SIMULATING MICROPHONE BLEED AND TOM-TOM RESONANCE IN MULTISAMPLED DRUM WORKSTATIONS

Alice Clifford, *

Centre for Digital Music
Queen Mary University of London
London, UK
alice.clifford@eecs.qmul.ac.uk

Henry Lindsay Smith

FXpansion Audio UK
London, UK
henry@fxpansion.com

Joshua Reiss

Centre for Digital Music
Queen Mary University of London
London, UK
josh.reiss@eecs.qmul.ac.uk

ABSTRACT

In recent years multisampled drum workstations have become increasingly popular. They offer an alternative to recording a full drum kit if a producer, engineer or amateur lacks the equipment, money, space or knowledge to produce a quality recording. These drum workstations strive for realism, often recording up to a hundred different velocity hits of the same drum, including recordings from all microphones for each drum hit and including bleed between these microphones. This paper describes research undertaken to investigate if it is possible to simulate the snare and kick drum bleed into the tom-tom microphones and the subsequent resonance of the tom-tom that is caused, with the aim of reducing the amount of audio data that needs to be stored. A listening test was performed asking participants to identify the real recording from a simulation. The results were not statistically significant to reject the hypothesis that subjects were unable to distinguish the difference between the real and simulated recordings. This suggests listeners were unable to identify the real recording in the majority of cases.

1. INTRODUCTION

Recording a full drum kit comes with many challenges, from simply finding a space big enough to adequately record a drum kit to dealing with issues that occur with the large amount of separate instruments in close proximity. There are products on the market, known as multisampled drum workstations (MDWs), that allow amateur and professional engineers the opportunity to recreate the sound of a full kit recorded in a professional studio simply from a laptop, for example FXpansion's BFD[1].

The premise of an MDW is to go one step further than a simple sampler or synthesiser. A drum kit is laid out in a studio with a standard microphone setup and each drum is recorded in isolation and struck at many different velocities and positions. An interface is then developed to access these samples and allow the user to program their own drum beats and render all the individual recordings together to create a studio quality emulation of a real drummer. Ideally every microphone would be recorded for every drum hit to reproduce the bleed between microphones. For example when a user listens to a particular drum of interest the sound of the other drums in the kit will be heard due to bleed into the microphone used to record the drum of interest. The problem with this method is a large amount of data needs to be stored.

It would be advantageous to be able to produce microphone bleed without having to provide the actual data. It may be possible to synthesise this missing data but this is at odds with the philosophy of creating an MDW from recorded samples. This paper investigates whether it is possible to simulate the bleed of a kick or snare drum into the tom-tom drum microphones and the audible effects this has on the tom-tom drum itself using the data already available as standard and evaluates how effective these simulations are compared to real data through listening tests.

2. BACKGROUND

Generally while recording a drum kit each drum has a dedicated microphone to reproduce the direct sound of a single drum. Bleed is considered the sound from a different drum arriving in this microphone and is inevitable in a multiple instrument, multiple microphone setup.

The bleed in a tom-tom microphone is primarily from two sources; the direct sound of the kick or snare drum arriving at the microphone and the tom-tom resonating due to this direct sound. For the case of the snare drum bleed this can be described as

$$x_t[n] = h_s * s[n] + h_t * \hat{t}[n] + w[n] \tag{1}$$

where x_t is the tom-tom microphone signal, s is the sound of the snare drum being struck, $\hat{t}$ is the tom-tom resonance excited by the snare drum, w is uncorrelated noise and h_s and h_t are room impulse responses between the snare drum and the microphone and the tom-tom resonance at the microphone.

Drums can be generalised as a circular membrane stretched over an air space [1]. When the membrane, or drum skin, is struck this causes it to vibrate at different modes. This also causes the air within the drum to resonate as well as the drum body itself, producing a characteristic sound. Drums can also resonate due to excitation from vibrations in the air due to other drums in the kit being struck, known as sympathetic resonance. Tom-tom drums are

* This work was funded by the EPSRC as part of the ImpactQM project EP/H500162/1.

[1] http://www.fxpansion.com

tuned to resonate at different, complementary frequencies. They are also notorious for resonating or "ringing" when other drums are played and may be tuned up or down to change the resonant frequency to avoid this. Although the ringing can be avoided it is an integral part of a real drum kit. In addition to this there are many different factors which will determine how the resonance of a tom-tom will sound in the microphone, including microphone type, the positions of the microphones, tom-toms, other drums, listening position, room characteristics and mechanical connections to other instruments.

It is unlikely that the details of all these factors are noted during a recording session and MDWs also allow users to place drums in almost any configuration. Assumptions therefore need to be made and the same algorithm needs to be able to simulate drums in a variety of configurations with a general approach.

3. SIMULATION

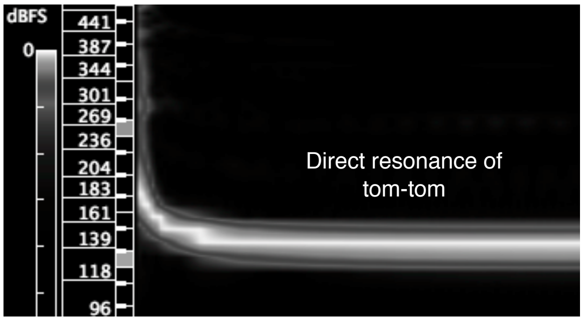
The direct kick or snare drum in the tom-tom microphone can be simulated from the direct recording of each instrument. The recording has to be processed to simulate the effects of sound travelling a distance through air [2]. It is unlikely the bleed will be heard in isolation therefore a simple simulation will suffice. A high shelving filter taken from [3] was used to simulate air absorption on the direct recordings. Equations are well established for modelling air absorption dependent on distance [4] but it is assumed the relative distances between drums are unknown. The gain of the filter was then taken from analysis of previously recorded data, leading to a filter specification of -8dB gain at a 5kHz cutoff. In addition to this the source instrument was attenuated by an amount that would not cause noticeable positive reinforcement when the bleed signals were mixed together.

3.1. Extracting tom-tom resonance

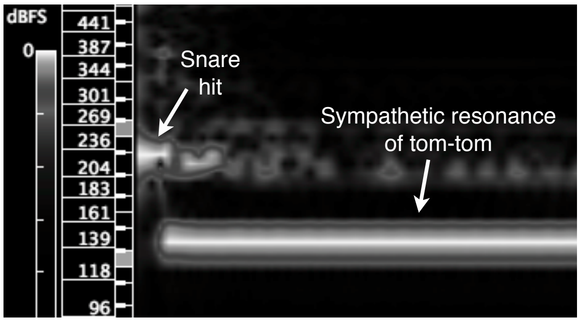
To simulate the resonance of the tom-tom due to the bleed instrument, the resonance is extracted from the direct recording of the tom-tom drum as this also excites the resonant modes. The modes of an ideal circular membrane can be predicted [1], although real tom-toms appear to diverge from the ideal case. It is known that the modes of a tom-tom will rise if struck with a large force. Figure 1a shows a spectrogram of a tom-tom hit recorded at the tom-tom microphone, showing the fundamental mode of 138Hz. At the beginning of the hit the mode is at a higher frequency. Figure 1b shows a spectrogram of a snare hit in the tom-tom microphone. The resonance of the fundamental mode of the tom-tom can clearly be seen at the same frequency but it is delayed due to the delay of the sound of the snare arriving at the tom-tom.

It is therefore not appropriate to simply use the direct tom-tom to reproduce the tom-tom resonance due to the fall in frequency. The resonance can be extracted by measuring the spectral flux of the tom-tom signal [5]. Spectral flux is a measure of the change of spectral content over time and can be used for transient and steady state detection [3]. It is calculated by taking the Euclidean distance of the magnitude of subsequent frames of data. This is described by

$$S[n] = \sqrt{\sum_{k=0}^{N-1} [X[n,k] - X[n-1,k]]^2} \tag{2}$$



(a) Normalised spectrogram of direct tom-tom microphone while tom-tom is struck showing frequency over time



(b) Normalised spectrogram of direct tom-tom microphone while snare is struck showing frequency over time

Figure 1: Spectrograms taken from Sonic Visualiser [2]

where X is the microphone signal x in the frequency domain, k is the bin number from $0, \ldots, N-1$, N is the window size and n is the current time step. Once the fundamental mode of the tom-tom converges to a single value the spectral flux will also converge. Figure 2 shows the first derivative of the spectral flux of a direct tom-tom signal, S'. The initial attack and decay can clearly be seen. The point at which the resonance begins can be extracted by finding the point where the first derivative of the spectral flux crosses a threshold after the minimum, in this case when $S' > -10$. The position for this tom-tom is indicated by a dashed vertical line.

3.2. Snare drum

3.2.1. Resonance filter

For an object to sympathetically resonate, the resonant frequencies have to be excited. The consequence of this is that for a snare drum to resonate significant modes of a tom-tom it has to produce those frequencies [6]. After listening to and analysing real tom-tom bleed recordings it became apparent that for low tom-toms, the fundamental frequencies are not excited by the snare drum hit but are excited when the tom-tom is hit directly. Therefore simply using the resonance of the tom-tom from a direct hit will not be accurate for the simulation as it will contain frequencies not present on a real recording.

To mitigate this the extracted resonance is processed with a high pass filter with a cut off point taken from the peak frequency of the direct recording of a snare hit. In this implementation a 4th order Butterworth filter was used. The result of this is a more convincing low frequency tom-tom simulation where the fundamental frequencies are attenuated but the higher modes and any rattle of

[2] `http://www.sonicvisualiser.org/`

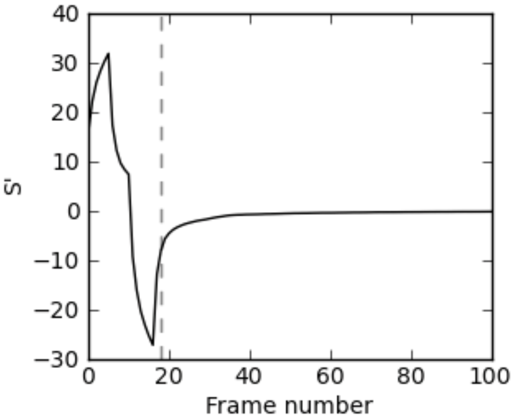


Figure 2: The first derivative of spectral flux S plotted against time. The beginning of the resonance as a dashed vertical line.

the tom-tom is retained.

3.2.2. Gain

Analysis of the real data shows that the peak amplitude of the snare has a linear relationship to the peak amplitude of the tom-tom bleed resonance. As mentioned previously, the position of the drums is unknown therefore the gain cannot be directly estimated. The extracted resonance is scaled by a gain factor that is proportional to half the difference in peak frequency of the snare drum and peak frequency of the extracted resonance. This means that a large difference in peak frequency will result in a large gain factor and more attenuation, also reducing the low frequency mode level.

3.3. Kick drum

The kick drum produces much lower frequencies than the snare drum and will resonate lower frequencies of the tom-tom therefore filtering of the extracted resonance is not required. The extracted resonance is scaled by a single value for all toms in comparison to the peak amplitude of the direct kick drum.

4. EVALUATION

The effectiveness of the simulations was established through a subjective listening test of the 4 full kit recordings available with tom-tom bleed recordings. For this test a single velocity layer of the kick and snare drums was used, resulting in 32 audio samples available to analyse and simulate.

4.1. Subjective analysis

4.1.1. Description

The listening test was designed to ascertain whether a participant was able to distinguish the real recording from the simulation. The null hypothesis is that participants are unable to discern between real and simulated recordings.

A pairwise comparison listening test was designed and implemented online[3]. The anonymous data from participants can be found at[4].The test was conducted online to reach a wider audience and to attract more participants. The url was only distributed to those considered experts in the field of audio who had experience of critical listening which resulted in 35 participants. The users were asked to indicate their experience in audio (audio engineer, software developer, student etc) and to rate their specific experience at listening to drum recordings on a scale of 1 to 10.

As a control test, the participant was firstly presented with two sounds; one direct snare signal and a snare signal mixed with the real tom-tom microphone with snare bleed and were asked to indicate which sound contained bleed. If the participant was unable to hear the bleed they were not included in the analysis. The majority of participants were able to detect the bleed. The participant was then presented with a training page to familiarise themselves with the sounds.

The participants were presented with a simple interface with 2 buttons labelled 'Sound A' and 'Sound B' and were given four options to choose from:

1. Sound A is a real recording
2. Sound B is a real recording
3. The sounds are different but either sound could be the real recording
4. The sounds are the same.

Option 3 was included after pilot tests suggested it was common for a participant to identify the sounds were different but that both sounded like a real recording. Option 4 was included to establish if any simulations were good enough to be considered the same sound. 10 additional pairs were included where sound A and sound B were the same sound files, randomly chosen from the dataset, as a control to ensure the participant could establish when the sounds were the same or different. The user was also given the opportunity to add any other comments about each pair. The order of pairs was randomised and therefore the test was double-blind.

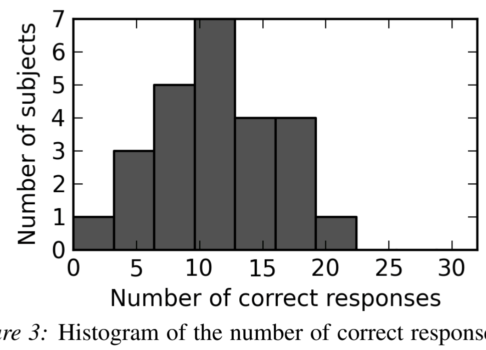


Figure 3: Histogram of the number of correct responses per subject.

4.1.2. Results

The results were analysed assuming a Binomial distribution as an adaptation of traditional ABX listening tests [7]. 25 of the partic-

[3]http://webprojects.eecs.qmul.ac.uk/alicec/fx_listeningTest/pairwise_ver1

[4]http://webprojects.eecs.qmul.ac.uk/alicec/fx_listeningTest/pairwise_ver1/anon_userData

Response	*PCC*	*p-value*
Correct	-0.387	0.029
Incorrect	-0.046	0.804
Same	0.380	0.032
No preference	0.054	0.771

Table 1: Pearson's Correlation Coefficient of each response against SDR for each pair of sounds.

ipants correctly identified 7 out of the 10 identical pairs and were used for the analysis.

Processing of the responses results in four possible outcomes for each pair trail: correct identification of the real recording, incorrect identification of the simulation as the real recording, incorrect identification that the sounds are the same or identifying the sounds are different but no preference which is real. To reject the hypothesis that participants are unable to distinguish between real and simulation recordings the users would have to correctly identify the real recordings with a high confidence.

For the 32 different pairs, the number of correct responses for each user is shown as a histogram in Figure 3. The mean number of correct response is 11.1, a probability of 0.35 of total responses, with a sample standard deviation 4.7

Taking the probability of correcting identifying the real recording as 0.25 by chance, 9 subjects, or 37.5%, correctly identified the real recording with a confidence interval of $p <= 0.05$. As the users have been filtered by those that could identify the equal pairs, it can be assumed that the participant is highly unlikely to incorrectly identify the sounds are the same. If the probability of a user selecting the correct answer is now 0.33, 5 subjects, or 21%, correctly identified the real recording with a confidence interval of $p <= 0.05$.

The results therefore fail to reject the hypothesis that users are unable to identify the real recording from the simulation as only 5 participants out of 32 are able to correctly identify the real recordings with a statistical significance higher than 95%. This leads to the conclusion that the simulation is convincing in the majority of cases.

Figure 4 shows the number of correct responses against the signal-to-distortion (SDR) ratio between the real and simulated signal. The SDR was calculated using a modified version of performance measurements used in blind source separation [8] and gives an indication of the difference between two signals. Table 1 shows the Pearsons Correlation Coefficient (PCC) and p-value for each pair. This shows there is a negative correlation between SDR and the number of correct responses and a positive correlation between the number of responses that the sounds are the same and SDR. This is as expected as it suggests that pairs that are very different i.e. the simulation sounds different to the real recording, are more likely to be correctly identified. Equally, if the SDR is high and the pair sounds similar, they are likely to incorrectly respond that the sounds are the same. There is little correlation of the other responses. Although this suggests the participants were able to hear the difference, it is a fairly weak negative or positive correlation at around $\pm$ 0.4.

The analysis was also run with filtering out participants that rated their experience as 6 out of 10 or higher. There was no significant difference between the results, which suggests the results are representative of audio experts with experience in drums and audio experts without.

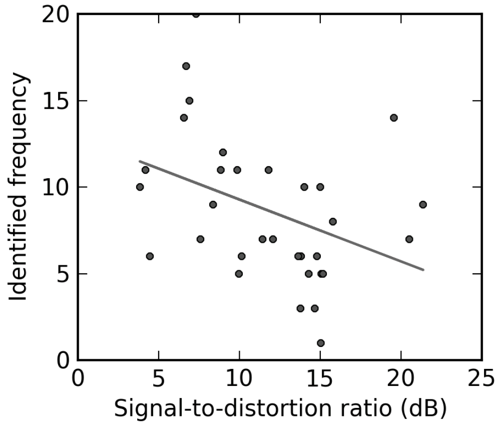


Figure 4: Peak amplitude of extracted tom-tom resonance against velocity

5. CONCLUSION

A method for simulating snare and kick drum bleed into tom-tom microphones from existing data has been described. The bleed instrument part of the bleed signal is simulated by attenuating and filtering the direct bleed instrument recording to simulate air absorption. The sympathetic resonance of the tom-tom by the bleed instrument is simulated by extracting the resonance from the direct tom-tom recording and applying a filter dependent on the frequency of the bleeding drum. Examples of the final algorithm can be heard at [5].

The simulation was subjectively tested using a pairwise comparison listening test and analysed using variations on analysis for ABX listening tests. Subjects were presented with pairs of sound, one of which was the real recording and one which was the simulation. The subjects were asked to indicate which sound was real or if the sounds were the same. The results were not statistically significant to reject the hypothesis that subjects were unable to distinguish the difference between the real and simulation. This suggests listeners were unable to identify the real recording in the majority of cases.

The simulation can be extended by simulating some of the finer details, such as rattle between tom-toms and the effects of groups of instruments on the resonance. A machine learning approach could be taken by processing recorded data to extract features that may be different between the direct recorded data and the bleed data.

The listening test can be extended by presenting subjects with the real and simulated recordings in a drum loop instead of single hits and simulating many different velocity layers.

6. REFERENCES

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer, 1998.

[2] H. Kuttruff, *Acoustics: An Introduction*, Spon Press, 2006.

[3] U. Zölzer, Ed., *DAFX - Digital Audio Effects*, Wiley, UK, 2002.

[5] `http://webprojects.eecs.qmul.ac.uk/alicec/fx_listeningTest/grooveExamples/`

[4] J. A. Moorer, “About this reverberation business,” *Computer Music Journal*, vol. 23, no. 2, 1979.

[5] O. Lartillot and P. Toiviainen, “A matlab toolbox for musical feature extraction from audio,” in *Proc. of the 10 th Int. Conference on Digital Audio Effects (DAFx-07),*, 2007.

[6] T. D. Rossing, “Acoustics of drums,” *Physics Today*, vol. 45, no. 3, pp. 40–47, 1992.

[7] J. Boley and M. Lester, “Statistical analysis of abx results using signal detection theory,” in *Proceedings of the 127th Audio Engineering Society Convention*, 10 2009.

[8] E. Vincent, R. Gribonval, and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, JUL 2006.