To appear in Proc. of Frontiers of Research on Speech and Music (FRSM)
15-16 December 2017, NIT Rourkela, Orissa, India.

RELIABILITY OF ACOUSTIC PERCEPTION OF TABLA STROKES IN DETERMINING THEIR QUALITY

Snehesh Nag, Rohit M A, and Preeti Rao

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

***Abstract**: The Tabla is one of the principal percussion instruments of the North Indian traditional form of music. A myriad of sounds can be produced by this instrument, by introducing slight changes in the position and manner of striking. This makes the precise understanding of the idiosyncrasies of every stroke imperative. Previous work has explored the possibility of building a system that can automatically identify, using suitably devised acoustic features, common misarticulations and their nature. In this work, we present a study of the reliability of human perception in deciding tabla stroke quality, in which the extent of agreement among three Tabla experts in ranking strokes of individual tabla bols based on their quality was analyzed. We prepared, for this task, a dataset comprising recordings of individual tabla strokes obtained from Tabla learners of different levels of experience. We also performed a study on the effectiveness of a simple acoustic feature like sustain duration in determining stroke quality.*

1. Introduction

The tabla is a percussion instrument used in North Indian music, consisting of two sealed membranophones with animal-skin heads: the smaller, wooden-shell ‘dayan’ (played with the right hand and tuned to a higher pitch) and the larger, metal-shell ‘bayan’ (played with the left hand). The different regions of the tabla (applies to both membranophones) are the rim(Kinar), head(Maidan) and the patch(Syahi). Out of the set of basic ‘bols’ associated with the tabla, an analysis on 8 bols was carried out. These are - the resonant bols ‘Ta’, ‘Tin’, ‘Tun’, and damped bols ‘Ti’, ‘Re’, ‘Tak’ on the dayan and the resonant and damped bols ‘Ghe’ and ‘Ke’, respectively, on the bayan.

Spectral analysis of tabla strokes has been previously carried out in the context of bol recognition for tabla music transcription([7], [8]). In recent work ([1]), a thorough harmonic and timbral feature analysis was performed for several bols and it was noted that each of the strokes was characterized uniquely by subsets of certain harmonic and timbre based features. This highlights the precise nature of the acoustic characteristics of the various bols and the clear distinctions present between them.

The present work was motivated primarily by the work carried out in [4], where a stroke's harmonic content was analysed in view of providing feedback on the stroke's correctness. However, work in [4] was limited in the sense that all the analysis was carried out on a single tabla and incorrect strokes were simulated by an expert. With the ultimate goal of making the system universal, an investigation was carried out in the present work to assess the reliability of acoustic perception in a task of comparing the goodness of strokes collected from different tabla sets and players of varying skill levels. This was intended to motivate the use of the aforementioned acoustic characteristics in the automatic evaluation of tabla strokes. Such a perception testwas last reported in [5], where an informal survey was conducted to validate the quality of strokes produced by a mechanical arm.

Lastly, motivated by the use of the half-decay time of the amplitude envelope in [6] to distinguish the resonant bol ‘Ghe’ from the non-resonant bol ‘Kat’, a similar quantitative attribute - 'duration of sustain', was explored for its feasibility as a measure of goodness of Ghe strokes.

2. Dataset Collection

A dataset consisting of 128 isolated strokes of the 8 basic bols – Ghe, Ke, Ta, Tin, Tun, Ti, Re, and Tak, was created by recording strokes played by 5 players on 3 different tabla-sets (at a sampling rate of 44.1 KHz, on an Edirol R-09H Stereo microphone recorder). The players were of various levels of experience, and the tabla-sets were tuned to different pitches, intentionally, to bring in variability. The tabla-sets were also of different qualities. The recordings were carried out in normal household environments with steps taken to minimize disturbances. The players were asked to play several instances of every bol as naturally as possible, such that a new stroke was played only after the previous stroke was no longer audible. Table 1 provides a description of the dataset collected.

It is evident from the table that the dataset consists of a diverse set of players in terms of their experience, as well as a variety of tabla-sets in terms of their quality and pitch of tuning. In some cases the same tabla-set has been used by two people of different levels of expertise – players 2 and 3, and players 6 and 7. This makes possible the study of the effect of tabla quality on the judgement of the quality of strokes produced by players of different levels of expertise. Also, a 3-point scale (Good, Medium and Poor) was used to rate the quality of the tabla-sets. The quality is based on parameters like the duration of sustain of resonant strokes and the sharpness of non-resonant strokes, as obtained by an expert player.

<table>
<tr><th rowspan="2">Tabla Player ID</th><th rowspan="2">Tabla Set ID</th><th rowspan="2">Duration of experience</th><th colspan="2">Tabla Pitch (Hz)</th><th rowspan="2">Tabla Quality</th></tr>
<tr><th>Bayan</th><th>Dayan</th></tr>
<tr><td>1</td><td>Set 1</td><td>2 years</td><td>88.5</td><td>257.5</td><td>Good</td></tr>
<tr><td>2</td><td rowspan="2">Set 2</td><td>10 years</td><td rowspan="2">80.5</td><td rowspan="2">262.2</td><td rowspan="2">Good</td></tr>
<tr><td>3</td><td>13 years</td></tr>
<tr><td>6</td><td rowspan="2">Set 3</td><td>2 months</td><td rowspan="2">100.75</td><td rowspan="2">272.2</td><td rowspan="2">Poor</td></tr>
<tr><td>7</td><td>4 years</td></tr>
</table>

Table 1: Description of the prepared dataset

3. Perception Testing

The focus of the work was to test the reliability of using audio alone for the accurate assessment of tabla stroke quality, by analysing the extent of agreement between various Tabla experts in judging stroke quality. We sought to study this by conducting a stroke-quality ranking task, where experts had to rank a given set of strokes in the order of quality (in other words, sort them). While a more thorough alternative could be a task where experts would have to closely analyse each stroke and label them Good or Bad objectively, a ranking task serves as a faster way to obtain responses, since rating strokes in detail is a more trying task than arranging them in an order of perceived quality. Furthermore, the objective of studying the presence of any correlations between experts on their perception of stroke quality is accomplished equally well in a ranking task.

For the ranking task, the dataset collected was split into separate sets of bols (i.e., bol-sets), with each bol-set containing 4-6 strokes of a particular bol produced on a particular tabla-set. 8 such bol-sets were created for each of the 3 tabla-sets. 3 expert tabla players were then asked to number the strokes in each of these bol-sets in the order of quality (with 1 representing 'Best' and the last number representing 'Worst'). They were encouraged to listen to the audio files in every set in no particular order, and as many ever times as necessary. And although they were asked to avoid ties in ranking, they were not forbidden from including tied ranks if ever they found strokes to be too similar to rank apart. The experts were in no way related to each other and did not all follow the same style ('*gharana*') of playing either (2 of them practiced the Farrukhabad Gharana, while the other followed a mix of different styles).

To measure the level of agreement between the raters, a statistic known as Kendall's Tau-b[2], was made use of. 'Kendall Rank Correlation Coefficient', commonly referred to as Kendall's tau coefficient, a statistic introduced by Maurice G Kendall, is used to measure the ordinal association between two measured quantities, and hence, can serve as a measure of rank correlation between a pair of ranked variables. While the coefficient in its original form does not account for ties, an alternate version (also proposed by Kendall) known as Tau-b, which does take into account the presence of ties, was used, since a few cases with tied ranks were observed. (The exact implementation is the function available in python's scipy library [9]).

Kendall tau-b takes valuesin the range -1 to 1, with a value of 1 indicating complete agreement between a pair of raters, and a value of -1 indicating complete disagreement, i.e., exactly opposite rankings. A value of 0 indicates the lack of any correlation between the rankings.

The value of tau-b was computed for every pair of rankings of a bol-set, and then the mean of the three resulting tau-b values was taken as the average measure of correlation among the three raters for that bol-set. Table 2 shows the values of the correlation coefficient computed this way for every bol-set.

Tabla Set	**Bol**							
	Ghe	Ke	Ta	Tin	Tun	Ti	Re	Tak
Set 1	0.73	0.2	-0.19	-0.11	-	0.465	-	0.067
Set 2	0.86	-0.33	0.73	0.59	0.465	-0.47	0.59	-0.24
Set 3	0.067	0.86	0.73	-0.11	0.33	**0.2***	0.73	**-0.33***

Table 2: Kendall's tau-b values for experts' rankings in the ranking task
(* indicates cases where only two sets of rankings were available due to one of the raters finding all the strokes to be too similar to unambiguously rank. Also, the missing values indicate cases which were not considered for this task due to the lack of sufficient usable data in that particular bol-set.)

4. Audio Analyses

In this work, a simple feature like the 'duration of sustain' was tested for its effectiveness as an objective measure of stroke quality in case of the resonant bol 'Ghe'. This was motivated by the fact that a common misarticulation in case of 'Ghe', as opined by Tabla Gurus, is the application of excessive hand pressure which causes the stroke to sound damped

To compute the duration of sustain of a stroke, an envelope of the stroke's audio signal was first obtained by selecting the points of local maximum in consecutive non-overlapping 50ms - segments of the absolute representation of the audio signal. This was followed by computing two time-durations as measures of the decay of the signal – the durations from the time of maximum amplitude of the envelope to the times when the amplitude has fallen to ten per cent and one per cent of the maximum. Figure 1 shows a plot of the absolute amplitude of a 'Ghe' stroke (versus sample number) along with the extracted envelope and dashed lines marking the points of the two durations.

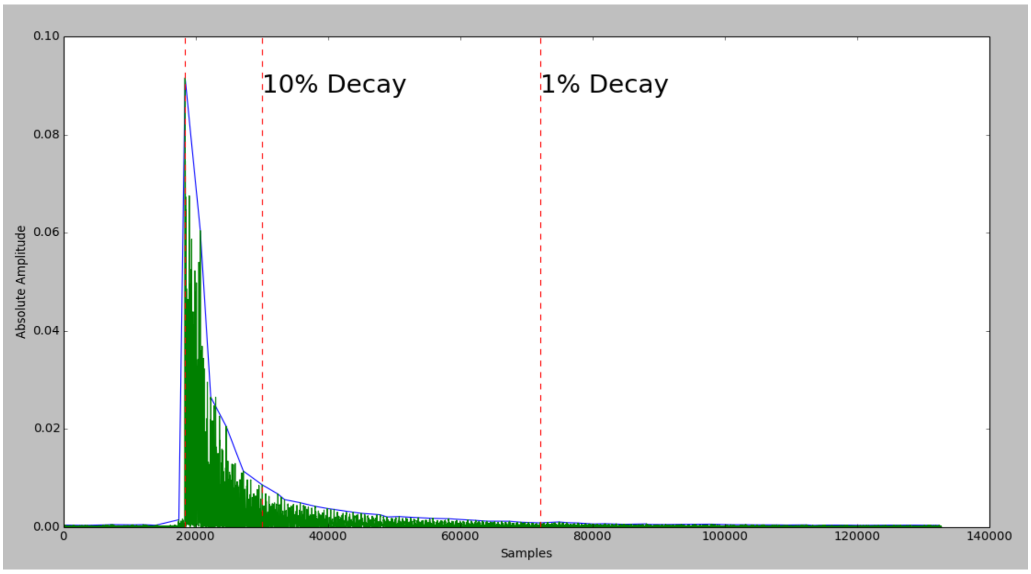

Figure 1: A plot (amplitude vs sample number) showing the calculation of the duration of sustain – The absolute amplitude of the audio signal is marked in green, the envelope in blue and the durations using red dashed lines

The values of the different sustain durations were determined in this fashion for all the strokes, and the strokes of every bol-set (bol-sets as used in the ranking task) were then sorted in the order of highest to lowest sustain duration, separately based on the 10% and the 1% durations (i.e., out of say, 5 strokes, the stroke with the longest sustain duration was ranked 1 and the stroke with the least duration ranked 5). These 'feature-based' rankings were then compared with each of the three experts' rankings for the two Ghe bol-sets with high correlation value in Table 2, and the correlation coefficient (tau-b) was computed for each case. The mean of the three resulting coefficient values was again taken as the average measure of correlation between the feature and the experts for that bol-set. Table 3 shows these values computed for the sets Set1 and Set2 of the bol Ghe. The values of average correlation among expert raters (reported in Table 2) have been reproduced for the sake of comparison.

Tabla Set	**Kendall's Tau (Correlation Coefficient)**		
	Among expert raters	Among experts and 10% sustain duration	Among experts and 1% sustain duration
Set1	0.73	0.8	0.73
Set2	0.86	0.73	-0.13

Table 3: Comparing rank correlation among expert raters and among the 'sustain duration' feature and expert raters for two sets of the bol 'Ghe'

5. Discussions

5.1 Expert Ranking Task

From Table 2 it can be observed that 'Tak' is the only bol with poor correlation values in all three sets. All the other Bols have high correlation coefficient values in at least one of the three sets. This suggests the possibility of an inherent difficulty in judging the bol 'Tak', and hence the requirement of a reference stroke to accurately judge it. This however needs a more thorough study with more raters involved in a similar ranking task. For the rest of the bols however, the poor correlation in some of the cases can be explained as follows:

Ghe – Set1 and Set2 have very high correlation coefficient values (0.73 and 0.86) whereas Set3 has a much lower value of 0.067. This could be due to the following reasons – while both Set1 and Set2 were recorded on good quality tabla-sets, the quality of Set3 was comparatively poorer. Furthermore, a significantly high amount of natural reverberation was noticed in the room where Set3 was recorded. The effect of this on the audio files may have caused difficulties in telling good from bad strokes. The fact that Ghe is a resonant bol only adds further ground to the claim.

Ke – Here, Set3 has a very high and Set1 a low positive correlation coefficient value, while Set2 shows negative correlation. A possible explanation to this is that both Set1 and Set3 have a good mix of correctly and incorrectly articulated strokes, while that is not quite the case for Set2. Being a damped bol, it is difficult, as expressed by raters, to accurately point out subtle differences between strokes. And a bad mix makes a good correlation less likely as subtle differences are more likely to be perceived differently.

Ta – The strong disagreement in Set1 is because of the presence of two variants of the stroke Ta [3] - one that is played closer to the kinar ('*kinarkabaj*') and another played closer to the maidan ('*surkabaj*'). Although these two versions are acoustically quite different, both of them fall under the bol Ta, and are not always set apart. Upon interviewing the raters about this, we found that while one of the raters made a more serious distinction between these two kinds of sounds, and considered one of them wrong, the others did not base their decisions so much on this distinction. However it is evident from their rankings that while the first rater preferred the *surkabaj*, the other two preferred the *kinarkabaj*.

Tin – The reason for a low coefficient value for Set3 is believed to just be the quality of the tabla itself. Raters have agreed that rating bols(especially resonant bols) played on a poor quality tabla was a difficult task.

5.2 Correlation between Duration of Sustain and Experts' Rankings

Before proceeding to a discussion on the correlation coefficient values of Table 3, we first discuss the acoustic significance of the 1% and 10% sustain durations in this work. It was observed that the amplitude level of the background noise in all the recordings was between 40% and 100% (averaging at about 64.5%) of the signal envelope amplitude at the 1% decay point. This made the 1% decay point a point in the signal when the envelope amplitude had faded to a level comparable to that of the background noise. Also, the 1% duration was, on an average, at least 1 second long from the point of the maximum signal amplitude. On the other hand, the 10% decay point was, on an average, a mere 300 ms after the point of maximum amplitude, and hence described the decay of the signal envelope close to the time when the drum was struck.

From the correlation coefficient values in Table 3, it is observed that the 10% sustain duration agrees well with the expert raters in both the sets, while the 1% duration correlates well with the rankings of only Set1 and is in fact negatively correlated with the rankings of Set2. On closer analysis of the exact ranks assigned by the raters to the strokes in Set2, it was observed that the stroke with the least value of 1% sustain duration was ranked in top 2 by all the experts. This stroke has a significant amount of modulation (pressure applied by the palm on the drum surface around the time of striking, as observed from the audio) present. Furthermore, in general, a stroke with a lesser value of 1% sustain duration but a higher degree of modulation was consistently ranked better by the experts, than another stroke with a longer 1% sustain duration but lesser modulation. This goes to show that for the bol 'Ghe', experts considered modulation more seriously even if it happened to reduce the overall sustain duration.

In contrast, the 10% sustain duration, which describes the signal envelope near the time of striking the drum, seems to capture information about modulation - it is higher for the strokes with more modulation, and hence in well agreement with the rankings of the expert raters, which indicates that the presence of modulation is actually causing the signal to decay more slowly near the time of striking.

As for the good correlation of both the duration measures with expert rankings in the case of Set1, it was observed that the strokes in Set1 had no considerable palm modulation.

6. Conclusions

A significant degree of agreement about relative stroke qualities was observed among the three raters in two of the three sets in case of the bols Ghe, Ta, Tun and Re, and in one of the three sets in case of Ke, Tin and Ti, while the overall level of agreement was very poor in the case of Tak. The high agreement between the raters in the sets mentioned despite the raters being from different backgrounds and the tabla-sets being of different qualities suggests the presence of acoustic features common to those bols across tabla-sets, which could also be computed and used to rate strokes computationally. One such feature – the duration of sustain was used to rank strokes of the bol Ghe and these rankings were then compared with the rankings of the raters. The duration of decay of signal amplitude from its maximum to one per cent of this value, which is indicative of the net duration for which the sound lasts, was found to be a measure of quality that is concordant with experts only in the case when the strokes did not contain any modulation. In the presence of modulation, the duration of decay from the maximum to ten per cent of the maximum was observed to be a more successful measure of quality that was concordant with experts' rankings. In this work, we considered only Ghe strokes for determining the correlation between duration of sustain and experts' rankings. For future work we would like to continue this research using the rest of the bols and more features.

7. Acknowledgements

The authors would like to thank Shri Sandip Ghosh and Shri Soumen Sarkar for sharing invaluable information about the tabla and for their help with the different recording sessions. We are equally grateful to Swarup Das, Sameel Sikkani, Bishwajeet Mitra for providing us with recordings, and Sreejan Deb, Prajwal Bharadwaj and Anoop Kashyap for participating in the ranking task. We are also indebted to Kaustuv Kanti Ganguli for providing his insights on Indian Classical music.

8. References

[1] Anirban Patranabis, e. a. (2005). *Harmonic and Timbre Analysis of Tabla Strokes.*eprint arXiv:1510.04880.

[2] Kendall, M. G. (1938, June). A New Measure of Rank Correlation.*Biometrika*.

[3] Kippen, J. (1988). *The Tabla of Lucknow: A Cultural Analysis of a Musical Tradition.*Cambridge University Press.

[4] Narang, K., & Rao, P. (2017.). Acoustic Features for Determining Goodness of Tabla Stroke.*18th International Society for Music Information Retrieval Conference.* Suzhou, China.

[5] P. Persad, J. B. (2008). Investigating the Use of a Robot with Tabla Education.*Digital Games and Intelligent Toys Based Education 2008 Second IEEE International Conference.*

[6] Patel, A., &Iversen, J. (2003). 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).* Barcelona.

[7] O. Gillet and G. Richard.(2003). Automatic labelling of tabla signals.*4th International Society for Music Information Retrieval Conference*

[8] Chordia, P. (2005).Segmentation and Recognition of Tabla Strokes.*6th International Society for Music Information Retrieval Conference*

[9] Jones E, Oliphant E, Peterson P, *et al.***SciPy: Open Source Scientific Tools for Python**, 2001-, http://www.scipy.org/