**	**F measure**
Mridangam	0.972	0.974	0.973	0.964	0.941	0.952
Ghatam	0.968	0.924	0.946	0.968	0.943	0.956
Kanjira	0.936	0.914	0.925	0.98	0.972	0.976
Morsing	0.925	0.907	0.916	0.936	0.937	0.937
Thavil	0.95	0.805	0.872	0.991	0.82	0.897
Ensemble	0.927	0.886	0.906	0.956	0.921	0.938

TABLE II: Performance measures for the proposed algorithm vs convolutional neural networks based state-of-the-art algorithm, on various Carnatic percussion instruments

[3] W. A. Schloss, "On the automatic transcription of percussive music: From acoustic signal to high level analysis," Ph.D. dissertation, Stanford University, CA, USA, May 1985. [Online]. Available: http://ccrma.stanford.edu/STANM/stanms/stanm27/stanm27.pdf

[4] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *In Proc. Digital Audio Effects Workshop (DAFx*, September 2004, pp. 1–7.

[5] P. Masri, "Computer modeling of sound for transformation and synthesis of musical signals," Ph.D. dissertation, University of Bristol, UK, 1996.

[6] J. P. Bello and M. Sandler, "Phase-based note onset detection for music signals," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP). 2003 IEEE International Conference on*, vol. 5. IEEE, 2003, pp. V–441.

[7] J. P. Bello, C. Duxbury, M. Davies, and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 553–556, 2004.

[8] C. Duxbury, J. P. Bello, M. Sandler, M. S, and M. Davies, "A comparison between fixed and multiresolution analysis for onset detection in musical signals," in *In Proc. Digital Audio Effects Workshop (DAFx*, October 2004, pp. 3–7.

[9] E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini, and B. Schuller, "Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*, 2014, pp. 2164–2168.

[10] L. Gabrielli, F. Piazza, and S. Squartini, "Adaptive linear prediction filtering in dwt domain for real-time musical onset detection." *EURASIP Journal on Advances in Signal Processing*, 2011.

[11] W.-C. Lee and C.-C. Kuo, "Musical onset detection based on adaptive linear prediction," in *Multimedia and Expo, 2006 IEEE International Conference on*, July 2006, pp. 957–960.

[12] E. Marchi, G. Ferroni, F. Eyben, S. Squartini, and B. Schuller, "Audio onset detection: A wavelet packet based approach with recurrent neural networks," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, July 2014, pp. 3585–3591.

[13] J. Schlüter and S. Böck, "Improved Musical Onset Detection with Convolutional Neural Networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, May 2014.

[14] S. Dixon, "Onset detection revisited," in *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx06)*, 2006, pp. 133–137.

[15] S. Böck, F. Krebs, and M. Schedl, "Evaluating the online capabilities of onset detection methods." in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2012)*, 2012, pp. 49–54.

[16] S. A. Shanmugam and H. A. Murthy, "A hybrid approach to segmentation of speech using group delay processing and hmm based embedded reestimation," in *INTERSPEECH*, 2014.

[17] ——, "Group delay based phone segmentation for HTS," in *National Conference on Communications 2014 (NCC-2014)*, Kanpur, India, Feb. 2014.

[18] A. Holzapfel, Y. Stylianou, A. Gedik, and B. Bozkurt, "Three dimensions of pitched instrument onset detection," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1517–1527, Aug 2010.

[19] E. Benetos and Y. Stylianou, "Auditory spectrum-based pitched instrument onset detection," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 1968–1977, Nov 2010.

[20] S. Böck and G. Widmer, "Local group delay based vibrato and tremolo suppression for onset detection," in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, Curitiba, Brazil, November 2013.

[21] C. Raman, "The indian musical drums," in *Proceedings of the Indian Academy of Sciences-Section A*, vol. 1, no. 3. Springer, 1934, pp. 179–188.

[22] R. Siddharthan, P. Chatterjee, and V. Tripathi, "A study of harmonic overtones produced in indian drums," in *Physics Education*, 1994, pp. 304–310.

[23] A. Anantapadmanabhan, A. Bellur, and H. A. Murthy, "Modal analysis and transcription of strokes of the mridangam using non-negative matrix factorization," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 181–185.

[24] D. Gabor, "Theory of communication," *The Journal of the Institution of Electrical Engineers*, vol. 93, no. 26, pp. 429–457, 1946.

[25] B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," *Acoustical Society of America*, vol. 63, pp. 1638–1640, 1979.

[26] R. M. Hegde, H. A. Murthy, and V. R. R. Gadde, "Significance of the modified group delay features in speech recognition," *IEEE Transactions on Audio,Speech and Language Processing*, vol. 15, pp. 190–202, January 2007.

[27] S. Parthasarathi, R. Padmanabhan, and H. A. Murthy, "Robustness of group delay representations for noisy speech signals," *International Journal of Speech Technology*, vol. 14, no. 4, pp. 361–368, 2011.

[28] R. Padmanabhan, S. H. K. Parthasarthi, and H. A. Murthy, "Robustness of phase based features for speaker recognition," in *Proceedings of Int. Conf. Spoken Language Processing*, September 2009, pp. 2355–2358.

[29] T. Nagarajan, V. K. Prasad, and H. A. Murthy, "Minimum phase signal derived from the root cepstrum," *IEE Electronics Letters*, vol. 39, pp. pp.941–942, June 2003.

[30] V. K. Prasad, T. Nagarajan, and H. A. Murthy, "Automatic segmentation of continuous speech using minimum phase group delay functions," in *Speech Communications*, vol. 42, Apr. 2004, pp. 1883 – 1886.

[31] T. Nagarajan, V. K. Prasad, and H. A. Murthy, "The minimum phase signal derived from the magnitude spectrum and its applications to speech segmentation," in *SPCOM*, July 2001, pp. 95–101.

[32] T. Nagarajan, H. A. Murthy, and R. M. Hegde, "Segmentation of speech into syllable-like units," in *Proceedings of EUROSPEECH*, Geneva, Switzerland, September 2003, pp. 2893–2896.

[33] F. Eyben, S. Böck, B. Schuller, and A. Graves, "Universal onset detection with bidirectional long short-term memory neural networks." in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 589–594.

[34] S. Handel, *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, 1989.