Composing Recorded Tabla Sound to Accompany Musicians

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

***Abstract*—The primary percussion instrument extensively used in the music of many Asian countries is *Tabla*. Although, this cheap instrument has high sense of pitch and can create variety of pleasant sounds, it hard to play. Consequently, there is a huge lack of percussionists. Most of the musicians like to perform with a few known players. Pitch tuning requires long time and its range is also limited. As a result, for quick switch-over from one scale to another, players are to carry many sets of instruments tuned at different pitch.**

In this paper, we have proposed a scheme to synthesize pre-recorded *tabla* sounds to assist soloists. The synthesized sound seems like real performed sound. The pitch, tempo and rhythm can be customized based on musician's requirement. In addition to the performance expression of the traditional *tabla* interaction, it can also provide non-standard sounds and alternative musical expressions.

Keywords—Tabla; Sound; Signal; Synthesis; Rhythm; Pitch; Tempo

I. INTRODUCTION

Musical instruments have often been the prime accompanying components of music to be more attractive. *Tabla* [Fig. 1] is 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. That is because it is a very low-cost instrument, but has a 'strong sense of pitch' and can create a wide variety of different pleasant sounds and rhythms.

However, playing *tabla* is not easy at all. It involves extensive use of the fingers and palms in various configurations. This is one of the primary reasons that there are quite a very few players in musician community. Although it is not difficult to buy this cheap instrument, there is a huge lack of players that makes the novice singers very uncomfortable and they often give up learning singing. Tuning *tabla* takes significantly long time and its range is also limited. Consequently, accompanists are to carry many sets of instruments tuned at different pitch for quick switch-over from one scale to another. Although, a very few attempts [Sathej & Adhikari] have been made to model this instrument, due to eccentric structure of *bayan* and lack of circular symmetry, they failed to produce realistic sound.

The tuning *tabla* takes significantly long time and its range is also limited. Consequently, accompanists are to carry many sets of instruments tuned at different pitch for quick switch-over from one scale to another.

Moreover, most musicians prefer to play with a few known percussionists. So, the percussionists are to travel a lot with the these instruments.

Fig. 1. Percussion musical instrument 'Tabla'. The bigger one is 'dagga' and smaller one is 'tabla'.

This paper proposes a methodology to synthesize pre-recorded *tabla* sounds to accompany soloists. Since, we use recorded sound; it has no difference from real performed sound. The rhythms (*taals*), pitch and tempo (*lay*) can be customized one-the-fly arbitrarily as and when required. Moreover, non-standard and alternative musical expressions can be achieved while maintaining the performance expression of the traditional Tabla interaction. A low cost, small, portable *tabla* synthesizer can be produced using this proposed scheme that will replace a percussionist and produce table sounds with arbitrary rhythm, scale and speed to accompany musicians.

II. 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 of pre-synthesized rhythms. Some allow composing new but limited (only 2 to 8) rhythms. Although, they allow

978-1-4673-6540-6/15/$31.00 ©2015 IEEE

limited pitch/tempo changing, but does 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 a too simple design and cannot even produce moderately realistic sound; hence cannot be used as professional controller.

In [ukr], we already proposed a scheme for synthesizing recorded table syllables. In this paper, instead of using artificial model, we have used recorded *tabla* syllables to produce arbitrary rhythms with customizable pitch and tempo. The performance of the previous method has also been improved significantly.

III. PROPOSED SCHEME

A special strategy is used in *tabla* to obtain harmonic overtones by loading the central part of the membrane with greater thickness (Fig 2) resulting increase clarity of pitch and variety of tonal possibilities unique to this instrument [Sathej & Adhikari]. Further 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 technique involves 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 stokes:

a) 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)

b) Tabla stokes:

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)

c) 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'.

B. Rhythms(Taals)

A *tabla* rhythm consists of sequence of syllables (called *bols*) of different durations (1, ½, ¼ note etc.). However, the rhythmic structure (called *taal*) can be quite complex. The basic rhythmic structures can have a large variety of beats (for example 6, 7, 8, 10, 12, 16,..) which are grouped in measures (called *Vivhaga*). Table 1 shows some of the popular *taals.*

TABLE I. SOME POPULAR TABLA TAALS

Name	Beats	Division	Vivhaga
Dadra	6	3+3	X 0
Keherwa	8	4+4	X 0
Tintal (or Trital or Teental)	16	4+4+4+4	X 2 0 3
Jhoomra	14	3+4+3+4	X 2 0 3
Tilwaada	16	4+4+4+4	x 2 0 3
Dhamar	14	5+2+3+4	X 2 0 3
Ektal and Chautal	12	2+2+2+2+2+2	X 0 2 0 3 4
Jhaptal	10	2+3+2+3	X 2 0 3
Rupak (Mughlai/Roopak)	7	3+2+2	X 2 3

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

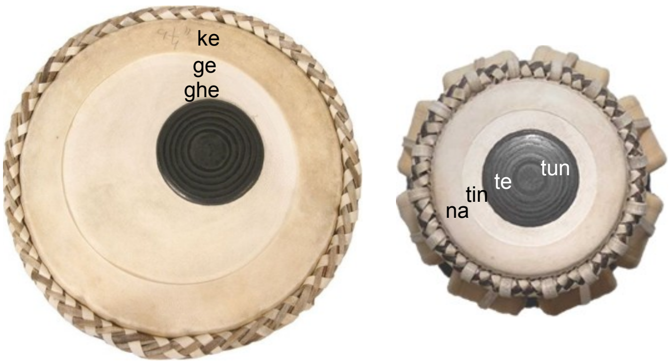


Fig. 2. Positions for plyaing some basic tabls mnemonics

C. Processing Rhythms

Traditional Indian music is primarily practice-oriented and 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 notation is not standardized. Fig. 2 shows the names of some basic mnemonic syllables with the striking locations to play them.

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.

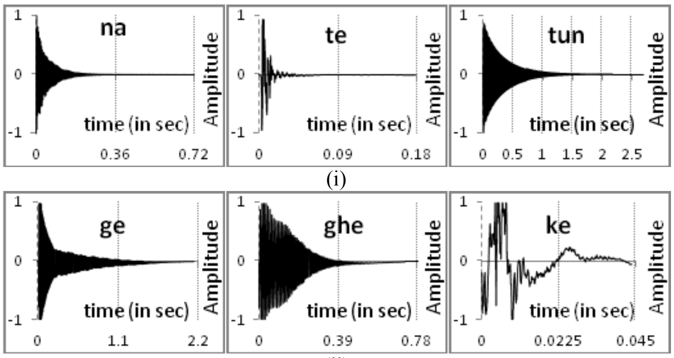


Fig. 3. Signals corresponding to (i) tabla (ii) dagga mnemonic syllables

It may be noted that signals do not have equal play duration. For example, 'tun' has largest duration (more than 2.5 sec) where as 'ke' has 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 notation:

|"ghe - te -" "na ke" "- ke" dhin|"tun - - te" "na ke" "- ke" dha|

This is the notation for popular 8-beat (*matra*) rhythms called *keherwa*. Note that this not the basic notation for *keherwa*; but one of its many variations. Since, it uses many syllables, it is good for demonstration. The 8 beats (written as 4/4) consist of two segments (called *bibhag*) each consisting of 4 beats and 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.

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

Time------------------------------->

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

D. Separation of strikes

Note that strokes on the *dagga* and *tabla* can be combined, like in the bols : *dha* (*na* + *ge*), *dhin* (*tin* + *ghe*). Due to these characteristics, even if two drums are played simultaneously, the transcription is monophonic - a single symbol is used even if the corresponding stroke is compound. So, it is necessary to separate *tabla* and *dagga* transcription as follows:

```
Initialize t and d as empty sequences;
for each bol in tranacription
    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
      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'

E. 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 ¼. Following shows the durations of all *tabla* syllables:

-	-	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

Similarly, following shows the durations of all *dagga* syllables:

ghe	-	-	-	-	ke	-	ke	ghe	-	-	-	-	-	ke	-	ke	ge
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

F. 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

ge	-	-	-	-
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 (¼+¼+¼+¼+½) 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:

<table>
<tr><td rowspan="2">tabla</td><td>-</td><td>te</td><td colspan="4">na</td><td colspan="2">tin</td><td>tun</td><td>te</td><td colspan="4">na</td><td colspan="2">na</td></tr>
<tr><td>1/2</td><td>1/2</td><td colspan="4">2</td><td colspan="2">1</td><td>1/2</td><td>1/2</td><td colspan="4">2</td><td colspan="2">1</td></tr>
<tr><td rowspan="2">dagga</td><td colspan="3">ghe</td><td colspan="2">ke</td><td>ke</td><td colspan="5">ghe</td><td colspan="2">ke</td><td>ke</td><td colspan="2">ge</td></tr>
<tr><td colspan="3">3/2</td><td colspan="2">1</td><td>1/2</td><td colspan="5">5/2</td><td colspan="2">1</td><td>1/2</td><td colspan="2">1</td></tr>
</table>

Time------------------------------->

G. Combining components

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

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

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:

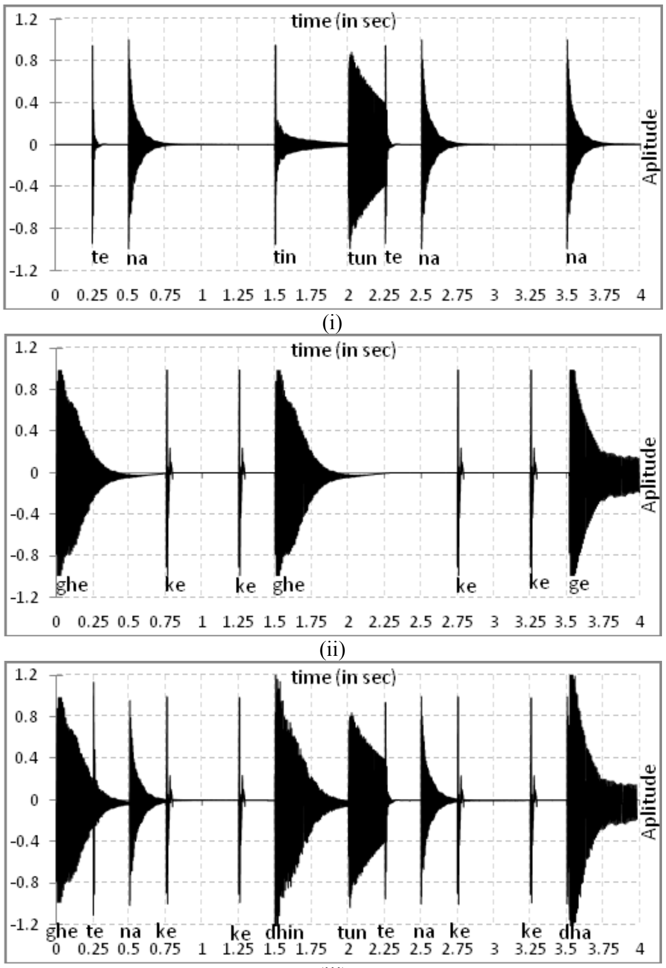
For sampling rate F_s, and *bol* duration D_b sec, number of samples to be used is $S_b=F_s.D_b$. Suppose, L_b is the actual number of samples present in the *bol*. Then there are two possibilities:

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).



(iii)

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

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.

H. 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 subdominant 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,.., x_{N-1}, its DFT is defined as:

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

And the inverse DFT is:

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

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

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

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

Let's quickly understand how MATLAB stores frequency domain data. In (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.

For example, if N=10 (even), there will be 5 +ve frequency components and 4 -ve frequency components. If N is 9(odd), then there will be an equal number (four) of +ve and -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), (iii) and (iv) show resultant FFT after a frequency shift of +200 Hz, +400 Hz and -200 Hz respectively.

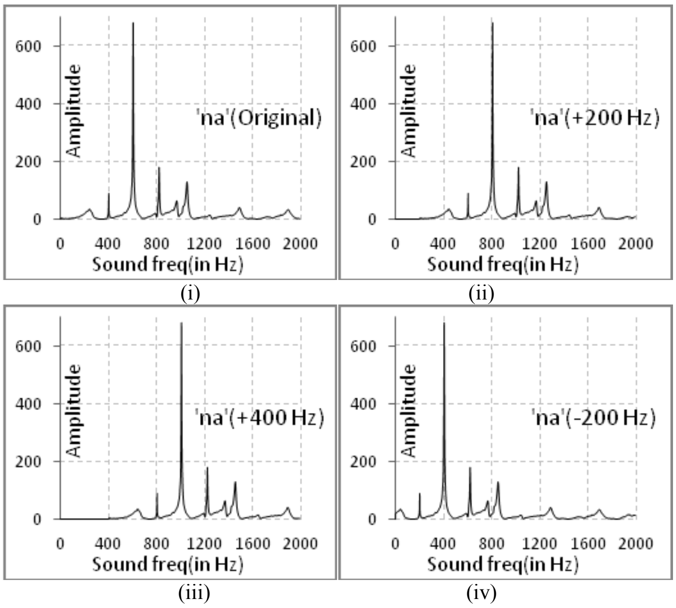


Fig. 5. Frequency drift for syllable ‘na’ (i) Original (ii) +200 Hz (iii) +400Hz (iv) -200 Hz

Fig. 6 shows the resultant signal for the notation considered.

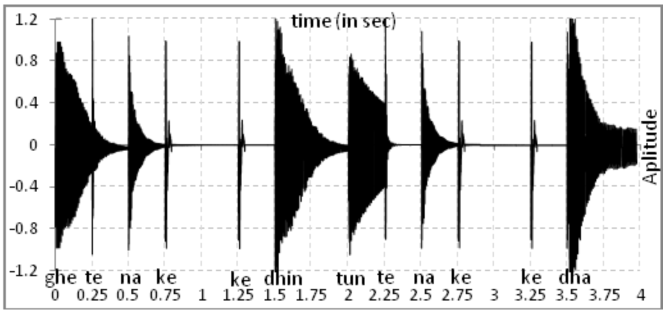


Fig. 6. A frequency increase of 10 Hz for all tabla syllables

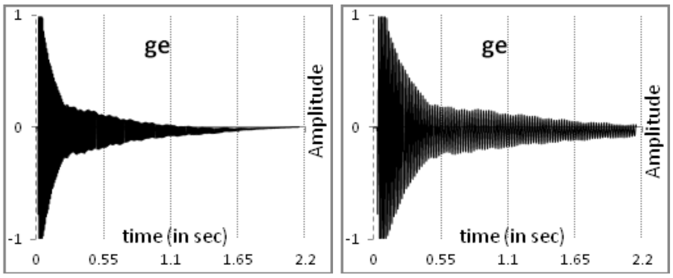
I. Changing speed

Not all music is played in equal speed. So speed regulation is another fundamental requirement of music synthesis. Although, media player often offers facility to control speed, it either degrades sound quality or introduces noise at very high/low speed. This happens since media player, without knowing anything about the sound content, deletes/introduces some samples. In general, deleting/adding up some samples causes deletion/introduction of some frequencies. This is popularly known as ‘lossy re-sampling’ which may loss some frequencies.

Ideally, change of speed should be done such way that ‘rate of the length’ and ‘pitch’ of audio remains unchanged. So, a careful ‘re-sampling’ is required. For speeding up, the samples are "squashed" into a shorter time and for slowing down, the samples are "stretched" to a longer duration.

Fortunately, ‘re-sampling’ for tabla sound can be done very efficiently. Since ‘Tabla’ is percussion instrument, the sound amplitude gets maximum value at the moment of strike and decays gradually. This implies, former part of the signal contributes more to frequency domain than the later. So, re-sampling should keep the former part unchanged as much as possible. If some samples have to be removed/added, better to do it from trailing end.

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.



(i) (ii)

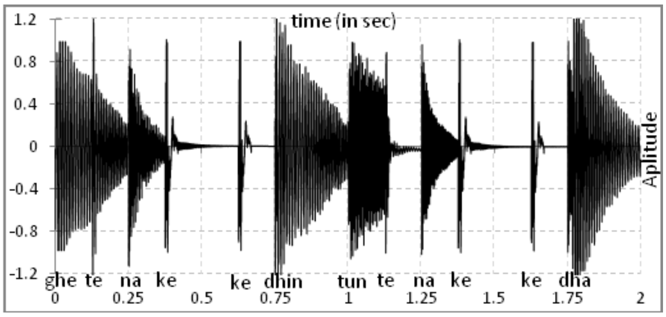
Fig. 7. Resampling (i) original signal (ii) after doubling the speed

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:

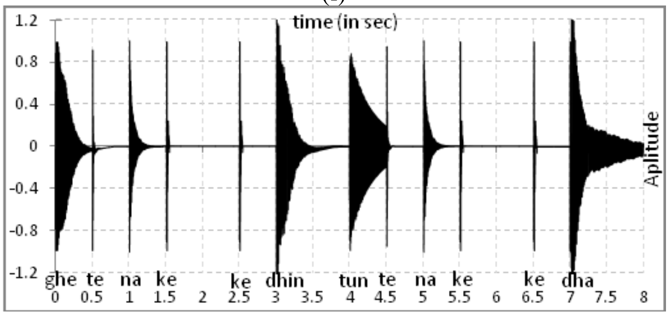
```
new_rate = original_rate/speed_factor;
```

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, consider the Fig 7 (i). The signal 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 signal is shown in Fig. 7 (ii).

The resultant combined signal is shown in Fig 8 (i) and (ii) with half and double speed respectively.



(i)



(ii)

Fig. 8. Resultant signal with (i) double speed (ii) half speed

IV. 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, 191-1996(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

[14] Uttam Kumar Roy, "Concatenated Tabla Sound Synthesis to Help Musicians", Research in Computing Science[ukr]