Tabla Gyan: An Artificial Tabla Improviser

Parag Chordia and Alex Rae

Georgia Institute of Technology, Atlanta GA, 30332 USA
ppc@gatech.edu, arae3@gatech.edu
http://paragchordia.com/

Abstract. We describe *Tabla Gyan*, a system capable of improvising tabla solo music, a sophisticated percussion tradition from North India. The system is based on a generative model of the qaida, a central form in tabla solo based on thematic variation. The system uses a recombinative process of variation generation, and filters the results according to rhythmic and timbral characteristics of each phrase. The sequences are used to generate audio in realtime using pre-recorded tabla samples. An evaluation of the system was conducted with seventy users, primarily experienced tabla performer and listeners. With respect to qualities such as musicality, novelty, adherence to stylistic norms, and technical ability, the computer-generated performances compared favorably with performances by a world-class tabla player.

1 Introduction

This work aims to explore computational models of creativity, realizing them in a system designed for realtime generation of improvised music. This is envisioned as an attempt to develop musical intelligence in the context of structured improvisation, and by doing so to enable and encourage new forms of musical control and performance. A model of qaida, a traditional north Indian solo tabla form, is presented along with the results of an online survey comparing it to a professional tabla player's recording on dimensions of musicality, creativity, and novelty. The model is based on generating a bank of variations and filtering according to musical qualities.

2 Background

2.1 Theories of Creativity

This work is fundamentally motivated by an interest in exploring computational models of creativity. There have been many attempts to characterize the basic nature of creativity, and here we identify some key insights.

Mihaly Csikszentmihalyi [4] outlined a theory formulating creativity as a concept arising from the interaction of a *domain*, such as music or a particular musical genre, the *individual* who produces some possibly creative work, and the *field* within which the work is judged. One significance of this is that it moves creativity from being a purely individual characteristic to one largely the product of external interactions; notably, the final determination of whether the individual has been creative rests on the judgement of peers.

Many theories are based in the idea of multiple creativities. *Geneplore* [6], for example, models creativity as comprised of a generative phase in which a large set of potential materials is amassed, and an exploratory phase in which this set is explored and interpreted. There is notable similarity between this and elements of our system described in Sections 4.1 and 4.2. Sternberg presents a theory[13] that represents creativity in terms of three processes for finding insights in large quantities of information: selective encoding, combination, comparison. Insights found by filtering information are then combined to generate new insights, which in turn are compared to previous or distant insights to create yet another insight. Gardner [7] also addresses creativity, characterizing it as the production of novelty within a domain, similarly to Csikszentmihalyi's approach.

More practical but equally valid definitions have focused on the concept of novelty. A common formulation defines creativity as an action or process which produces novel output that satisfies the constraints of context [3]. Addressing the basis for judging whether an artificial system could be considered creative, Pereira [11] identifies the requirements that when given a problem, answers produced by the system should not replicate previous solutions of which it has knowledge, and should apply acceptably to the problem. These are notably similar conceptualizations of creativity, and share the idea that the existence of creativity can, and should, be evaluated on the basis of the product.

2.2 Machine Musicianship

Many systems have been developed which can claim to involve computational creativity. We mention here a few in order to indicate the range of approaches and goals which others have undertaken.

The Continuator [9], developed by François Pachet tries to come up with improvisatory responses to human pianist's playing, using weighted random draws from a prefix tree built from phrases detected in the audio input. Arne Eigenfeldt's multi-agent "Kinetic Engine" [5] models the interactions between networked improvising agents in terms of both musical features and social dynamics, allowing shared parameters such as tempo and overall contour to be controlled by a "conductor" agent. David Cope's long-running project Experiments in Musical Intelligence (EMI) focuses on faithful emulations of styles in the Western classical canon [1]. His approach focuses on analyzing a large corpus of works to extract patterns which encode the main elements of the style, recombining them to create derivative works [2]. Cope has written and worked extensively in this field, and identifies a number of basic elements which he determines to be central to computational creativity, specifically calling out pattern-matching and recombinance [3].

3 Introduction to Tabla

Tabla is the predominant percussion instrument of North India. Physically, tabla is actually a pair of drums, as seen in Figure 1. It is played with the hands and fingers, and each drum is associated with one hand. The right-hand drum, called the *tabla* or *dayan*, is higher in pitch than the left-hand drum, or *bayan*. Both drums are capable of producing a variety of distinct timbres, ranging from ringing sounds with a clear pitch to short sharp sounds with a high noise content. There are specific striking techniques for producing each of the different timbres, known generally as strokes, and each is named. There are three broad classes of strokes: resonant strokes with a clear pitch and ringing tone, shorter non-resonant noisy strokes, and bass strokes produced on the *bayan*. Individual strokes and common short phrases are known as *bols*, and form the building-blocks of larger phrases. Improvisation in tabla music takes places within a rhythmic cycle which defines a large-scale periodicity, consisting of a set number of beats. The most common cycle is *Teental*, consisting of sixteen beats. To make the cycle easier to perceive, *bayan* strokes on certain beats are damped, and are referred to as "closed". Strokes in which the bass is allowed to sound are referred to as "open".

Fig. 1. A tabla. The drum on the left is the *bayan*; the drum on the right is the *dayan*.

There is a rich tradition of solo tabla performance. The tabla is then usually accompanied by a melodic instrument that plays a repeated figure known as *nagma* which occupies the role of a timekeeper. One of the most prominent compositional forms presented in a solo tabla performance is qaida, a structured improvisation consisting of a theme and variations form [14]. The theme upon which a given qaida performance is built is composed of a series of subphrases, and is taken as a fixed composition. The macroscopic form of qaida follows a fairly simple structure: introduction of the theme, development of variations at an increased tempo, conclusion. Within the main body, variations are presented in a structured manner: a variation is introduced, the theme is reiterated, the same variation is repeated with closed *bayan*, and finally the theme is played again with closed *bayan*, often re-opening it shortly before the end of the cycle.

While qaida themes are part of the shared repertoire of solo tabla, variations are improvised according to some basic principles. The most important guiding principle of qaida variation is a restriction: only *bols* which appear in the qaida theme may be used in the variations. This is intended to preserved the essential character of the given qaida. Given this limitation, one common and effective variation technique is to rearrange subsections of the theme.

4 Methods

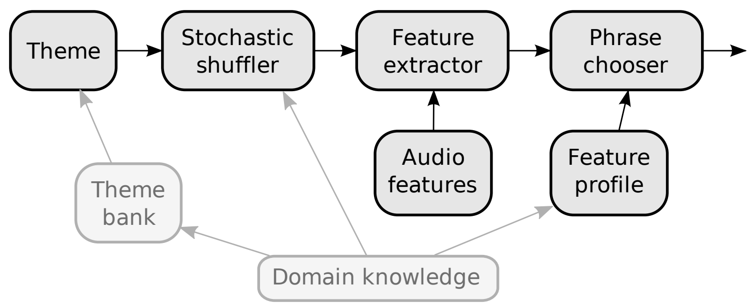

Fig. 2. Overview of the Qaida variation architecture. The theme bank is greyed-out because the choice of theme is made only once, initially. Domain knowledge, specific knowledge about qaida and tabla, is shown being incorporated at specific points in the process.

The design of the system centers around complementary processes of variation generation and variation selection. A database of potential variations is built through a stochastic process, and phrases are selected from that set based on certain criteria. This bears some semblance to the technique known in algorithmic composition as "generate-and-test" [12], however in our case the criteria are treated more probabilistically, as a basis for the system to make a choice with a some indeterminacy but weighted heavily towards a desired outcome.

Consistent with the fact that qaida themes are not themselves improvised, and rarely even composed by the performer, no attempt was made to generate new thematic material. Instead, a number of traditional themes were transcribed manually and annotated with partition bounds. A bank of of these themes is stored in XML format, and one theme is chosen at start-up which remains the only source material for the duration of the qaida improvisation.

The core of the system was coded in Python, relying on the NumPy and SciPy [8] packages for performance intensive computation. Audio output was generated using Pure Data (Pd) [10]. An overview of this system is shown in Figure 2.

4.1 Variation Generation

A bank of phrases is generated from the chosen theme by applying transformations consistent with qaida theory, and then stochastically applying another set of operations to bias the population towards more stylistically appropriate content. An overview of these operations is shown in Figure 3. The size of the phrase database is set in advance, and is far smaller than the set of all possible variations given the transforms. Clearly, a larger database is preferable in that it will contain a greater diversity of material; however, the feature extraction and phrase selection processes described in Section 4.2 scale with the size of the database, and computational efficiency is critical within a real-time architecture. A bank of two thousand phrases was used during much of the development process, and it was qualitatively found that this size contained sufficient phrase diversity to support varied and novel output. A given variation is constructed by applying the transforms, and accepting or rejecting the result based on the constraint that the variation have the same metrical duration as the original. This process is repeated until a bank of the specified size has been constructed.

There are two main transforms used: re-ordering of the theme partitions, and repetition at doubled tempo. The first assembles a variation by sampling with replacement from the set of partitions. For efficiency, the number of possible partitions in the new phrase is limited to the range within which generated phrases of the required length are possible. The second transform simply selects a partition at random and repeats it twice at double the speed. A parameter controls the relative likelihood of applying one or the other of these operations.

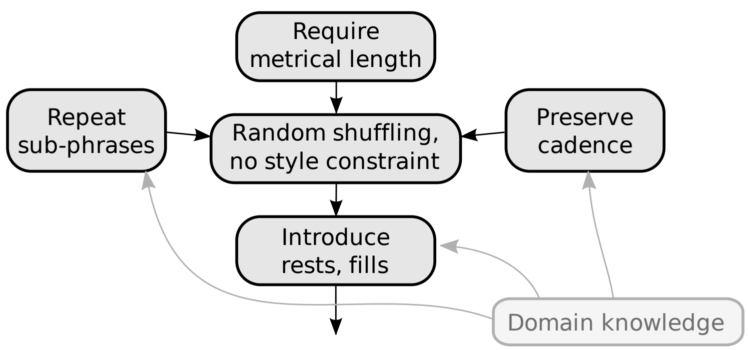

Fig. 3. Detail of the Qaida variation generating architecture. The reordering routine is depicted here, constrained by metrical length and incorporating domain knowledge in the form of added transformations.

An additional set of three transformations may then be applied, each with an independent probability. They function to bias the phrase bank toward more style-specific characteristics. They are intended to favor multiple occurrences of the same partition (non-consecutive repetition), consecutive repetitions of a partition and preservation of the final partition (cadence).

Lastly, a final transform that introduces short rests into the phrases may be applied at any time. This operation is essential to break the homogeneity which tends to emerge over time, but it can also disturb the coherence of a phrase. For this reason it is reserved for use in the more "complicated" sections of qaida development, and may be applied to an existing phrase bank.

4.2 Variation Selection

Selection of a phrase from the bank is initiated by a request for a phrase with a desired set of features. In response, phrases in memory are compared against the request, a close match is selected, and the system returns a single phrase for playback.

Immediately after the phrase bank is first built, features are calculated over each phrase in the set. It was found that a relatively small set of features could provide a surprisingly flexible handle into the character of the returned phrases, though a larger set would no doubt improve the range of performance. The currently calculated features are distribution over each stroke type, by frequency of occurrence and by time, ratio of open to closed strokes, by frequency of occurrence and by time, rhythmic density, spectral centroid, and spectral spread.

Note that these are not all of equivalent dimensionality — rhythmic density, open/closed ratios, and spectral centroid are scalar values, while the distributions over stroke types are vectors. For the most part, these are in effect timbral features, due to the correspondence between stroke types and timbre. The spectral centroid and spread require more explanation. The features themselves are uncomplicated, but up to this point we have been dealing with symbolic data only. However, the sequences are destined for playback on a known set of sounds, so in this step we calculate average values over the same audio database of segmented tabla strokes which is used in playback. This gives us a quantitative estimate of the timbre we expect when a phrase is synthesized.

The feature preferences defined in the request for a variation can describe any subset of the above features, and specify three values for each: the target value, a relative weighting for this feature, and a "flexibility" measure. The target value, expressed in the range 0 to 1, is normalized to the range present in the current bank of variations. The flexibility parameter functions as a sort of distance metric, an alternative to simple linear distance. It defines the width of a Gaussian centered on the target value, which is then used as a look-up table to get the unweighted score for that phrase and feature.

A score is calculated for each phrase in the bank of variations. Rather than always choose the best match, which would lead to a deterministic output, the choice is made probabilistically. The two most successful algorithms are to rescale the scores to emphasize the higher-scoring phrases and choose randomly from the full bank using scores as probability weightings, or to take the set of top scorers and make a choice among those based on their normalized probabilities. This procedure serves as a way to balance the creativity and novelty of the system's output with its responsiveness to the demands of context.

4.3 Macroscopic Structure

The macroscopic structure is simpler and largely deterministic, following the basic qaida form outlined above. Playback is implemented in Pd, and is described further in Section 4.4. The patch controls the alternation between theme and variation, requests variations from the Python generator, controls the periodic opening and closing of the *bayan* strokes, and generates the audio. An accompanying *nagma* marks the cycle. Feature preferences for the variation requests are specified manually with a set of sliders. Modeling of longer-term structure is minimal; the manual controls provided allow a user to take the place of a fuller model. It should be noted, however, that the user need not be highly skilled, or even particularly knowledgeable with respect to tabla or qaida.

4.4 Audio Output

Synthesis of the generated qaida was accomplished using high-quality isolated samples of tabla strokes, played by a professional tabla player and recorded specifically for this project. Several timbrally consistent samples were recorded for each stroke type, one of which was selected at random at each playback command. Amplitudes were scaled by durations, to mimic the lighter touch that is generally used when playing fast sequences. The quality and consistency of the recordings was reflected in the audio output; the only significant shortcoming remains a lack of *bayan* modulation.

5 Evaluation

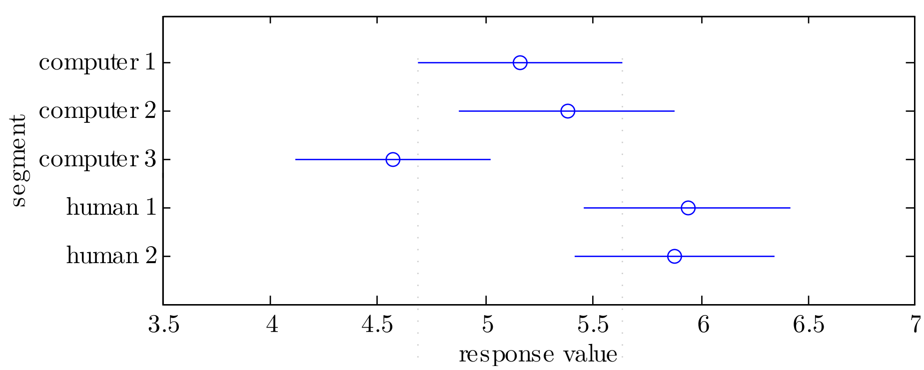

Fig. 4. Plot showing mean values and confidence intervals for responses to Question 1: "To what extent would you say that this recording demonstrates a feeling of musicality?"

An online survey was conducted, in which three recordings of generated output were presented alongside two recordings by a world-class tabla player, without indication of the origin of the recordings; participants were simply asked to make a series of judgements, unaware that the survey involved comparison of human playing and computer modeling. The survey can be found at `http://paragchordia.com/survey/tablasurvey/`, and the audio clips of both computer-generated output and professional tabla performance can be heard separately at `http://www.alexrae.net/thesis/sound/`, the first three being the qaida model's output, as in the results presented here. The recordings of model output were "played" via the user interface implemented in Pd, and were recorded without subsequent editing.

A total of 70 participants responded to the survey. A majority claimed moderate to high familiarity with tabla music, and many reported themselves to be practicing tabla players. The mean age was 35.2, with a standard deviation of 12.2. The order of presentation of audio segments was randomized, and participants were asked to rate the examples along several dimensions. Judgements were on a scale of 1 to 7, reflecting answers ranging from "very little" to "a lot", except in case of the last two questions, phrased as ranging from "poor" to "excellent". A higher value corresponded to a more favorable judgment. Respondents were invited to supplement their quantitative judgements with further comments.

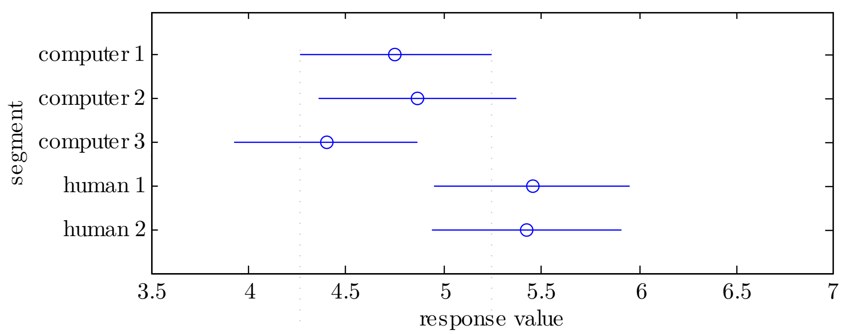

Fig. 5. Plot showing mean values and confidence intervals for responses to Question 2: "To what extent would you say that this recording demonstrates musical creativity?"

Participants were asked the following questions:

1. To what extent would you say that this recording demonstrates a feeling of musicality?
2. To what extent would you say that this recording demonstrates musical creativity?
3. To what extent would you say that this recording adheres to qaida form?
4. To what extent would you say that this recording is novel or surprising, given the qaida theme?
5. To what extent would you say that the improvisations in this recording are appropriate to the style and the theme?

6. If told that this recording were of a tabla student, how would you rate his/her overall TECHNICAL abilities?
7. If told that this recording were of a tabla student, how would you rate his/her overall MUSICAL abilities?

Figures 4–6 show mean values and confidence intervals of the judgement scores for each audio segment, adjusted for multiple comparisons using the Dunn-Sidak correction ($p < 0.05$). A trend is visible in the average values of the data across the examples, showing the computer generated output to be rated slightly lower than the human generated excerpts. However, the differences do not reach statistical significance given the sample size, except in the case of the third generated qaida, which in many cases is rated somewhat lower than the other model outputs. Judgements of musical creativity, question 2, are notable, as two of the qaida model's outputs were ranked on par with the human performer. The model was rated similarly highly on judgements of novelty. These results are encouraging: the computer-generated qaida performed quite well in comparison to very high-quality human-played examples.

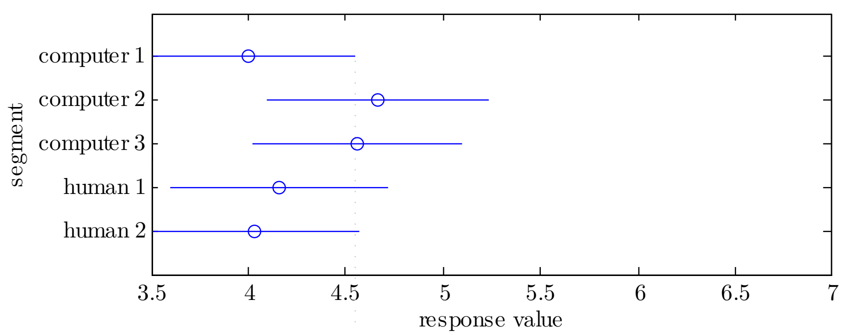

Fig. 6. Plot showing mean values and confidence intervals for responses to Question 4: "To what extent would you say that this recording is novel or surprising, given the qaida theme?"

It is also interesting to note from the comments that many respondents remained unaware that three of the examples were computer-generated. One, for example, wrote in response to example 3: "Again this recording demonstrates that the Tabla player has excellent abilities in playing the right drum with crisp tonal quality. Left drum (Baya) needs some improvement as I stated in the first two Qaidas." Some comments focused more directly on the style or quality, for example "Good presentation of Purab / Benaras style kayda. Great speed. Nice overall sound" (excerpt 2), and "Very nicely done" (excerpt 3). Only one respondent clearly deduced the origin of the model's output, writing simply "The synthesized nature of this piece limits its ability to be musical." Criticism was not reserved for the generated recordings. One respondent commented that excerpt 4 "sounded too mechanical and devoid of emotion," and another that "The Tirakitas at the start [of example 5] sound very odd and clumsy!" Most comments for examples 4 and 5, however, were clearly positive.

6 Conclusion

The results of our survey suggest that the qaida model has been successful in producing improvisatory music which is heard as creative. There is, of course, much work to be done, ranging from addressing deficiencies in playback cited by a number of respondents, such as the lack of *bayan* modulation, to incorporating a more robust model of sculpting a larger contour. However it is encouraging and quite interesting to see how effective the methods employed in this model have been.

References

1. David Cope. *Experiments in Musical Intelligence.* A-R Editions, Madison, WI, 1996.
2. David Cope. *Virtual Music: Computer Synthesis of Musical Style.* MIT Press, Cambridge MA, 2001.
3. David Cope. *Computer Models of Musical Creativity.* MIT Press, Cambridge MA, 2005.
4. Mihaly Csikszentmihalyi. *Creativity: Flow and the Psychology of Discovery and Invention.* Harper Collins, New York, 1996.
5. Arne Eigenfeldt. The creation of evolutionary rhythms within a multi-agent networked drum ensemble. In *Proceedings of the International Computer Music Conference*, pages 267–270, Copenhagen, Denmark, 2007.
6. Ronald A. Fink, Thomas B. Ward, and Steven M. Smith. *Creative Cognition: Theory, Research, and Applications.* MIT Press, Cambridge, MA, 1992.
7. Howard Gardner. *Intelligence reframed: multiple intelligences for the 21st century.* Basic Books, New York, 1999.
8. Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–present.
9. François Pachet. The Continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–41, 2003.
10. Pure Data — PD community site. `http://puredata.info`, (accessed March 2009).
11. Francisco Cmara Pereira. *Creativity and artificial intelligence: a conceptual blending approach.* Walter de Gruyter, 2007.
12. Curtis Roads. *The Computer Music Tutorial.* MIT Press, Cambridge, MA, 1998.
13. Robert J. Sternberg and Janet E. Davidson. The mind of the puzzler. *Psychology Today*, 16:37–44, October 1982.
14. Gert-Matthias Wegner. *Vintage Tabla Repertory.* Munshiram Manohalal Publishers Pvt. Ltd., New Delhi, India, 2004.