

Concatenated Tabla Sound Synthesis to Help Musicians

Uttam Kumar Roy

Dept. of Information Technology,
Jadavpur University, Kolkata, India.
u_roy@it.jusl.ac.in

Abstract. *Tabla* is the prime percussion instrument used in the music of many Asian countries. This cheap instrument can produce variety of pleasant sounds and has high sense of pitch. However, it is difficult to play and consequently there is a huge lack of tabla percussionists. Majority of musicians like to perform with a few known players. Since tuning requires long time and its range is also limited, players are to carry many sets of instruments tuned at different pitch for quick switch-over from one scale to another.

This paper proposes a scheme to synthesize pre-recorded *tabla* sounds to help musicians. Since recorded sound is used, the clarity and quality of the sound is unquestionable. The rhythm (taal), tempo or even pitch can be customized depending on musician's requirement. The method can also synthesize some table sounds which otherwise would be impossible to play by table percussionist.

Keywords: Sound, Signal, *Tabla*, Rhythm, Pitch, Tempo, Synthesis.

1 Introduction

To make music more promising, musical instruments often accompany soloists. One of the prime membranophone percussion instruments extensively used in classical music and traditional music of many Asian countries (such as India, Pakistan, Afghanistan, Nepal, Bangladesh, Sri Lanka, Indonesia etc.) is *Tabla* [Fig. 1]. This happens since it can create a wide variety of different pleasant sounds and rhythms, has a 'strong sense of pitch' and is also a very low-cost instrument.

Bad news is that playing *tabla* is not easy at all. It requires extensive and hard use of the fingers and palms in various configurations. This might be one of the primary reasons that there are quite a very few table percussionists in musician community. Novice singers often buy this cheap instrument expecting that some player will accompany him/her. However, they often don't find a player that makes them very uncomfortable and they often give up learning singing.

Tuning *tabla* does not only take moderately long time but also its range is limited. So, accompanists are to carry many sets of table tuned at different pitch to switch-over from one scale to another quickly.



Fig. 1. Percussion musical instrument ‘Tabla’. Bigger one is ‘dagger’(or bayan) and smaller one is ‘tabla’ (or dayan).

A very few attempts [Sathej & Adhikari] have been made to model this instrument. However, due to eccentric structure of *bayan* and lack of circular symmetry, all attempts have been failed to produce accurate sound.

In this paper, we have proposed a scheme to synthesize pre-recorded *tabla* sounds to accompany soloists. Since we use recorded sound, resultant synthesized sound has no difference from real performed sound. The rhythms (*taals*), pitch and tempo (*lay*) can be customized on-the-fly arbitrarily based on soloist’s requirement. Some non-standard sounds and alternative musical expressions can be produced while keeping the performance expression of the traditional Tabla interaction. Using this proposed scheme, a low cost, small, portable *tabla* synthesizer can be produced that is capable of replacing a percussionist and produce table sounds with arbitrary rhythm, scale and speed to accompany musicians.

2 Related Work

Although a few electronic *tabla* (Radel’s Taalmala Digi-60Dx and Digi-108, Sound Lab’s Sangat, Pakrashi’s Riyaz,) are available in Indian market, they can only produce limited number (60/100/108) of pre-synthesized rhythms. Some allow composing new but limited (only 2 to 8) rhythms. Although some of them allow limited pitch/tempo changing, they do not allow incorporating new *tabla* sounds.

The perfect model of *tabla* is not yet known. Raman [Raman, 1934] made the first scientific study of this family of drums. He and his coworkers obtained through a series of experiments, the eigen modes and eigen values of the *mridangam*. Ramakrishna and Sondhi [Ramakrishna & Sondhi] subsequently modeled the drum but agree with Raman’s experimental values to within 10%. The approximate solutions were provided by [B_S_Ramakrishna, Sarojini & Rahman], but the agreement with experimental values is also very poor.

In [Lehana_Dubey], a method for the separation of *tabla* sound from a mixer of vocal and *tabla* is presented. Although the separated *tabla* sound didn't contain any residual of vocal sound, the quality of the sound was poor.

In [Kapur_Ajay], authors describe the design of a simple electronic Tabla controller (*ETabla*). This is too a simple design and cannot even produce moderately realistic sound; hence cannot be used as professional controller.

In this paper, instead of using artificial model, we have used recorded *tabla* syllables to produce arbitrary rhythms with customizable pitch and tempo.

3 Proposed Scheme

Tabla manufacturers use a unique strategy to obtain harmonic overtones by loading the central part of the membrane with greater thickness (Fig. 2) resulting increased clarity of pitch and variety of tonal possibilities unique to this instrument [Sathej & Adhikari]. Further, the elements on the physics of the *tabla* can be found in [Fletcher and Rossing, 1998] or in the early work of Raman [Raman, 1934].

Playing *tabla* requires extensive use of the fingers and palms in various configurations to create a myriad of different sounds and rhythms, reflected in the mnemonic syllables (*bol*). To understand the technical part, we are describing some of the basic strokes:

Dagga strokes:

ge: holding wrist down and arching the fingers over the *syahi*, the middle and ring-fingers then strike the *maidan* (resonant)

ghe: similar to 'ge' except the heel of the hand is used to apply pressure or in a sliding motion on the larger drum to change pitch during the sound's decay (resonant)

ke or kath: striking with the flat palm and fingers (non-resonant)

Tabla strokes:

na: striking sharply with the index finger at the rim(resonant)

tin: striking gently with the index finger between *syahi* and rim(non-resonant)

te: striking the center of the *syahi* with the index finger (non-resonant)

tun: striking the center of the *syahi* with the index finger (resonant)

Combined strokes:

Some syllables are produced by striking both *tabla* and *dagga* simultaneously and are often called *combined syllables*. For example, 'dha' is a combination of 'na' and 'ge' where as 'dhin' is a combination of 'tin' and 'ghe'.

3.1 Rhythms (Taals)

A rhythm consists of series of syllables (*bols*) of different durations (1, $\frac{1}{2}$, $\frac{1}{4}$ note etc.). However, the rhythmic structure (called *taal*) can be quite complex. The basic rhythmic structures can have a large variety of beats (e.g. 6, 7, 8, 10, 12, 16,..) which are grouped in measures (called *Vivhaga*). Table 1 shows some of the popular *taals*.

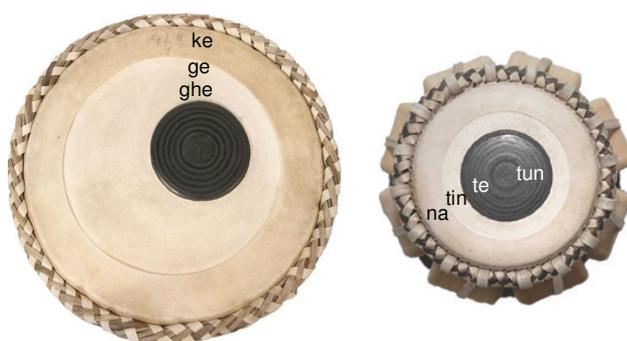
In this paper, we shall use *Keherwa* for demonstration.

Table 1. Some frequently used tabla rhythms.

Name	Beats	Division	Vivhaga
Dadra	6	3+3	X 0
Kewarha	8	4+4	X 0
Tintal	16	4+4+4+4	X 2 0 3
Ektal	12	3+3+3+3	X 2 0 3
Jhaptal	10	2+3+2+3	X 2 0 3
Rupak	7	3+2+2	X 2 3

3.2 Processing Rhythms

Indian music is primarily practice-oriented. This means the rules of compositions themselves are taught from teacher to disciple, in person. Accordingly, although oral notation for *tabla* stroke names is very developed, written transcription is not standardized. Fig. 2 shows the names of basic mnemonic syllables with their striking locations.

**Fig. 2.** Locations for playing basic mnemonic syllables.

In fact, the set of all syllables is fairly large and requires expertise to understand them. For quick understanding, we have used only seven basic syllables; four for *tabla* 'na', 'te', 'tin', 'tun' and three for *dagga* 'ge', 'ghe', 'ke'. However, the proposed scheme works for arbitrary syllables. These syllables were recorded in lossless WAV files. The signals corresponding to these basic syllables are shown in Fig. 3.

It may be noted that signals may not have equal play duration. For example, 'tun' had largest duration (more than 2.5 sec) where as 'ke' had shortest duration (0.045 sec). This means some sound echoes longer period of time than others. To understand how to generate resultant sound that can be played behind a song, let us consider a very simple transcription:

This is the notation for popular 8-beat (*matra*) rhythms called *keherwa*. The 8 beats (written as 4/4) consist of two segments (called *bibhag*) each consisting of 4 beats and

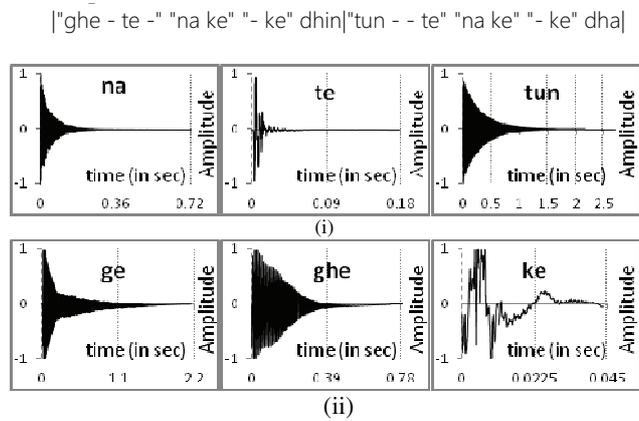


Fig. 3. Signals corresponding to (i) tabla (ii) daga mnemonic syllables.

is delimited by ‘|’ character. A beat may consist of one or more syllables. If a beat has multiple syllables, they are written within ‘ and ’ characters. A silence is represented by ‘-’ character.

Since notation uses some characters (such as ‘|’) other than the syllables, to divide the entire rhythm into parts, we extract only syllables before processing them further. Following shows the same using only syllables.

'ghe - te -' 'na ke' '- ke' 'dhin' 'tun - - te' 'na ke' '- ke' 'dha'

The timing diagram for the rhythm is shown here.



This shows that the rhythm consists of 8 beats. The first beat consists of a ghe, a silent, a te and again a silent.

3.3 Separation of strikes

Tabla transcription is said to be *monophonic*. This means a single symbol is used even if the corresponding stroke is compound. For example, *dagga* and *tabla* strokes are combined in *dha* (*na + ge*), and *dhin* (*tin + ghe*). So, it is necessary to separate *tabla* and *dagga* transcription as follows:

```
Initialize t and d as empty sequences;
for each bol in trancription
    if it is a tabla bol then append it to t;
    if it is a dagga bol then append it to d;
    if it is a combined bol then
```

Uttam Kumar Roy

```

begin
    [tb, db]=split (bol);
    append tb to t;
    append db to d;
end
endfor

```

Here is the separated transcription.

Tabla → '- - te -' 'na -' '- -' 'tin' 'tun - - te' 'na -' '- -' 'na'
 Dagga → 'ghe - - -' '- ke' '- ke' 'ghe' '- - -' '- ke' '- ke' 'ge'

3.4 Serialization

We now calculate the duration of each syllables assuming duration of a beat is unity. For example, since the first beat has four syllables, each of the syllables has duration 1/4. Following shows the syllables with their durations.

Tabla	-	-	te	-	na	-	-	-	tin	tun	-	te	-	na	-	-	-	na	
	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1	
Dagga	ghe	-	-	-	-	ke	-	ke	ghe	-	-	-	-	-	ke	-	ke	ge	
	1/4	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1

3.5 Normalization

A notation may contain (probably many) silences. During that time, no striking happens but echo of the previous syllable continues. So, if syllable has one or more silences, the duration for the syllable has to be re-calculated. For example, the following sequence of syllables

ghe	-	-	-	-
1/4	1/4	1/4	1/4	1/2

The, duration of each syllable is shown under it. The entire sequence can be replaced by a 'ge' with duration 1.5 (1/4+1/4+1/4+1/4+1/2) units. The number of samples required for this duration is 1.5*FS, where FS is the sampling frequency. If the signal 'ge' has at least those many samples, we take 1.5*FS samples from the beginning. Otherwise, we append necessary zeros at the end of 'ge' signal to make total number of samples equal to 1.5*FS. We call this process as *normalization*. Following shows the normalized duration:

tabla	-	te	na		tin	tun	te	na		na
	1/2	1/2	2		1	1/2	1/2	2		1
dagga	ghe		ke	ke	ghe		ke	ke	ge	
	3/2		1	1/2	5/2		1	1/2	1	

Time----->

3.6 Combining Components

This is where we get the resultant signal R which can be expressed as:

$$S_{\text{result}} = S_{\text{tabla}} + S_{\text{dagga}}$$

where S_{tabla} and S_{dagga} are the signals of *tabla* and *dagga* respectively. The signal S_{tabla} is obtained by concatenating all signals of *tabla bols*. The signal of a *bol* is obtained as follows:

For sampling rate F_s , and *bol* duration D_b sec, number of samples to be used is $S_b = F_s \cdot D_b$. Suppose, L_b is the actual number of samples present in the *bol*. Then there are two possibilities:

Case 1: $S_b \leq L_b$

This means, *bol* has enough samples to be taken. So, we take first S_b samples from L_b .

Case 2: $S_b > L_b$

This means, *bol* does not have enough samples to be taken. So, first take all L_b samples and append $(S_b - L_b)$ number of null samples (having amplitude zero). The signal S_{dagga} is obtained in a similar way. These two signals are then combined to form the resultant signal. Fig. 4(i) and Fig. 4(ii) shows *tabla* and *dagga* signals respectively whereas Fig. 4(iii) shows the resultant signal.

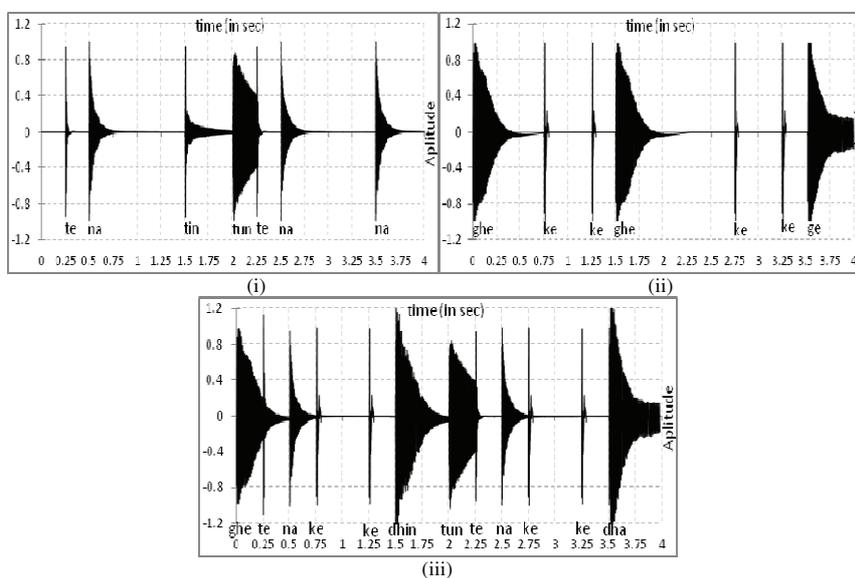


Fig. 4. Resultant signal (i) Tabla (ii) Dagga (iii) Combined.

3.7 Tuning Pitch

It refers to increase or decrease of sound frequencies to match with singer's voice or other instruments. Following factors should be considered:

- Two drums are tuned separately.
- The smaller drum is tuned to a specific note, usually the tonic, dominant or sub-dominant of the soloist's key.
- Changing the pitch must not change tempo.

Tuning must be done at frequency domain. Since input signals are in time domain, we first convert them to frequency domain, perform desired frequency shift and get it in the time domain again.

Since sound samples are uniform, we can use Fast Fourier Transform (FFT) which is an efficient implementation of Discrete Fourier Transform (DFT) to get frequency domain. Mathematically, for a set of N samples (complex numbers) x_0, x_1, \dots, x_{N-1} , its DFT is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, 1, 2, \dots, N-1 \quad (1)$$

And the inverse DFT is:

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{i2\pi n \frac{k}{N}} \quad n = 0, 1, 2, \dots, N-1 \quad (2)$$

Since vector indices start from 1 in Matlab, Equations (1) and (2) may respectively be rewritten as:

$$X_k = \sum_{n=1}^N x_n e^{-i2\pi(k-1) \frac{(n-1)}{N}} \quad k = 1, 2, \dots, N \quad (3)$$

$$x_n = \frac{1}{N} \sum_{k=1}^N X_k e^{i2\pi(n-1) \frac{(k-1)}{N}} \quad n = 1, 2, \dots, N \quad (4)$$

Let's quickly understand how MATLAB stores frequency domain data. In Equation (3), X_1 is the amplitude of the DC component and X_k ($1 < k \leq N$), is the complex amplitude corresponding to k^{th} frequency. The non-DC component has two parts; the first half X_k ($2 \leq k \leq N/2+1$) is the set of amplitudes for +ve frequencies and the second half X_k ($N/2+2 \leq k \leq N$) is the set of -ve frequencies. With these points in mind, we can increase the frequency as follows:

- Take the Fourier Transform
- Keep the DC component unchanged
- Shift the +ve part of the spectrum to the right

- Shift the -ve part of the spectrum to the left(or get reversed complex conjugate of the +ve part)
- Combine DC component, shifted first and second half
- Take inverse Fourier Transform.

To decrease the frequency, the direction of the shift has to be reversed. Fig. 5 (i) shows the FFT of original syllable ‘na’ and (ii) and (iii) show resultant FFT after a frequency shift of +200 Hz and -200 Hz respectively.

Figure 6 shows the resultant combined signal with a pitch increase of 10 Hz.

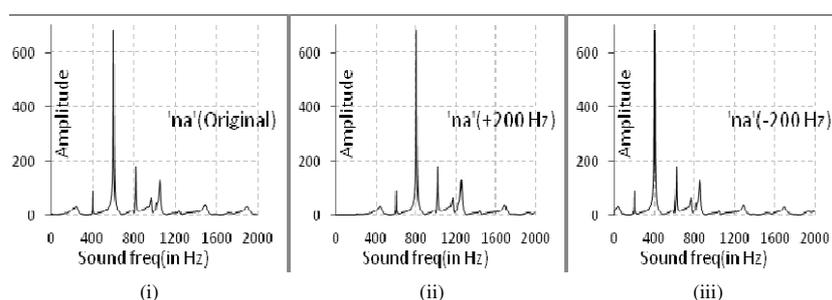


Fig. 5. Frequency drift for syllable ‘na’ (i) Original (ii) +200 Hz (iii) -200 Hz.

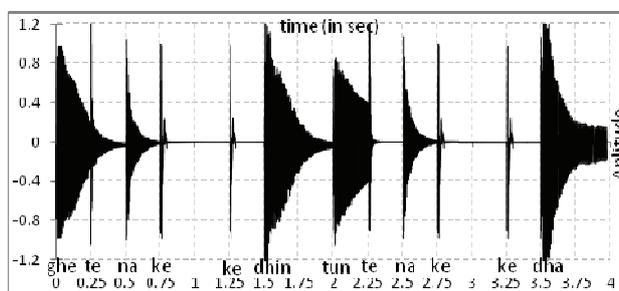


Fig. 6. A frequency increase of 10 Hz for tabla syllable sequence.

3.8 Changing Speed

For speeding up, we downsample (discard some samples) and for slowing down, we upsample (add/interpolate extra samples) so that the sample rate of the track is brought back to its original rate. We defined a factor called *speed_factor* that indicates the amount of speeding up/slowing down required. A +ve *speed_factor* implies speeding up and -ve *speed_factor* implies slowing down. So, the re-sampling rate may be calculated as:

$$\text{new_rate} = \text{original_rate}/\text{speed_factor};$$

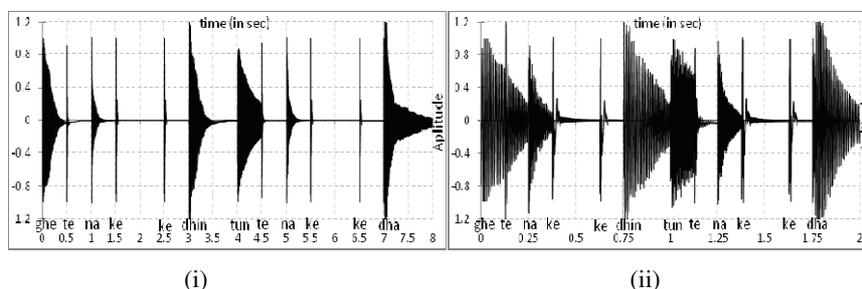


Fig. 7. Resultant signal with (i) half speed (ii) double speed.

For example, for a sampling rate (original_rate) 44.1 KHz, a signal having d sec duration has $d*44.1$ K samples. To speed up it a factor of 2, we have taken first $d*44.1/2 = d*22.05$ K samples to be played in d sec. More specifically, the signal ‘ge’ has the duration 2.1 sec and has the total $2.1*44.1 = 92.61$ K samples. To play it in double speed, we take first $92.61/2 = 46.205$ K samples and play it in 2.1 seconds.

The resultant combined signal is shown in Fig. 7 (i) and (ii) with half and double speed respectively.

Conclusion

In this paper, we proposed a concatenated synthesis pre-recorded *tabla* sounds to accompany soloists. Since we use recorded sound, it sounds like real performed one. The rhythm (*taal*), pitch and tempo (*lay*) can be customized arbitrarily. Nonstandard sounds and alternative musical expressions can be achieved while maintaining the performance expression of the traditional *tabla* interaction.

References

1. Fletcher, N.H., & Rossing, T.D. (1998). *The physics of musical instruments*. Second Edition. New York: Springer Verlag. [Rossing][Fletcher and Rossing, 1998]
2. Raman, C. (1934). The indian musical drum. In Science, P. I. A., editor, Reprinted in Musical Acoustics: selected reprints, ed. T.D. Rossing Am. Assn. Phys. Tech., College Park, MD, 1988. [Raman, 1934]
3. Sathej, G, and R. Adhikari. “The eigenspectra of Indian musical drums.” The Journal of the Acoustical Society of America 125.2 (2009): 831–838. [Sathej & Adhikari]
4. B. S. Ramakrishna and M. M. Sondhi, “Vibrations of Indian Musical Drums Regarded as Composite Membranes”, The Journal of the Acoustical Society of America 26/4, 523–529 (1954). [Ramakrishna & Sondhi]
5. B. S. Ramakrishna, “Modes of Vibration of the Indian Drum Dugga or Left-Hand Thabala”, The Journal of the Acoustical Society of America 29/2, 234–238 (1957). [B_S_Ramakrishna]
6. T. Sarojini and A. Rahman, “Variational Method for the Vibrations of the Indian Drums”, J.Acoust.Soc.Am 30/3, 191–196 (1958). [Sarojini & Rahman]

7. Parveen Lehana, Neeraj Dubey & Maitreyee Dutta, "Separation of Tabla from Singing Voice using Percussive Feature Detection", *International Journal of Computer Science & Communication*, 1/1, 219–222, 2010. [Lehana_Dubey]
8. Kapur, Ajay, *et al.* "The electronic tabla controller." *Journal of New Music Research* 32.4 (2003): 351–359. [Kapur_Ajay]
9. Radel Electronics Pvt. Ltd, <http://www.radelindia.com/>
10. Pakrashi and Co., <http://www.pakrashi-harmonium.com/>
11. Sound Labs, <http://www.soundlabs.in/>
12. RafAel Ferer, "Timbral Environments: An Ecological Approach to the Cognition of Timbre", *Empirical Musicology Review* Vol. 6, No. 2, 2011.
13. Mihir Sarkar, Barry Vercoe, "Recognition and Prediction in a Network Music Performance System for Indian Percussion", *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07)*, New York, NY, USA.