Interspeech 2018
2-6 September 2018, Hyderabad

Acoustic-Prosodic Features of Tabla Bol Recitation and Correspondence with the Tabla Imitation

Rohit M A, Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology, Bombay
rohit.ma@iitb.ac.in, prao@ee.iitb.ac.in

Abstract

In the Indian classical drumming tradition, the different strokes on the tabla are named by spoken syllables(bols) in a case of onomatopoeia. The recitation of a tabla composition using vocalic syllables(bols) plays an important role in the oral tradition of pedagogy in North Indian classical music. Previous studies have considered the phonetic features of isolated bol utterances with the corresponding isolated strokes produced on the tabla. In this work, we investigate the acoustic properties of bol recitation beyond the segmental measurements related to the phones or syllables. The recitation of a tabla composition, apart from conveying the basic score in terms of the sequence of stroke name and onset times, is typically quite expressive in nature, being marked by pitch variations, loudness dynamics and voice quality variations across a sequence or phrase. Given the distinct spaces of acoustic variation of the voice and tabla, we study acoustic-prosodic variations in the recitation and investigate the corresponding (time-aligned) supra-segmental acoustic variations in the drumming. An available large dataset of recordings of selected tabla compositions by an expert tabla player, each aligned with the corresponding bol recitation, is employed in the analyses. We find that while the recitation reliably encodes intensity variations across bols in a cycle, the observed pitch variations are meaningful only for the pitch-varying drum strokes of the left drum.

Index Terms: Speech prosody, Tabla Bols, Onomatopoeia, North Indian Classical Music

1. Introduction

In the Indian classical drumming tradition, the different strokes on the tabla are named by spoken syllables (known as bols) in a case of onomatopoeia [1]. The recitation of a tabla composition using these vocalic syllables plays an important role in the oral tradition of pedagogy in North Indian classical music [2]. The syllables are well established and accepted by practitioners of the artform as means of recall and practice of the sophisticated rhythms. Further, the vocal recitation of tabla-compositions (structurally grouped sequences of bols) is very expressive, replete with pitch and intensity variations, much like natural speech or the recitation of poetry.

Using vocables to represent musical instrument sounds is a practice seen in several cultures across the world. In some cases it has arisen from the use of oral transmission of music and in others it has made communication easier during collaborations between musicians. The system of representing instrument sounds using speech syllables is well standardized in some cultures like the system of bols in the tabla, the *canntaireachd* in Scottish bagpipe music [3], and the vocal imitations of Church bells in traditional Lithuanian texts [4]. In all these cases, the syllables serve as an aid to memory and in some cases also capture the properties of the sounds they represent. However, in the case of Western music, like the drums or the guitar, this representation has remained informal, motivating perceptual studies to propose and validate any system of vocables [5, 6].

In the case of tabla music, research has long focused on the problem of automatic recognition of tabla bols from played sequences [7]. The approach taken in these works has essentially involved the following steps - onset detection and segmentation, feature extraction, and finally, classification. In [8], the authors went a step further and attempted to also discover select patterns of bol sequences in the transcription output.

Of more relevance to us is [1], one of the first studies combining the bols and the corresponding tabla strokes, in which several acoustic features of pairs of bols/strokes were analyzed and compared. Measurements made on the spoken utterances, of attributes such as brightness, duration, frequency and power of the fundamental and formant frequencies of the vowels, spectro-temporal variations in sounds containing a combination of aspiration, voiced and unvoiced phonemes, etc., were compared with similar equivalent measurements made on isolated tabla strokes.

In [9], the improvisational aspects of tabla playing were addressed in an attempt to propose a linguistic model for tabla improvisation, using some rules of the underlying grammar learned from known correct sequences. This is similar in spirit to the work in the present paper, in that both offer a linguistic perspective of the different forms of tabla composition presentation, drawing on principles of language and speech. On the specific topic of bol recitation however, the closest studies are mostly musicological, such as [10], which is a thorough study of the system of vocables related to the *mridangam* - the South Indian counterpart of the tabla.

We note that there has been little acoustical analysis of bol recitation at the supra-segmental level. Given its communicative function, of interest also are the relationships that may exist between the prosodic variations in the recitation and the corresponding imitation on the instrument which itself is an acoustic system very different from the human voice. The present work aims to study the nature and extent of prosodic variations during bol recitation and look for correspondences, if any, in the stroke sequence acoustics in the reproduction by the tabla. This could help develop a model for any supra-segmentals that may exist in tabla music, along similar lines as the models for speech prosody. Such models could help improve the quality of synthesized tabla and increase its naturalness.

2. Dataset Description

We carry out our analyses on a parallel corpus of bol recitation and tabla rendered audio recordings. In this section, we

10.21437/Interspeech.2018-1692

Dha Dha - Dha; Dha Dha - Dhin; Ghi,Da Na,Ga Ti,Ra Ki,Ta; Taa Taa Ti,Ra Ki,Ta
Taa Taa - Taa; Dha Dha - Dhin; Ghi,Da Na,Ga Ti,Ra Ki,Ta; Taa Taa Ti,Ra Ki,Ta
Dha Dha - Dha; Dha Dha - Dhin; Ghi,Da Na,Ga Ti,Ra Ki,Ta; Taa Taa Ti,Ra Ki,Ta
Taa Taa - Taa; Dha Dha - Dhin; Ghi,Da Na,Ga Ti,Ra Ki,Ta; Taa Taa Ti,Ra Ki,Ta

Figure 1: *Score of a composition spanning one cycle of 16 beats (beats are separated by semi-colons).*

describe our dataset after providing the needed background on tabla music.

2.1. Background

The tabla is a principal percussion instrument in traditional music forms of India. It comprises two drums - a treble (*Dayan* or Tabla) and a bass drum (*Bayan*). Both drums are tuned prior to performance with the treble drum at a chosen tonic and the bass drum at a lower pitch. The two drums together give rise to a variety of sounds, each corresponding to a specified place of striking on the drum head and the precise hand gesture employed [11]. A classification of bols in terms of the characteristics of the sound produced appears in Table 1. The single resonant sounds produced on the *dayan* have a clear sense of pitch and produce slowly decaying bell-like sounds, while those on the *bayan* can realise a chosen pitch in a narrow range in either a flat or a rising/falling pitch contour even though the exact pitch may not be easy to perceive [12]. Sounds produced using a combination of these resonant strokes by striking both the drums simultaneously(such as, *Dha = Ta + Ghe*, *Dhin = Tin/Tun + Ghe*, *Thin = Tin/Tun + Ke*), can be timbrally quite rich since they combine the characteristics of both the drums. Damped or closed sounds on either drum are produced by damping the membrane immediately after the strike by a gesture of pressing on the membrane. These sounds are hence neither pitched nor sustained.

Tabla compositions, like melodic compositions, are always set to a tala (a cyclic pattern of beats), and while they can start on any beat of the cycle, they mostly end on the *sam* (the first beat of the cycle). The score of a composition is just the sequence of bols in written notation, occasionally with the only additional information supplied relating to timing, e.g. certain bols grouped together to indicate the bols occurring on every beat, and groups separated by a punctuation mark(e.g., a long space) [9]. There is, generally, no information about any dynamics to be employed in the rendering, unlike in the scores of Western classical music. Despite the skeletal description embedded in the score, the bol recitation of the composition is highly expressive and packed with prosodic variations, and this information is passed on orally from teacher to pupil.

2.2. Compositions and Audio Recordings

From Gupta et al. [8], we have the audio recordings of recitation and tabla realization, written scores and time aligned syllabic transcriptions for a set of 38 tabla compositions obtained from the educative video DVD by Pt Arvind Mulgaonkar, titled Shades of Tabla [1]. These are traditional compositions of various *gharanas* (styles), all set to *Tintal* - a 16-beat cycle, which have been recited and played by the artist, along with presentations about the nuances of every *gharana*, making the DVD a valuable pedagogical resource.

Figure 1 shows the score of a composition as obtained from

[1]underscorerecords.com/catalog/detail/604/Shades-of-Tabla

Table 1: *Classification of tabla strokes*

Drum struck	Nature of sound	Bol names	Whether Pitched (pitch range in Hz)	Whether Pitch can be varied
Bayan (bass)	Resonant	Ghe/ Ghi/Ga	Yes (80-100)	Yes (≈ an octave)
	Damped	Ke/Ki, Kat	No	No
Dayan (treble)	Resonant	Na/Ta, Tin,Tun	Yes (220-440)	No
	Damped	Ti,Ra,Te Tak, Dhe, Re	No	No
Both	Resonant	Dha, Dhin	Yes	Yes
	Resonant	Thin	Yes	No

the dataset. The composition spans one cycle of the *tintal*, with each line representing a *vibhaag* (a sub-division of the cycle comprising 4 beats). Bols which occupy one full beat are written between two semicolons, which mark the beat-boundaries. A hyphen indicates a pause of duration equal to one-fourth of a beat and a comma between two bols indicates bols which together occupy one-fourth of a beat, and hence occur at twice the speed. Audio recordings of the recitation and the playing of the same can be found in the attached media or at this link[2].

For the present work, we needed the audio recordings of the bol recitation as well; of the original 38 compositions, only 27 had both recitation and playing parts. These were extracted from the video and time-aligned with the already extracted tabla audios. The recitation and playing, although clearly audible in the recordings, are accompanied by a harmonium playing the *lehera* (cycle-marking melodic motif), making signal analysis such as tracking the intensity and pitch of the voice or tabla challenging. We therefore arranged for the dataset to be re-recorded by getting a senior disciple of the original artist to recite and play the chosen compositions. The recordings were performed in a sound-proof studio at a sampling rate of 48 kHz, with separate microphones capturing the recitation and the playing. The *dayan* was tuned to a frequency close to 257 Hz, and the *bayan* close to 80 Hz. The written scores and the original audio recordings of all the compositions were provided to the artist prior to the recording session. Of the 27, a total of 20 compositions which the artist was comfortably familiar with were finally recorded. The recording was carried out in sessions, with short rests between compositions and slightly longer breaks after every few compositions. Every composition was first recited and then, following a short pause, played on the tabla. During the recording, the artist listened over headphones to a background track of harmonium *lehera* and a metronome to match his tempo to that of the original version.

2.3. Statistical Description

The dataset of 20 compositions contains a total duration of 6 minutes each of the recitation and the playing audios with approximately 2400 syllables in each. Of these, 194 *Ghe*-based

[2]https://goo.gl/APh92A

bols, and 157 pitched *dayan* bols which were at least 100 ms long were chosen for analysis.

The compositions are between 8 and 30 seconds long, and are not all at the same tempo. The number of 16-beat cycles spanned by any composition is between 1 and 8. The number of bols per cycle ranges from 10 to 60, with the exception of 4 single-cycle compositions which have 85 bols in the cycle. In all, there are 70 cycles for our analyses.

3. Analysis Methods

A *tala* establishes strong beats which act as boundary markers. In the *tintal*, the *sam* (the first beat) and *khali* (the half-way mark) are the strongest beats in the cycle. These structurally accented positions mark the positions where phrases are most likely to begin and end [12]. In this work, we consider the segmentation of the audio of every composition into cycle-length durations, bounded by the *sam* as confirmed by the corresponding harmonium *lehera* track available to us. We investigate the supra-segmental features of each such cycle-length segment in the recitation and in the corresponding tabla playing. We also study bol-specific supra-segmental features. The signal processing steps to the extraction of pitch, intensity and duration features are presented here.

3.1. Onset Detection and Time-alignment

Short-term magnitude spectra were computed across the audio over windows of size 20 ms and a 10 ms hop-size. The magnitude spectra, differenced across adjacent frames, provides a measure of 'spectral flux' commonly used in musical note onset detection [13]. We applied the same to the bol and tabla recordings. Peaks in the spectral flux occur at the syllable (and stroke) onsets. To avoid false alarms, adjacent peaks closer than 40 ms were combined into a single onset. In portions of fast recitation, some onsets were missed, and a manual correction stage was employed to add the missed onsets. The onset detection function of each cycle was time-aligned with the corresponding bol sequence obtained from the written composition to achieve the automatic segmentation and labeling of both bol and tabla recordings. The acoustic features were next computed on each segment.

3.2. Feature Extraction

Motivated by the prosodic elements observed in the recitation, pitch and intensity features were computed from the audios. The pitch contours were obtained using a script in Praat [14] using windows of size 60 ms with a hop size of 10 ms. For the tabla audios, in the case of *Ghe*-based strokes where both drums are struck, pitch values were extracted from the low-passed signal in the 0-250 Hz range. This region in the spectrum is dominated by the *bayan* and the analysis of the low-passed signal provides estimates of the intensity and pitch of the *bayan* component of the stroke. The intensity contours were obtained as the set of short-time energy(squared amplitude) values computed over windows of size 20 ms, with a hop size of 10 ms.

The contours were then normalized by dividing the F0 and intensity values of every composition audio by the respective mean F0 and intensity values , and then representing the ratios on a logarithmic scale to obtain the intensity in dB and F0 in semitones.

Figure 2 shows the continuous pitch and intensity contours of the recitation and the playing of the first line of the composition in Fig 1. The solid line(in blue) denotes the pitch and the dashed line(in green) the intensity. Part (a) in the figure is the transcription label track and part (b) shows the contours extracted for the recitation. Parts (c) and (d) correspond to the low and high-pass components of the playing audio, and essentially capture the *bayan* and the *dayan*, respectively. (The low and high pass filters were basically Hann-shaped windows in the frequency domain with a transition bandwidth of 50 Hz and a cut-off frequency of 250 Hz). As is evident, the pitch track in (c) exhibits significant variations wherever a *bayan*-involving stroke occurs, and in (d) it mostly stays constant at the detected F0 of the *dayan*. The recitation track, (b), has variations in F0 throughout the audio, except of course where either silence or an unvoiced sound is present. Also, note that although the '-' in (a) corresponds to a pause, a rising F0 is seen in (b) corresponding to the first pause because the player elongated the previous bol instead of cutting it off. Further, the intensity contour varies continuously in all three cases.

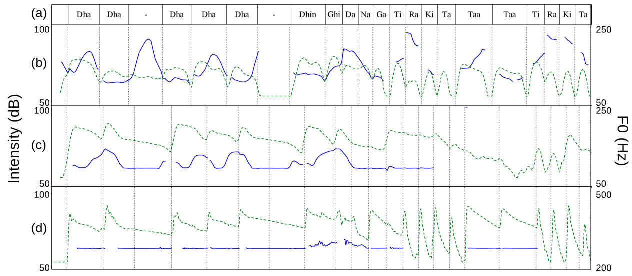

Figure 2: *Pitch and intensity contours (in blue and green resp.) of (b) the recitation (c) the low-pass band and (d) the high-pass band of the tabla audio. (a) is the transcription label track. Note that although the '-' in (a) corresponds to a pause, a rising F0 is seen in (b) corresponding to the first pause because the player elongated the previous bol instead of cutting it off.*

Using the bol/stroke segment boundaries obtained earlier, corresponding portions from the normalized pitch and intensity contours were extracted and a feature vector was constructed for every bol/stroke using the following pitch and intensity related features typical in speech prosody analyses - F0 span, mean F0, and mean intensity [15]. We thus obtained a sequence of feature values (for each feature category) corresponding to the sequence of bols in the composition.

3.3. Correlation Computations

The correlation between the recitation and the playing was sought at two levels - one at the cycle-level for every composition and the other at the individual bol/stroke level for all the pitched bols/strokes taken from across the dataset. Chiefly, the points of interest were the intensity variations observed over a cycle in every composition, and the pitch variations in the recitation of pitched strokes, whose pitch may or may not be variable on the tabla.

In the cycle-level analysis, every composition was first segmented into its cycles, resulting in a total of 70 cycles from all the compositions in the dataset. The Pearson Correlation Coefficient(r) was then computed for every cycle, between the sequence of mean intensity values of the bols in the recitation and the corresponding strokes in the playing of the cycle.

In the bol/stroke-level analysis, all the *Ghe*-based, and all the fixed-pitch *dayan*(Ta,Tin,Tun) strokes were extracted from the entire dataset. Feature sequences(for each feature category) of all the recited and played instances were constructed for each of the two categories of bols, and compared. Additionally, for the *Ghe*-based bols, the mean intensity of the low-pass band in the playing was compared with the mean intensity of recitation.

4. Results and Discussion

4.1. Bol characteristics

A comparison between the syllable rate, intensity range and F0 range values of the bol recitation in the dataset, and conversational English speech is provided in Table 2. The syllable rate of recitation depends on the composition and goes from a minimum of 4-5 per second to nearly 12-14 per second. The F0 and intensity ranges are found to be comparable to that of very lively speech [16, 17]. Interestingly, the F0 range is found to be high and comparable for the two categories of orally recited bols even though the *dayan* tabla strokes have a fixed pitch.

Table 2: *Comparing syllable rate, intensity span and F0 span of bol recitation with conversational English speech.*

Attribute	Bol Recitation		Conversational Speech
Syllable rate (bols/s)	Min: 4-5 Max: 12-14		4-5
Norm. Intensity range (dB)	≈20		≈20
Avg. F0 variation (SD in semitones)	Ghe-based bols	4.23	4.0
	Pitched *dayan* bols	5.1	

4.2. Bol-Stroke Correspondences

4.2.1. Cycle-level

Figure 3 shows a histogram of the *r* values obtained in the cycle-level correlation computation of mean intensity values. We note that the most frequently occurring r value is around 0.5, indicating a moderately strong positive correlation in most cycles. Among the cycles with weak correlation, some were found to contain phrases whose recitation had a large variation in the intensity level over the course of the phrase, which was not consistently observed in the playing. This is perhaps because of a minimum intensity level that every stroke on the tabla must possess in order to sound correct. Similarly, in some fast sequences, the recited bols were not found to be equally intense, while the corresponding tabla strokes had similar intensities.

4.2.2. Bol/Stroke-level

The correlation coefficients computed for each of the considered prosodic attributes between the recited and played bols are shown in Table 3. The mean F0 values are found to correlate well only in the case of the *Ghe*-based strokes, which is to be expected, since the *dayan* strokes have a fixed F0. The F0 span, however, was not found to correlate as well in the case of the *Ghe*-based strokes. The mean intensity values of recitation correlate well with the playing in both the categories of bols. However, they do not correlate as well with the low-pass band of the played *Ghe*-based strokes indicating that it is the total intensity of these strokes that corresponds best with that of the oral form.

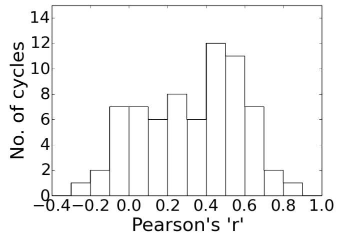

Figure 3: *Histogram of Pearson's 'r' values obtained in the cycle-level correlation of mean intensity.*

Table 3: *Correlation between the mean F0 and intensity values of the played and recited bols (Ghe-based and pitched dayan).*

Bol/Stroke	Attribute of the recitation	Attribute of the playing	Pearson 'r'
Ghe-based	$F0_{mean}$	$F0_{mean}$	0.37 ($p<0.05$)
	I_{mean}	I_{mean} of full band	0.39 ($p<0.05$)
	I_{mean}	I_{mean} of low-pass band	0.28 ($p<0.05$)
Dayan pitched	$F0_{mean}$	$F0_{mean}$	0.085 ($p>0.05$)
	I_{mean}	I_{mean}	0.35 ($p<0.05$)

5. Conclusion

Based on a dataset of recordings, by an expert tabla artist, of bol recitations and accompanying tabla renderings, we observed the significant presence of expressive prosody in the oral recitation of compositions. Given that the human voice and the tabla are very different acoustic systems, we studied the possible correspondence of F0 and intensity dynamics between the recitation and the tabla rendering. The tabla strokes are produced with a variety of articulatory mechanisms and include unpitched and fixed-pitch sounds, unlike the human voice where the vowels realise the prosodic variations irrespective of the bol. Our acoustic analyses demonstrate the positive correlation at the rhythmic cycle level between intensity variation across the vocalic syllables and that across the tabla strokes. The F0 variation correspondences are equally strong but restricted to the pitch-varying *bayan* strokes, as expected, although both varieties of vocables show F0 variation to the same extents. Future work will involve enhancing the audio processing methods to achieve better separation between the overlapping strokes both in time and frequency for more accurate analyses. Other musicological functions of the prosodic variation in the oral recitation including the marking of rhythmic accent in the cycle and demarcating rhythmic phrase boundaries will be studied once the dataset is manually annotated further for accents and phrases.

6. Acknowledgments

We would like to thank Shri Prasad Padhye for his help in reproducing the dataset, and Prof B. K. Chakravarthy of IDC, IIT Bombay, for making the studio available for the recording. Special thanks to Kaustuv Kanti Ganguli for discussions over the course of this work.

7. References

[1] A. D. Patel and J. R. Iversen, "Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: An empirical study of sound symbolism," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 2003, pp. 925–928.

[2] A. Pradhan, *Tabla: A Performer's Perspective*. BookBaby, 2011. [Online]. Available: https://books.google.co.in/books?id=42laDQAAQBAJ

[3] C. K. Chambers, "Non-lexical vocables in scottish traditional music," 1980.

[4] R. Ambrazevičius and R. Balsienė, "Imitations of bells: Correspondence between bell acoustics and onomatopoeic texts." *Journal of Interdisciplinary Music Studies*, vol. 4, no. 2, 2010.

[5] O. Gillet and G. Richard, "Drum loops retrieval from spoken queries," *Journal of Intelligent Information Systems*, vol. 24, no. 2-3, pp. 159–177, 2005.

[6] C. Traube and N. DAlessandro, "Vocal synthesis and graphical representation of the phonetic gestures underlying guitar timbre description," in *8th International Conference on Digital Audio Effects (DAFx05)*. Citeseer, 2005, pp. 104–109.

[7] P. Chordia, "Segmentation and recognition of tabla strokes." in *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, 2005, pp. 107–114.

[8] S. Gupta, A. Srinivasamurthy, M. Kumar, H. A. Murthy, and X. Serra, "Discovery of syllabic percussion patterns in tabla solo recordings," in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, 2015.

[9] B. Bel and J. Kippen, "Modelling music with grammars: formal language representation in the bol processor," 1992.

[10] D. F. Dineen, *Speaking Time, Being Time: Solkattu in South Indian Performing Arts*. Wesleyan University, 2015.

[11] K. Narang and P. Rao, "Acoustic features for determining goodness of tabla strokes," in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, 2017.

[12] P. Chordia, "Listeners Guide To Solo Tabla," https://ccrma.stanford.edu/~pchordia/soloTablaNotes.htm, [Online; accessed 19-March-2018].

[13] S. Dixon, "Onset detection revisited," in *Proceedings of the 9th International Conference on Digital Audio Effects*, vol. 120. Citeseer, 2006, pp. 133–137.

[14] "Paul Boersma & David Weenink (2012): Praat: doing phonetics by computer [Computer program].Version 5.3.16."

[15] K. Sabu and P. Rao, "Detection of prominent words in oral reading by children," in *Proceedings of the 9th Speech Prosody Conference, Poznan, Poland*, 2018.

[16] M. Gelfer and S. Young, "Comparisons of intensity measures and their stability in male and female speakers." *Journal of voice: official journal of the Voice Foundation*, vol. 11, no. 2, p. 178, 1997.

[17] H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults," *Unpublished manuscript*, 1995.