

Acoustic and Perceptual Comparison of Speech and Drum Sounds in the North Indian Tabla Tradition: An Empirical Study of Sound Symbolism

Aniruddh D. Patel and John R. Iversen

The Neurosciences Institute, 10640 John Jay Hopkins Dr., San Diego, CA 92121, USA

E-mail: apatel@nsi.edu, iversen@nsi.edu

ABSTRACT

North Indian tabla drumming is an oral tradition which uses a system of nonsense syllables (vocables) to name drum sounds. We tested the hypothesis that the vocables are a case of sound symbolism (onomatopoeia). Eight vocables and their corresponding drum sounds were collected from 6 professional tabla players in India. Analysis revealed that acoustic properties of drum sounds were reflected by a variety of phonetic components of vocables: spectral centroid, rate of amplitude envelope decay, duration between the releases of consonants in a cluster, fundamental frequency, and the influence of aspiration on the balance of low vs. high frequency energy in a vowel. We also describe a perceptual experiment which demonstrates that naïve listeners can match vocables to their corresponding drum sounds. Taken together, these results provide strong evidence of sound symbolism in the North Indian tabla drumming tradition.

1. INTRODUCTION

1.1 GENERAL BACKGROUND

Numerous cultures use nonsense syllables or ‘vocables’ in systematic ways to represent musical sounds [1,2,3,4]. A familiar Western example is solfège, in which the notes of the scale are represented by the syllables do, re, mi, etc. Two contrasting hypotheses can be proposed concerning vocables. The first is that their relationship to musical sounds is arbitrary, as appears to be the case in Western solfège [1]. The alternative hypothesis is that they acoustically resemble the musical sounds they represent, i.e. they are based on sound symbolism or onomatopoeia [5].

Recent empirical research has found evidence of sound symbolism in the vocables used to represent percussion instruments in Peking opera [3] and pitches in Japanese Noh flute music [1]. Thus the two hypotheses are not mutually exclusive; rather, the question is one of relative frequency. Another relevant question is how sound symbolism manifests itself in different cultures. Since phoneme inventories differ across languages, what similarities and differences emerge in the way people use the sounds of their language to represent musical sounds?

1.2 NORTH INDIAN TABLA DRUMMING

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). Teaching and composition on this instrument is based on oral tradition, and employs a linguistic syllabary in which particular drum strokes are named by particular Hindi syllables, known as ‘bols’ (for example, ‘Tin’, ‘Tun’, ‘Ta’, ‘Dha’, cf. Table 1 for IPA notation). The size of the bol inventory is between 10-20 items [2,6].

Past research has shown that a vocable may be pronounced with a different vowel or consonant depending on context, without altering the drum stroke [6]. Furthermore, bols and drum strokes vary somewhat between different schools of tabla (gharanas). Thus the mapping between syllables and drum sounds is not one-to-one. However, there are certain basic bols which are widely used and which we focus on in this study.

To our knowledge, there are no previous empirical studies comparing tabla sounds to the spoken syllables used to signify them. The goal of the current study is to determine if the mapping between drum sounds and speech sounds in tabla music has an acoustic and perceptual basis, i.e. if it is an example of sound symbolism.

2. METHODS

2.1 DETAILS OF SPEECH AND DRUM SOUNDS

Vocables and drum sounds were recorded by six professional tabla players (one female) in a sound recording studio in Bombay, India in April of 2002. After being asked to play all the basic strokes, a list of vocables (n = 11) was determined by the players in conference with each other. Each player produced 4-5 repetitions of each vocable and its corresponding drum sound in alternation, pausing long enough to avoid any overlap in sound. All players used the same set of drums; occasionally a player would re-tune the drum before recording. Sounds were sampled at 44,100 Hz and recorded directly to computer in .wav format. A digital video recording was also made. From this corpus, eight common vocables were chosen for this study, and are presented in pairs in order to focus on specific acoustic

contrasts. These pairs and the contrasts are: Tin-Tun (front vs. back high vowel), Kat-Ghe (voicing of stop consonants, presence of coda), Tra-Kra (different onset clusters), and Ta-Dha (voicing and aspiration of stop consonants).

Figure 1 shows a top view of the tabla drums, and Table 1 gives details on how the drum was struck for each of the vocables examined in this study.

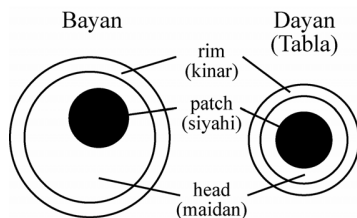


Figure 1: Top view of dayan (right) and bayan (left). Each drum surface is divided into three regions: a rim, a head, and a central circular patch of iron filings and paste. Indian names of the regions are also given, per ref [3]. Surface diameters: dayan ~14 cm, bayan ~22 cm.

Bol	Manner of playing
TIN /t̪ɪn/	<i>Dayan:</i> Index finger strikes head. Third finger damps the fundamental by resting on head.
TUN /t̪un/	<i>Dayan:</i> Index finger strikes patch; no damping.
KAT /kət/	<i>Bayan:</i> Damped stroke played with flat hand across rim, head, and patch.
GHE /g ^h e/	<i>Bayan:</i> Index finger strikes head at far edge of patch. Heel of hand, on head, can alter pitch.
TRA /t̪rə/	<i>Dayan:</i> Damped stroke: middle and index fingers strike patch in rapid succession, and remain on drum.
KRA /krə/	<i>Bayan:</i> Damped flat-hand stroke similar to KAT. <i>Dayan:</i> Damped stroke. Fingers strike patch together and remain in contact.
TA /t̪a/	<i>Dayan:</i> Index finger strikes rim. Third finger damps fundamental by resting on head.
DHA /d ^h a/	Simultaneous striking of TA and GHE.

Table 1: Drum stroke details (from video). Fingers bounce off drum (open stroke) unless the stroke is damped (closed). IPA in slashes. Examples: www.nsi.edu/users/patel/tabla/.

2.2 ACOUSTIC ANALYSIS

Spectral (FFT) and waveform measurements of sounds were made using SIGNAL (Engineering Design) on a modified PC. All spectral power measurements were made in units of decibels (dB) re a 1 Volt RMS signal. The number of tokens measured for each sound was typically 30, representing 5 tokens each from 6 speakers. Occasionally a speaker did not produce a given drum stroke or vocable, so the number of tokens was less (but never less than 20).

Tin vs. Tun (dayan). In Tin, damping with the third finger suppresses the fundamental frequency of the dayan (while allowing the higher harmonics to ring), yielding a brighter,

higher-pitched tone; in Tun the fundamental and upper harmonics all ring strongly [7]. To quantify these differences, we made spectral measurements of the frequency and power of the first four partials of each drum sound (16384 point (pt) FFT, starting at onset of sound).

Tin vs. Tun (speech): To examine differences in the frequency structure of spoken Tin vs. Tun, we examined the frequency and power of vowel fundamental frequency (Fo) and formants F1-F3 at the onset of stable formant structure. Fo measurement was based on a 2048 pt FFTs, while formant measurements were based on the average of 25 wideband spectra (256 pt FFT) spaced 2 ms apart.

Kat vs. Ghe (bayan): Kat is a damped stroke, while Ghe is an open stroke (the bayan is allowed to ring). To quantify the abrupt offset of Kat relative to Ghe, we computed an amplitude envelope for each sound and measured the duration between the envelope peak and the point where the envelope had decayed to 50% of the peak amplitude.

Kat vs. Ghe (speech): we performed the same measurement as for drummed Kat and Ghe as described above.

Tra vs. Kra (dayan and bayan): both of these strokes involve dual, damped impacts on the drums. In Tra, both hits are on the dayan, while in Kra, the bayan is hit first, followed immediately by the dayan. We measured the time between impacts using waveforms and wideband spectrograms (64 pt FFT).

Tra vs. Kra (speech): both vocables have complex onsets, with two consonant releases: one for the initial stop consonant, and one for the /r/. For both vocables we measured the duration between these two release points using waveforms and wideband spectrograms (256 pt FFT). Stop consonant release was identified via a burst of energy in the waveform and spectrogram. /r/ release was identified using amplitude and spectral cues (a rapid change in the proximity of F2 and F3).

Ta vs. Dha (dayan and bayan): Dha is a combination stroke consisting of simultaneous striking of Ta on the dayan and Ghe on the bayan. The Ta stroke is similar to Tin: both involve striking the dayan while damping the drum head with the third finger, giving the two sounds a very similar sound. Ghe is an open stroke on the lower-pitched bayan. Thus Ta and Ghe provide higher vs. lower frequency components of Dha.

Inspection of narrow-band spectrograms revealed that the fundamental of the low-pitched bayan rose and fell in frequency at the onset of Dha before attaining a steady value, likely due to subtle changes in the pressure of the heel of the hand after the left index finger strikes the drum. During this frequency glide the intensity of the bayan appeared stronger than during the later steady-state portion.

We quantified these patterns by measuring the duration of the bayan frequency glide in Dha from narrow-band spectrograms (2048 pt FFT), and by measuring the frequency and power of the bayan fundamental from

spectra taken at two points: 50 ms after the onset and offset of the frequency glide (2048 pt FFT). We also measured the frequency and power of the first strong partial (partial #2) of the dayan at these same points.

Ta vs. Dha (speech): The use of a voiceless vs. a voiced aspirated (breathy-voiced) stop in the initial consonant of these vocables had strong consequences on the temporal and spectral structure of the vocable. Notably, in Dha the region of heavy aspiration following consonant release resulted in a speech waveform lower in amplitude and more sinusoidal than later in the vowel, when aspiration ceases and modal (standard) phonation is achieved [cf. 8].

For Dha we measured the duration of aspiration from wideband spectrograms (256 pt FFT) as the time between the release burst and the onset of F1 in modal phonation. We measured Fo frequency and power of Dha at two points: 25 ms after the consonant release burst (i.e. during heavy aspiration) and 25 ms after the onset of F1 (1024 pt FFT); we also measured average spectral energy in the F1/F2 range (515 to 1551 Hz) at these two time points, based on the average of 12 wideband spectra (256 pt FFT) separated by 2 ms each. To compare Fo in Ta and Dha, we also measured the Fo of Ta 25 ms after F1 onset (1024 pt FFT).

2.3 PERCEPTION EXPERIMENT

We tested the ability of 7 non-Hindi speakers unfamiliar with tabla drumming (2 females) to match vocables to their corresponding drum sounds. Individuals were presented with two vocables and two drum sounds at a time on a computer: each sound was associated with a number (no images were presented). Individuals were told to listen to each sound as many times as they liked by pressing the associated number on the keyboard, and then decide which syllable went with which drum sound. Each vocable pair (Tin-Tun, Kat-Ghe, Tra-Kra, and Ta-Dha) and its associated drum sounds was presented 8 times, using different tokens from different speakers/players (within each trial, all tokens came from the same speaker/player), yielding 32 trials. All drum sounds were truncated at 550 ms (the duration of the longest vocable token), using a 25 ms tapered window. All sounds were normalized to 0.5 Volts RMS amplitude. The participants had no history of hearing problems.

3. RESULTS

3.1 ACOUSTIC ANALYSIS RESULTS

Unless otherwise stated, all significant differences reported below represent unpaired t-tests with $p < 0.001$.

Tin vs. Tun (Figure 2) Drummed Tin had a significantly higher spectral centroid than Tun (due to damping of the fundamental in Tin). Examination of the corresponding speech vocables revealed that the “formant centroid” (mean of F1 to F3 frequency weighted by power) was significantly higher in spoken Tin than Tun, largely due to the much higher F2 in /t/ vs. /u/. Analysis of Fo in spoken Tin vs. Tun revealed no significant difference.

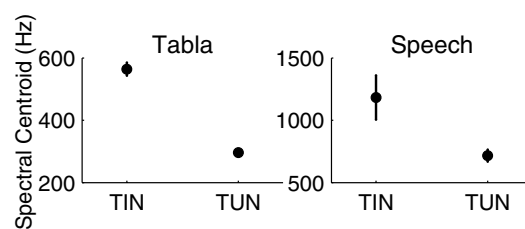


Figure 2: Spectral centroid (mean and se) of drum sounds and formant centroid of vocables for Tin and Tun.

Kat vs. Ghe (Figure 3) Measurement of the waveform amplitude envelope half-decay time of the drummed tokens revealed that Kat had a significantly shorter decay time than Ghe, as would be expected since Kat is a non-resonant (closed) stroke. The identical measure taken from the speech waveforms revealed that spoken Kat also had a significantly more abrupt offset than Ghe.

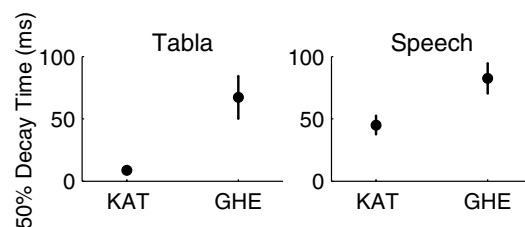


Figure 3: Amplitude envelope half-decay time of Kat and Ghe drum sounds and vocables.

Tra vs. Kra (Figure 4) Measurement of the time between impacts/releases of these dual strokes/consonant clusters revealed a significantly greater delay in both drummed and spoken Kra than in Tra.

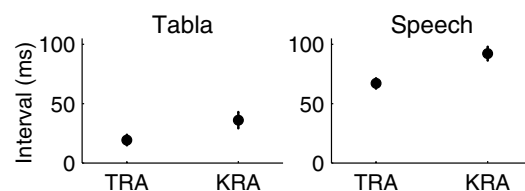


Figure 4: Time between impacts of drum sounds and between consonant releases of vocables for Tra and Kra.

Ta vs. Dha The Ta stroke is similar to Tin (cf. Table 1), and as a result, both are bright sounds with a very similar spectral centroid (data not shown). Compared to drummed Ta, Dha has lower frequency energy due to the use of the bayan. Analysis of spoken Ta vs. Dha revealed that Dha had a significantly lower Fo during the aspirated portion (151 vs. 113 Hz). Another difference between Ta and Dha was that the latter had two distinct acoustic stages following onset, both in drumming and speech. For drummed Dha, phase 1 was marked by frequency modulation (FM) in the bayan (mean duration 213 ms, se 7.8), during which energy in the bayan fundamental was relatively strong (cf. Methods). We quantified this by computing the ratio of the

energy in the bayan fundamental to energy in the dayan's first strong partial, during vs. after the FM. This ratio was significantly larger during the frequency glide (Figure 5).

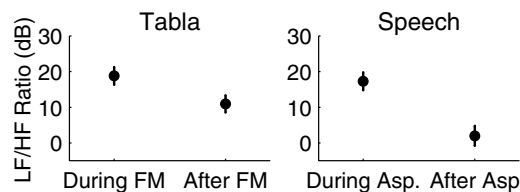


Figure 5: Ratio of low to high frequency energy during two stages of drummed and spoken Dha.

Spoken Dha also had a two-stage temporal structure, with the first phase marked by heavy aspiration (mean duration 88 ms, se = 6.5). The ratio of energy in Fo vs. in the F1/F2 region during and after aspiration was significantly larger during the aspirated portion (Figure 5).

3.2 PERCEPTION EXPERIMENT RESULTS

Figure 6 shows the percentage correct for the task of pairing vocables with their associated drum sounds. With the exception of Tra - Kra, naïve listeners were able to match syllables and drum sounds with better than chance performance (Chi Squared test, all p values < 0.0001).

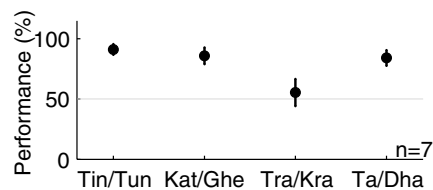


Figure 6: Percent correct (mean and se) in the perceptual task of matching vocables to corresponding drum sounds.

4. CONCLUSIONS

We have found strong acoustic and perceptual evidence of sound symbolism in the vocables (syllables) used to name drum strokes in North Indian tabla drumming. Empirical study of 8 vocables and drum sounds reveals a range of phonetic components involved in onomatopoeia, including spectral centroid, rate of amplitude envelope decay, duration between the release of consonants in a cluster, fundamental frequency, and the influence of aspiration on the balance of low vs. high frequency energy in a vowel.

A perceptual experiment revealed that naïve listeners could correctly match speech sounds to drum sounds in 3 out of 4 of the vocable pairs studied. The one exception, Tra vs. Kra, likely proved difficult because the difference in duration between the two impacts/consonant releases was small.

The role of aspiration in sound symbolism takes advantage of the phonetics of Hindi. In this case, aspiration serves to create a dominance of low-frequency energy early in the vocable Dha. This is because heavy aspiration leads to

incomplete glottal closure and hence poor excitation of the vowel formants [8]. In drummed Dha, there is also a dominance of low frequency energy early in the sound, during a short frequency glide by the bayan. In both cases, there is a division of the sound into two distinct regions with a different ratio of low to high-frequency energy.

Given the strong evidence for sound symbolism in tabla, it is likely that onomatopoeia played a role in the origin of tabla vocables. While other factors are also likely to have played a role in vocable choice (e.g. the need for contrastive syllables which could be spoken very rapidly as well as historical and cultural factors), an interesting question is whether sound symbolism confers an advantage in learning the system of sound-drum pairings and in encoding long drum sequences as sequences of vocables.

ACKNOWLEDGEMENTS

Deep thanks to Anuradha Pal for organizing the recording session in Bombay, and to all the tabla players for their participation. We also thank J. Rosenberg & A. Lerner for help with acoustic measurements, and B. Repp for comments. This research was supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham fellow and JRI is the Joseph Drown fellow.

REFERENCES

- [1] D. Hughes, "No nonsense: the logic and power of acoustic-iconic mnemonic systems," *British Journal of Ethnomusicology*, vol. 9, no. 2, pp. 95-122, 2000.
- [2] J. Kippen, *The Tabla of Lucknow: A cultural analysis of a musical tradition*, Cambridge: Cambridge University Press, 1988.
- [3] G. Li, *Onomatopoeia and Beyond: a Study of the Luogu Jing of the Beijing Opera*, Ph.D. dissertation, UCLA, 2001.
- [4] J. Sundberg, "Musical Significance of Musicians' Syllable Choice in Improvised Nonsense Text Singing: A Preliminary Study," *Phonetica*, vol. 54, pp. 132-145, 1994.
- [5] L. Hinton, J. Nichols, and J. Ohala, Eds., *Sound Symbolism*, Cambridge: Cambridge University Press, 1994.
- [6] A. Chandola, *Music as Speech: An Ethnomusicolinguistic Study of India*, New Delhi: Navrang, 1988.
- [7] D. Courtney, *Advanced Theory of Tabla*, Houston: Sur Sangeet Services, 2000.
- [8] P. Ladefoged, *Vowels and Consonants: An Introduction to the Sounds of Languages*, Malden MA: Blackwell Publishers Inc., 2001.