A passive approach to evaluating Mridangam transcription via perceptual experiment

Kaustuv Kanti Ganguli[1†], Akshay Anantapadmanabhan[2], Gautham Lali[1], and Carlos Guedes[1]

[1]*Music and Sound Cultures research laboratory, New York University Abu Dhabi, UAE*

[2]*Freelance Musician, India*

[†]kaustuvkanti@nyu.edu

Background

Analysis and characterization of percussive instruments using computational methods have recently got attention both in the context of Western and non-Western repertoire, specifically on automatic transcription of tabla and mridangam strokes (Anantapadmanabhan *et al.*, 2014; Guedes *et al.*, 2018). The transcription approach, however, is limited by its dependency on prior knowledge about the specific modes of the instrument. Another concern is the absence of a unique mapping between the strokes and their nomenclature in the vocabulary. The same stroke is often uttered differently in the konakkol vocalization based on contextual variations or grammatical impositions. Most notably, even an expert musician is often unable to resolve such ambiguities on isolated presentation of a stroke but uses contextual cues to possibly interpolate within and across phrases. Our previous work (Ganguli *et al.*, 2020; 2021) addressed this problem by proposing a combination of acoustic and semantic approaches for the contextual transcription of mridangam strokes. Grammatically-accurate labels for the transcribed strokes were obtained through the learned language model as knowledge-constraints.

Aims

An atypical behavioral experiment paradigm — listen and imitate — based on acoustic production, rather than listening, has shown to successfully test for Categorical Perception in speech intonation. The participant's task is to imitate a presented stimulus prompt as closely as possible. The goal is to test whether participants memorize the stimulus and tend to recall it from their working memory, or whether they perceive it as the closest prototype template already stored in their long-term memory and thus reproduce the prototype. Given the oral and aural nature of the pedagogy, we chose to present the stimuli in the form of konakkol instead of symbolic text, eventually to acquire mridangam recordings against the vocal konakkol stimuli.

Method

With a view to understanding whether the subjects closely imitate the stimulus prompt, the behavioral experiment is conducted by a domain expert tutoring the participants, forming two groups consisting of 4 individuals each with no age restrictions. The discriminating factor between the groups is 'experience': 4+ and <4 years of training respectively. The task is to play back the (in)congruent stimulus prompt on Mridangam, with(out) metronome clicks, after listening to the prompt as many times. The response is recorded in quiet environment, thrice in quick succession and repeated after two days. The judgment criteria are: (i) how close the content is to the original segments, (ii) were there any trends of systematic variation from the prompt, (iii) were there separate trends between the two control groups, (iv) what is the replacement by interpolation for incongruent or missing samples, (v) whether metronome has any effect on the playback if keeping to time is an additional constraint, and finally (vi) whether there is a (short-term) learning phenomenon that reflects on the renditions recorded on the later occasion.

Results

Upon initial experiments, it is observed that experience, vis-à-vis, a proxy for years of training, has an effect in interpolating gaps in the stimulus prompt by 'filling in'. This is allied to a language understanding when known idioms are interpreted even in ambiguous semantic fields. The other interesting aspect is the muscle memory or reflex, per se, which conforms to interpretation of standard rhythmic phrases as gestalts (Ganguli and Guedes, 2019), the manner of articulation showing artefacts.

Conclusions

We confirmed that while beginners focused on the actual 'uttered' stroke in the konakkol, the experts used their long-term memory to interpolate omitted/substituted strokes with respect to the 'commonly taught' phrases and use this semantic knowledge for the imitation task. This 'learned schema' can be conceived as a proxy for the 'experience' of a musician acquired through the years of training. Given the robustness of the experimental paradigm in speech and speech-like signals e.g. raga melodic motif perceptions (Ganguli and Rao, 2019), we consciously chose this task to evaluate the performance of our transcription algorithm.

References

Guedes, C., Trochidis, K., & Anantapadmanabhan, A. (2018). Modeling Carnatic Rhythm Generation: A Data-Driven approach based on Rhythmic Analysis. In Proceedings of the 15th Sound & Music Computing Conference, Limassol, Cyprus.

Anantapadmanabhan, A., Bello, J., Krishnan, R., & Murthy, H. (2014). Tonic-independent stroke transcription of the mridangam. In Audio Engineering Society Conference: 53rd International Conference: Semantic Audio. Audio Engineering Society.

Ganguli, K., & Guedes, C. (2019). An approach to adding knowledge constraints to a data-driven generative model for Carnatic rhythm sequence. *Trends in Electrical Engineering*, 9(3), 11-17.

Ganguli, K. K., Anantapadmanabhan, A., & Guedes, C. (2020). Questioning the Fundamental Problem-Definition of Mridangam Transcription. In Proceedings of Timbre Conference, Thessaloniki, Greece.

Ganguli, K. K., Anantapadmanabhan, A., & Guedes, C. (2021). An approach to adding knowledge constraints by fractal analysis on a generative model of Carnatic rhythm sequence. In Proceedings of Analytical Approaches to World Music (AAWM) Conference, Paris, France.

Ganguli, K. K., & Rao, P. (2019). On the perception of raga motifs by trained musicians. *The Journal of the Acoustical Society of America*, 145(4), 2418-2434.

Keywords: Carnatic rhythm, Mridangam transcription, Improvisation, Listen and imitate, Learned schema.

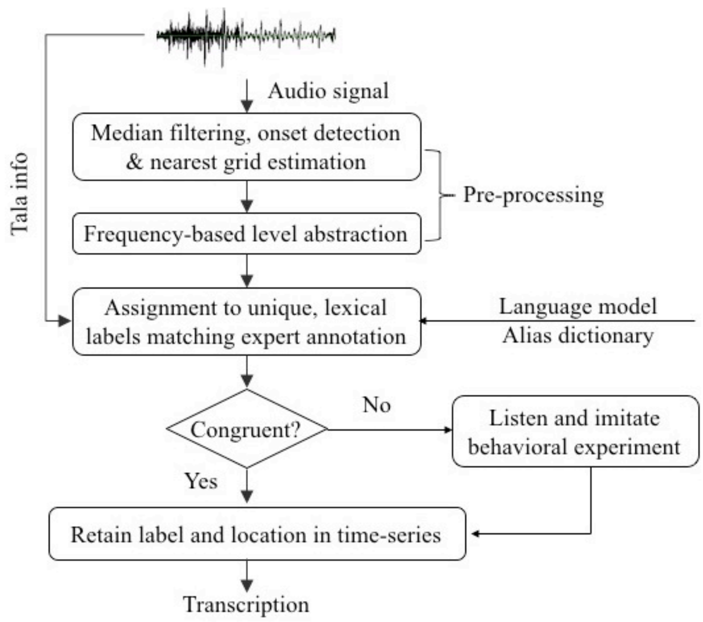

Figure 1. Schematic block diagram of the evaluation strategies for Mridangam transcription.