Characterizing Subtle Timbre Effects of Drum Strokes Played with Different Technique

Francesco Bigoni[1], Michael Grossbach[2], and Sofia Dahl[1†]
[1] Department of Architecture, Design and Media Technology, Aalborg University, Copenhagen, Denmark
[2] Institute of Music Physiology and Musicians' Medicine, Hanover University of Music, Drama and Media, Hannover, Germany
[†] Corresponding author: sof@create.aau.dk

Introduction

In instrumental playing, musicians control sound characteristics such as loudness and timbre. While the many years of training for professional musicians ensure that most changes in timbre are intentional, others arise from changes in playing position, technique applied, or timing (Danielsen et. al 2015). Although subtle, such changes in timbre may still be heard and add to the quality of the performance, also for brief, percussive sounds. Drum strokes where the drumstick is allowed to freely rebound from the drum head – "normal" strokes – appear to have different audible quality compared to "controlled" strokes, where a player restrains the stick from freely moving up after the hit (Dahl & Altenmüller, 2008). Although audible, these differences in timbre are not always captured by features traditionally used such as log attack time and temporal centroid. An additional problem is the brevity of percussive sounds, making the use of those descriptors that require frame-based processing (e.g. spectral flux, spectral contrast) difficult (Bigoni & Dahl, 2018). In this context, it is crucial to identify the signal phases (e.g. attack vs. decay or transient vs. steady-state) in a way that is perceptually relevant and meaningful for drum sounds. The goal of this study is to evaluate whether audio descriptors that capture characteristics of the transient part of the waveform differ between Normal and Controlled strokes. Similar to Danielsen et al. (2015), we define this transient part to occur between the onset and the temporal centroid of each stroke, but we also separate the initial attack.

Method

The data consisted of a set of 1102 drum strokes played on a 14-inch rototom with instructions that determined the grip of the drumstick for each stroke, as described by Dahl & Altenmüller (2008). Eight professional players were instructed to play *Normal* (N) strokes, where the stick was allowed to freely rebound, whereas for *Controlled* (C) strokes, the player was instructed to control the ending position of the drumstick, stopping it as close as possible to the drum head after a stroke. A trial typically consisted of 10-13 strokes, where each stroke was allowed to ring out before the next one. The striking area of the drumhead was defined by a circle, 5 cm in diameter. An omnidirectional condenser microphone was mounted at a distance of 50 cm and angled 45 degrees with respect to the drumhead surface.

We separated each drum stroke into separate files and extracted audio descriptors using MIRtoolbox 1.7.2 (Lartillot, Toiviainen & Eerola, 2008). In order to separate the early energy part of the waveform, which we believe to be perceptually relevant, from the later tonal and "ringing" phase of each waveform, we defined algorithms for detection of onsets and offsets of each stroke. Rather than relying on amplitude envelopes for detection of onset and offset time (c.f. Nymoen, Danielsen & London, 2017) we used a threshold peak-picking algorithm directly on the waveform. To avoid spurious peaks in the waveform, the maximum peak time was estimated from a smoothed signal envelope instead. These time events, together with the temporal centroid (see Figure 1), were used to define four signal phases: 1) attack (max peak time - onset time), 2) early decay (temporal centroid - max peak time), 3) late decay (offset time - temporal centroid), 4) total (offset time - onset time). In all, 25 audio descriptors were extracted using the MIRtoolbox algorithms: duration (x 4 phases), sound pressure level (SPL, x 4 phases), spectral centroid (x 4 phases), temporal centroid, temporal flatness (i.e. geometric mean / arithmetic mean, calculated on the amplitude envelope values, x 4 phases), spectral flatness (i.e. geometric mean / arithmetic mean, calculated on the spectral bin values, x 4 phases), and crest factor (i.e. max peak value / rms value, calculated on the waveform, x 4 phases).

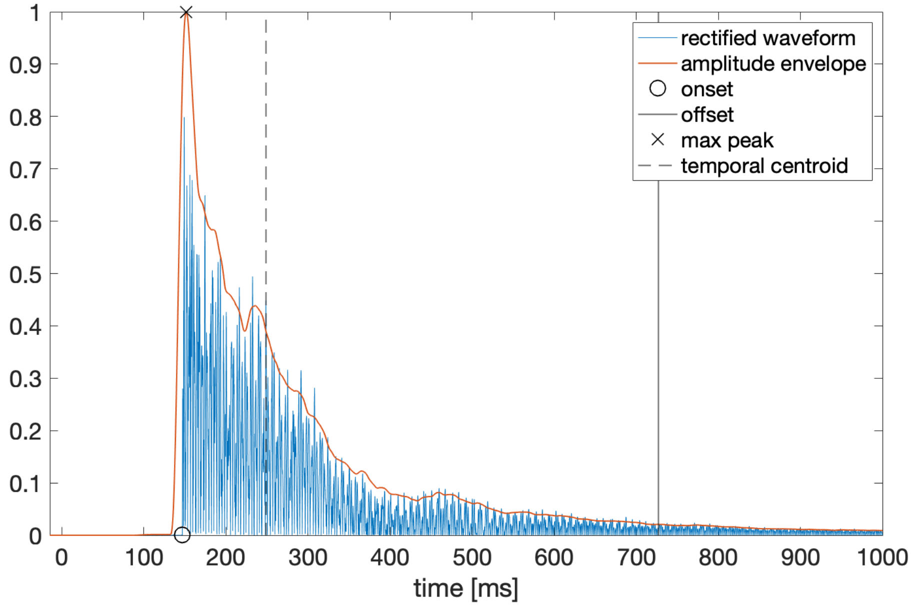

Figure 1: Rectified waveform of a stroke with the smoothed envelope. The markings show extracted onset (circle); maximum peak (cross); temporal centroid (dashed vertical line); and offset (full vertical line). For the modelling, we used descriptors based on the early decay phase (between max peak and temporal centroid).

Results

As expected, our descriptors differ between players in terms of range, and some are highly correlated. Judging from the density plots of the descriptors, we could not discern substantial differences between N and C strokes: that indicates that differences between players and playing arm influenced our data to a high degree.

Departing from our initial visual and auditory exploration of the data, as well as the reported results from Danielsen et al (2015), we started by modelling the spectral centroid of the early decay phase (earlyDecSC) in a multilevel Bayesian regression with a skewed normal link function and informative priors, using the package brms (Bürkner, 2017) in R (R Core Team, 2020). The best model according to a leave-one-out cross-validation (Vehtari et al., 2017) showed a clear difference in the estimated mean of earlyDecSC conditional on instruction (N or C), with the expected value for the dependent variable for N-strokes estimated to be 47.65 (Credible Interval [CI]: 58.31to 37.59) lower compared to that of C-strokes (Mean [CI]: 773.45 [759.67 to 788.26]). There was no interaction with the playing arm. Furthermore, the model suggests that the playing instruction had an effect on skewness as well as on spread, both increasing for C-strokes. Playing with the non-dominant arm also increased skew and spread, but to a lesser degree.

Discussion

Based on the model, we can infer that the playing instructions cause differences in the mean, the skewness, and spread of spectral centroid calculated across the early decay phase (between max peak and temporal centroid). The results show a substantially lower mean spectral centroid for normal strokes, with no effect from the playing arm. Since spectral centroid is commonly considered as a perceptual correlate of brightness, our model agrees with our informal listening explorations, where we found controlled strokes to sound slightly harsher/brighter around the hit point and getting faster to the tonal, ringing phase compared to normal strokes. The listening test in Dahl & Altenmüller (2008), using a subset of the strokes from one of the players, reported the controlled strokes being rated as more flat or dull compared to normal strokes despite having higher peak force and shorter contact durations. Intuitively, one would think that stopping the drum stick as soon as possible after the impact would prolong the contact time, but the opposite was found for this player. Rather, a firm, "cramped" grip around the drum stick could alter the vibrational modes that influence the force pulse during contact with the drum head, thus altering the spectral centroid and perceived brightness.

The increase in skewness and spread for controlled strokes seem reasonable in that both inter- and intra-player variability would be more likely to increase in their execution of these strokes, but also that some players may have enforced the stopping of the drumstick to a higher degree than others. A rather high intra-class correlation coefficient for a null model, including only the subjects as grouping structure, indicates that more than 70 percent of the variation can be explained by the individual participants, most likely an effect of the small sample size in combination with high inter-individual variation.

Although collected from a small sample, our results provide further insights to how subtle timbre changes can be described for brief, percussive sounds. We will proceed to include temporal flatness and crest factor into a multivariate model over the early decay phase for normal and controlled strokes. A later listening tests will aim to verify that the investigated descriptors are useful as perceptual predictors. Although subtle, we argue that the changes are perceivable and that investigating suitable descriptors provide important knowledge on the link between action and perception in the control of musical instrument playing. For example, percussionists often train to produce even loudness and timbre across many repeated strokes, a skill very much needed in performance of pieces such as Ravel's *Bolero*.

Acknowledgments

The authors would like to thank Olivier Lartillot for sharing the updated version of *mirevents* before the release of MIRtoolbox 2.0. The original data was collected during a MOBILITY-2.1 Marie Curie Intra-European Fellowship (EIF) awarded to Dahl. The contribution by Dahl is partially funded by NordForsk's Nordic University Hub Nordic Sound and Music Computing Network (NordicSMC), project number 86892.

Author SD conceived and designed the study; SD & MG collected the data; FB preprocessed the data, performed feature extraction and initial analysis; MG made the statistical modelling; all authors participated in the interpretation of results and writing of the final manuscript.

References

Bigoni, F., & Dahl, S. (2018). Timbre Discrimination for Brief Instrument Sounds. In E. Gómez, X. Hu, E. Humphrey, & E. Benetos (eds.), *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018),* (pp. 128–134). Paris, France.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan, *Journal of Statistical Software,* Foundation for Open Access Statistic, 80.

Dahl, S., & Altenmüller, E. (2008). Motor control in drumming: Influence of movement pattern on contact force and sound characteristics. In *Proceedings of Acoustics'08, Intl. Meeting Acoustical Society of America, ASA, the European Acoustics Association, EAA, and the Socité Francaise d'Acoustique, SFA.*, (pp. 1489–1494). Miami, Florida.

Danielsen, A., Waadeland, C. H., Sundt, H. G., & Witek, M. A. (2015). Effects of instructed timing and tempo on snare drum sound in drum kit performance. *The Journal of the Acoustical Society of America*, *138*(4), 2301–2316.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, B., Decker, R. (eds), *Data analysis, machine learning and applications* (pp. 261-268). Springer, Berlin, Heidelberg.

Nymoen, K., Danielsen, A., & London, J. (2017). Validating attack phase descriptors obtained by the Timbre Toolbox and MIRtoolbox. In: Tapio Lokki, Jukka Pätynen, and Vesa Välimäki (eds), *Proceedings of the SMC Conferences* (pp. 214-219). Aalto University, Finland.

R Core Team (2020). R: A Language and Environment for Statistical Computing.

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, *Statistics and Computing*, *27*, 1413–1432.