A new approach to the timbral representation of a nine-piece jazz drum set

John R. Taylor
Sydney Conservatorium of Music, The University of Sydney, NSW, Australia
john.taylor@sydney.edu.au

Abstract

Many commercial software-based percussive performance tools employ multisampling and programming techniques to simulate human percussive performance. However, several of these simulations aim to produce a level of timbral consistency which is, paradoxically, only achievable using a computer. This is because existing models fail to consider the micro-timbral nuances of human percussive performance, and the complexities of drums timbres. Consequently, opportunities exist in the intersection of music, science and technology to sonically represent a nine-piece jazz drum set comprising membranophones and idiophones. This presents challenges ranging from technological implementation and perceptual organisation, to the simulation of music/performance context; and opportunities include the exploration of percussive timbre for compositional purposes. This study, which is part of a wider research project¹, describes these challenges and opportunities and proposes a new approach to percussive timbre classification for the purpose of creating a timbral representation of drums for use in compositional software tools.

Keywords: percussion, music, technology, performance, model, simulation.

Acknowledgements: The author would like to thank Leon Gross, Dom Blake, Adam Wilson and Andrew Humphries.

Instrumental Mechanics

Percussion instruments can be classified in many ways, ranging from their sound production characteristics, their role in musical contexts, “whether or not they convey a sense of pitch” (Rossing 2000, 1) or by cultural derivation (Blades 1992, 33-36). A typical drum set configuration consists of a bass drum, snare drum, hi-hat, tom-toms (including floor tom), ride cymbal and crash cymbal (Sweeney 2004, 4), although the configuration of individual components (e.g. drum size) and the positional configuration of the set, can be extended by personal preference (Strong 2006, 12-14; 65-74). The standard jazz drum set can be divided into two groups by virtue of their sound generation methods, with the bass drum, snare and tom-toms belonging to the family of membranophones, and the cymbals belonging to the

¹ The aim of this research project is to create a compositional software tool that simulates variation in human percussive performance, by modelling both the timbral variation of drums, together with the biomechanical considerations of human percussive performance.
family of idiophones (Rossing 2000, 26-46; 89). Regarding membranophones, these three elements of a drum set are similar, insofar as they all have clamped circular edges, whereas idiophones are unclamped circular plates supported in the middle. The hi-hat differs from the ride and crash cymbals as it comprises of two opposing circular plates, supported in the middle with a clamping force between the two plate edges and controlled by the performer, in either an ‘open’ position (no contact), ‘closed’ position (forced contact) or somewhere in between.

**Vibrational characteristics of membranophones**

One of the main starting points for the theoretical consideration of membranophones is the membrane. Fixed at the edges, the membrane serves as a primary vibrator which can be described as two-dimensional, where vibrations travel in both radial (concentric) and azimuthal (diametric) directions (Moravcsik 2001, 188), and “the resonant vibrator is the air column inside the drum” (Moravcsik 2001, 191). It is these complex vibrations that produce “inharmonic overtones” that “give percussion instruments their distinctive timbre” (Rossing 2000, 2). In order to analyse these vibrational characteristics, many existing studies describe not only the membrane, but the environmental conditions in which the membrane will operate (e.g. an ideal membrane); for example, where the membrane is wholly flexible and vibrating in a vacuum (Rossing 1992, 84), where the membrane has “zero thickness” and is “perfectly elastic” (Moravcsik 2001, 189; Raichel 2006, 111), has “no stiffness” (Rossing and Fletcher 2004, 70) and is not subject to damping (Raichel 2006, 120-129).

Fletcher and Rossing (1998) identified fourteen different concentric and diametric modes, each corresponding to a different relative frequency, where mode \((0, 1)\) represents \(f_0\). The inharmonic overtones produced by these vibrations are represented by the frequencies given in Figure 1, which display non-integer relationships with the fundamental, in contrast to harmonic tones whose frequencies are integer multiples of the fundamental (Roads 1996, 18).

![Figure 1](image_url) / The vibrational modes of an ideal membrane, and their relative frequencies (Fletcher and Rossing 1998, 75).
Frequency modes in a real membrane are different to an ideal membrane, because real membranes are subject to air loading, bending stiffness and shear stiffness, in which the latter two raise the modal frequencies, while the former lowers them (Fletcher and Rossing 1998, 84). Furthermore, it is the excitation of a real membrane in different locations that causes different combinations of these effects (Moravcsik 2001, 189), with the creation and subsequent decays of various modes being unequal. In addition to the inequality of modal decay, different strike locations are more efficient in exciting modes that display a similar vibrational distribution to the strike (Hall 1991, 169). A strike on a nodal line or point will not excite that mode. Where an impact occurs comfortably within a region of a natural mode, that mode is excited efficiently in the same (positive) direction – see figure 2 (a) and (b) –, whereas an impact that overlaps several modes, excites all the modes within that region, and vibrations at different phases can causes cancellation – see figure 2 (c).

In the interaction between two membranes (heads) on a single drum (e.g. a snare drum has a batter head – top membrane – and a resonant head – the bottom or snare head), a coupling effect arises through either the enclosed air between the membranes or through the shell (Rossing et al. 1992). The effect of air loading in an enclosed two-head system differs from a single head, in that while the air loading of single head will lower the modal frequencies, enclosed air loading raises the axisymmetric modal frequencies (Fletcher and Rossing 1998, 75). Furthermore, coupled heads can either move in the same direction (in phase) or in opposing directions (out of phase) (Rossing 1992, 84-94; Tindale 2004, 15).

A comprehensive investigation into the properties of the bass drum sound was undertaken by Fletcher (1978), in which strikes of a bass drum were measured in an anechoic chamber, and compared with theoretical (ideal) values. These values were used to digitally synthesize bass drum tones, and a listening test was conducted with thirty-one adults to determine the real from the synthetic tones. The main finding of this investigation was a general agreement between theoretical and observed frequencies, although some anomalies were reported. Other findings included how, for a hard blow, the frequency decreases over time, the timbre is affected by strike strength and location, and the decay rate is dependent on the characteristics of the drum rather than the location or type of strike. The listening test revealed that bass drum sounds played on tape through loudspeakers are different from that of an actual drum, although it is worth noting that this might have been a technical limitation.
Another aspect of the bass drum sound, can be attributed to the moving parts of the bass drum pedal, which operate in the pre-attack phase of the sound. The release of the mechanism during the decay of the bass drum sound could also contribute to a perceived difference in the timbre of sound captured by microphones (Huber and Runstein 2005, 164).

One of the first studies to empirically measure the acoustical properties of a snare drum was that of Henzie (1960), whose focus was on the relationship between the amplitude and duration of snare drum tones, by using varying drum stick gauges with strokes at varying heights. The main findings of this study showed that by equating the durations of snare drum tones produced at different heights in inches (e.g. 1, 2, 3, 4...8) to note values at 120 beats per minute, there is a ratio of 4:1 between strike height, and amplitude and duration. Interestingly, this experiment omits strike location as an important variable in either amplitude or durational characteristics. During this investigation, Henzie described how different variables could affect tone production in drums, the first being the drumheads themselves (e.g. size, thickness, tension, and age and condition).

Henzie also describes how the construction of the drum (e.g. shell depth, materials and air ventilation) can affect tone production, particularly in relation to the coupling of two-head systems. Where coupling effects are propagated through shell vibrations, Rossing (1992) noted that energy transfer from a snare drum shell to the stand increased the decay rate of the lowest (0, 1) mode, compared to a freely supported snare drum, where the decay rate of this mode was similar to that of other modes. It was also noted that drum shell mass had an “appreciable effect” on the decay rate, and subsequently the timbre (Rossing 1992, 93).

Another variable described by Henzie affecting snare drum tone was the construction of the snare. The snare is a series of metal wires that are strung across the resonant (snare) head, can be set to different tunings, and are engaged with the resonant head by using a “strainer” mechanism (Tindale 2004, 20). When the batter head is struck, the resonant head moves due to the coupling effect causing the snare wires to bounce and interact with the resonant head membrane, which produces the ‘snare’ sound. Tindale describes how there is an “activation time” between the initial strike and the snare movement, although he asserts that a lack of investigation regarding elasticity and mass on the snares has failed to provide a mathematical representation of this phenomenon (Tindale 2004, 21).

One method used by percussionists to dampen the sound of a snare drum is the use of tape applied to the batter head, either across the diameter (Koblick 2007, 26) or as a small piece of tape near the rim (Rogers 2011, 7). This was computed by Yu and Wang (2001), with a later investigation into the placing of small strips of tape in arbitrary positions on a circular membrane undertaken by Yu (2004). It was found that the fundamental frequency decreased as the small strip moves towards the boundary of the membrane.

Tom-toms are available in many different sizes, and can be incorporated in a variety of ways into a standard drum set. Some of the smaller tom-toms (eight to sixteen inches) are usually mounted on the bass drum via metal shafts, while the larger floor tom (typically fourteen to eighteen inches) is supported by legs that are mounted on the side of the shell (Sweeney 2004, 4-5). Like other membranophones, the tom-toms are classified as un-pitched, although they do impart a perception of pitch. This sense of pitch is more prominent in tom-toms with only a single head (Fletcher and Rossing 1998, 515), compared to double headed tom-toms.
whose tendency is to produce an “indefinite pitch” (Berg and Stork 1982, 343). Describing how drum depth is a determining factor in tone quality, Berg and Stork state that double headed tom-toms with greater cylindrical shell lengths have “longer” standing waves and, as a result, produce a lower tone than double headed toms with smaller drum depths. It is these differences in tone and pitch conveyance, that distinguish tom-toms from the bass and snare drums, allowing a drum set to have a wider range of timbres.

There are also variations in tom-tom membranes, with some membranes being manufactured with large round dots in the centre, thus thickening the membrane in that location. The acoustical effect of these dots is described by Fletcher and Rossing, as “shifting the lowest partials into a more nearly harmonical relationship” (1998, 606), resulting in greater perception of pitch. In addition to the change in harmonicity of the membrane, the greater thickness of the dot also increases all modal decay times. The effect of strike force on the spectral characteristics of different sized two-headed tom-toms was investigated by Dahl (1997), who found an increase in strike strength resulted in an increase in modal excitations – including a change in spectral slope in frequencies above 1kHz – with a typical decrease of 9dB/octave as strike strength increases from soft to hard. Dahl also noted a frequency shift with stronger strikes, and describes this frequency shift as a characteristic of loud playing, rather than a pitch glide as observed in bass drum tones (Fletcher 1978, 1570-1576).

The decay times of snare drums are affected by its support mechanism, in which decay rates decrease due to vibration transmission through the stand (Rossing et al. 1992, 89). The implications of this in relation to tom-toms are significant, as different drum set configurations allow tom-toms to be mounted either on the floor, on the bass drum via an arm mechanism or via a clamp to a cymbal stand, with each of these mounting configurations having different effects on the decay rates of the drums.

In the case of tom-toms supported by a bass drum arm mount, a strong strike will transmit vibrations through the supporting arm mechanism, thus exciting either the shell or the membrane. In some instances, vibrations can cause rattles from lugs (Schroedl 2003, 19) and squeaks from moving parts; while other moving parts that can cause squeaks are necessary for drum striking (e.g. the bass drum pedal mechanism) (Huber and Runstein 2005, 164), this can overlap the sound during the recoil of the mechanism for the next strike, affecting tone perception. Strong vibrations can also result in the sympathetic production of resonance (Jaffe and Smith 1983, 66) of other supported tom-toms – for example, a double tom stand on the bass drum – and, depending on the construction and consequent sensitivity of the bass drum, both energy transmission and loss through the mechanism can affect the perception of the tone of the primary instrument (in this case, the tom-tom).

**Vibrational characteristics of idiophones**

One of the main features of an idiophone is that the “vibrating material is the same object that is played (free from any applied tension)” (Schloss 1985, 48), which includes xylophones, bells and cymbals etc., and can be either tuned or un-tuned (Fletcher and Rossing 1998, 533). This definition is based upon the classification system proposed by Hornbostel and Sachs (1961; 1914), which presents top-level classifiers based on the method of excitation, from struck and plucked to friction and blown (Benson 2007, 91; Kartomi 1990, 170). Idiophones
(cymbals) in the drum set are directly struck-upon percussion vessels (or sub-class 111.24) (Hornbostel and Sachs 1914, in Kartomi 1990, 170). Physically, this subclass of idiophones can be described as circular plates, or mechanically as “a membrane with stiffness” (Fletcher and Rossing 1998, 76), where the plate is a two-dimensional primary and resonant vibrator, which can also become three-dimensional with the planar deviations of the plate resulting from any striking action (Moravcsik 2001, 188). The plate shifting to three-dimensions is only relevant to plates with a free edge (e.g. a ride/crash cymbal), though there are different boundary conditions, such as clamped edges and simply supported edges (e.g. a hi-hat). The vibrational modes of a circular plate with the different boundary conditions are shown in figure 3. Note the fundamental mode (2,0) in the free edge compared to the (0,1) mode of a clamped or simply supported edge, and the similarities of the first four modes in figure 3 (b) in comparison to the vibrations of membranes in figure 1.

Although cymbals display vibrational characteristics similar to circular plates, Fletcher and Rossing (1998) describe how there are differences in behaviour between low and high frequency modes, in which lower modes are similar to flat circular plates, and at higher frequencies the modes merge together and become difficult to identify. Such behaviour can be seen in figure 4, which shows the Chladni patterns of twenty-three vibrational modes of a crash cymbal and the corresponding vibrational modes in a flat circular plate using electronic TV holography.
Although ride and crash cymbals behave mechanically in a similar way, there are some immediate dimensional differences that affect the overall characteristics of the sound, resulting in different uses of these two cymbal types. This can be seen in figure 5.
Typically, the ride cymbal is thicker than other cymbals (Black 2003), with a smaller “taper” (Pinksterboer 1992, 72) (a change in thickness from centre to edge), which produces more sustain, and is used mainly to play ostinato patterns (Black 2003, 24). In contrast, the crash cymbal is typically smaller and thinner (Black 2003, 25) with a larger taper (e.g. thinner edges) in order to produce more sustain, and is used primarily for accentuation and phrasing (Pinksterboer 1992, 26) thus being played less frequently and at a louder volume compared to the ride cymbal2.

Rossing (2000) notes that there are three key features in the sound of a cymbal. The first feature relates to the strike sound, after initial excitation, due to the initial wave propagation in the first millisecond. This is followed by a frequency increase of between 700-1000Hz for approximately 20 ms, with the third phase of the sound, occurring at around a second afterwards, containing frequencies of mostly 3-5kHz. It is this final phase of the sound that provides the “shimmer” effect (2000, 92).

Using double-pulsed TV holography, Schedin et al. (1998) captured the wave propagation as it passed through a cymbal. A laser pulse excited the cymbal at two points: one millimetre from the edge and at half radius. It was found that waves with longer wavelengths were more pronounced, and more reflected from the central dome and edge of the cymbal, compared to waves with shorter wavelengths. This transient behaviour will occur irrespective of how the cymbal is excited, due to the nonlinear coupling of the vibrational modes (Fletcher and Rossing 1998, 92-96; Touze and Chaigne 2000, 557-567). The nonlinear behaviour produced harmonics and then sub-harmonics before becoming chaotic in nature. This can be seen in phase plots for a crash cymbal (Wilbur and Rossing 1997; as cited in Rossing 2000, 95).

The hi-hat comprises two cymbals facing each other (Black 2003, 24), typically with a thinner (lighter) cymbal on top, and a thicker (heavier) cymbal on the bottom (Pinksterboer 1992, 72). These cymbals are mounted on a rod passing through the centre of each. The top cymbal is mounted to a foot-operated clutch which clamps the cymbals together, by lowering the top cymbal, thus providing an additional method of excitation and changing the boundary

2. For a comprehensive discussion on the history and differences between cymbals, see Pinksterboer (1992).
conditions of the cymbals to adjust the timbre. The spacing between the cymbals in the open position can be varied by using the clutch, from resting on the lower cymbal to being completely devoid of contact. Pinksterboer (1992, 88) describes how “a very tight clutch will deaden the sound of the hi-hat,” while Black (2003, 22) suggests that the optimum space is one to two inches to allow closing of the hi-hat with the foot.

An open hi-hat displays the same characteristics as a typical cymbal with a free edge, with the exception that the mounting on the rod prevents the third dimensional planar deviation. In the clamped position (closed) the two cymbals have a damping effect on each other, which decreases the overall decay time, with the coupling of the cymbals causing vibrations that are normally only reflected from the edge (in the case of a free edge boundary condition) to be transferred into the edge of the counterpart cymbal. In addition, vibrations are also transmitted through the rod mounting, producing a damping effect. The amount of decay and vibrational transmission through contact is dependent on two variables: the strike strength and the clamping force between the two cymbals. Both of these variables give rise to significant differences in sound.

**Drum tuning uniformity**

The vibrational characteristics of an ideal membrane in relation to the bass drum, snare drum and tom-tom were described in order to provide an overview of the complexities of tone production in membranophone percussion and included some of the differences between a real and ideal membrane. The complexity of these vibrational systems become compounded when one considers how an ideal membrane is mathematically modelled using Bessel functions which inherently assume a “uniformly stretched uniform circular membrane” (Bowman 1958, 20), which disregards the notion of membrane tuning dis-uniformity. Such dis-uniformity can occur in new membranes in which striking stretches the membrane causing a perceived detuning (Schroedl 2003, 20). Tuning can be defined as “the process of adjusting a musical instrument such that the tones produced by the instrument obey certain relations” (Christensen and Jakobsson 2009, 5). This definition, although describing relative tuning and aimed at stringed instruments (e.g. violin or guitar), is also relevant to membranophones, due to the relationship between modes caused by cross-tensional forces. In addition, the material properties of the idiophones described cannot be easily adjusted (e.g. tightened) to manipulate tonal production.

Drum tuning can be done in two ways: cross-tensional and clockwise (Black 2003, 4; Sweeney 2004, 7), although only cross-tensional tuning “maintains even tension throughout the tuning process” (Sweeney 2004, 7). In an investigation into the acoustics of snare drums, Rossing et al. (1992, 85) notes the lack of standard practice for tuning snare drum heads, while emphasising how the investigation relied upon achieving the most uniform tension possible. However, drum tuning is not only important when undertaking empirical acoustic research, but is also important for live performance and studio-based music production. From a live performance perspective, repeatability of drum setup is important when on tour, with tuning being important to tone quality and instrumental context in the recording studio (Toulson

---

3. This excludes the use of tape and other dampening techniques.

4. For an overview on drum tuning, see Schroedl (2002).
et al. 2008, 2). The investigation by Toulson et al. found that advanced musicians were able to understand and manipulate drum tuning, compared with amateur performers who appreciated the drum tuning, but could not tune their drums. Unsurprisingly, both Toulson et al. and Rossing et al. are in agreement regarding standard tuning practice, although Toulson et al. focuses on benchmarking different tuning setups for different musical genres, while the focus for Rossing et al was on experimental validity.

An acoustical analysis of the tuning of snare drums was undertaken by Richardson (2010), whose findings showed that detailed and accurate tuning was possible, and that modal frequency ratios that were previously considered fixed could be managed with tuning and damping to create new modal frequency ratios, thus creating a desired tone.

An early mathematical investigation into the vibrations of circular membranes with “non-uniform tensile forces at the edge” was done by Mei (1969), who attempted to identify the vibrational behaviour for a non-ideal membrane using a finite element method. In this research project, Mei found that lower modes were the same as an ideal membrane, though nodal patterns associated with higher modal frequencies became distorted, suggesting that dis-uniform tuning can have an effect on the timbre of a drum. This is apparent in later research into the simulation of a kettledrum by Rhaouti et al. (1999) in which, during a comparison of simulations and experiments, they state that: “the simulations were used in order to check whether or not this feature [beating] is due to imperfect tuning of the membrane, as it is usually assumed” (Rhaouti et al. 1999, 3556). In order to quantify this assumption, Rhaouti et al. simulated a kettledrum at both uniform and non-uniform tension, and presented a comparative example of tension distributions, shown in figure 6. The authors also noted that an adjustment in membrane tension within the model could yield similar results to modifications in other parameters of the model.

![Figure 6](image)

**Figure 6** / A comparison of (a) uniform and (b) non-uniform tension distribution. Higher areas of tension are shown in white, while lower areas of tension are shown in black (Rhaouti et al. 1999, 3556).

In order to assist the approximation of an ideal membrane during the tuning of drums, Worland (2008) analysed the modal patterns of a single-headed drum under non-uniform tension, using electronic speckle pattern interferometry (ESPI) in order to “image the mode shapes on the drumhead and identify corresponding frequencies” (Worland 2008, 5). In one experiment, the tightening of a lug by two turns caused the (1,1) mode to split and curve away from the tuned lug, as the opposing lug retained the original tension (e.g. a two-fold
perturbation), thus creating “perpendicular fast and slow axes on the membrane” (Worland 2008, 11). Worland also notes that the modal curvature due to irregular tuning is not “directly related” to the frequency splitting (2008, 7), and that higher modes (those outside of his research) “can be split by higher order perturbations in the applied tension” (Worland 2008, 11). Further research by Worland (2010) saw the expansion of this approach to include time-averaged ESPI on the (1,1) mode. The main finding of this study was that the splitting of the (1,1) mode via a two-fold perturbation was the “largest contributor to the sound of a drum not being in tune with itself” (Worland 2010, 533).

Drum set configuration

The typical drum set configuration has constantly evolved since the first drum kits in approximately the early 1900s (Starr 2009, 263), with changes resulting from either economic drivers or through stylistic changes in musical tastes over the decades (e.g. Be-bop jazz to Rock) (Aldridge 1994, 28-30). A typical modern drum kit configuration is described in Huber and Runstein (2005, 162), Murakami and Miura (2008, 450) and Strong (2006, 13), and consists of only the basic elements of a drum set, compared to extended configurations that cover “all of the instruments potentially used” (Murakami and Miura 2008, 450; illustrated in Murakami and Miura 2008, 451). Taking these configurations into account, Strong (2006, 65) describes both left and right handed configurations, the difference being that in a right handed configuration, the hi-hat is struck with the lead (right) hand, and the drum with the lead (right) foot, compared to the left hand or foot in the left handed configuration. Within different drum kit configurations, there are many ways to arrange the drums, usually based upon personal preference. However, the main criteria for positioning the drums are comfort, ease of use, and injury avoidance (Starr 2009, 12-13), although Black (2003) describes how positioning to “minimize reaching, stretching and twisting” is dependant on the performers’ “physical size and technical ability,” and how correct positioning “will help to assure optimum sound quality and volume while minimizing the possibility of damage to the cymbal” (Black 2003, 22).

Previous empirical research reveals that intra-class differences in vibrational characteristics and the subsequent timbral variation are the result of many different factors, most notably the construction of the instrument, either through differences in shell size or plate, membrane configuration (single or double head) or through supporting mechanisms. Despite the timbral variation afforded by the differences in vibrational characteristics of these factors, the vibrational behaviour of equivalent drums with similar or slightly deviating properties is largely the same. This is of particular relevance in the case of a performer’s ability to predict and exploit the vibrational behaviour of a drum on an unfamiliar drum set. From a theoretical standpoint, these similarities facilitate the synthesis of these instruments by implementing these similarities as generalised parameters in the synthesis process. This is particularly relevant to physical modelling synthesis that aims to simulate the physical behaviour of an instrument in order to produce a sound.

5. For a comprehensive history of the drum kit, see Aldridge (1994).
6. For further discussion on the proper positioning of a drum kit, see Starr (2009, 12-14).
Conversely, the generalised parameterisation of context in synthesis is more problematic, in that there are some physical characteristics unique to different drum sets. These are usually due to the configuration of the drum set and drums; one example relates to the choice and location of each drum’s supporting mechanisms (e.g. bass drum mounted tom-tom or an open/closed hi-hat), however, this example could also be related to preferences in individual configuration. By far, the most significant factor in the production of timbral variation lies with the level of uniformity of the tuning of a membrane. A dis-uniformly tuned drum can alter the vibrational behaviour of a drum set, and at a fundamental level affect the timbre production of the drum (particularly the micro-timbre) by exciting different modes than theoretical ideals. Ultimately, this leads to the potential for an infinite number of micro-timbral variations across the surface of a membrane for a single instrument, and has significant implications on the repeatability and timbral consistency of performance, and the computational methods associated with synthesis paradigms such as some physical modelling techniques.

Significant challenges occur when attempting to use a single synthesis technique to produce not only a sonic representation of the constituent drum types (membranophones and idiophones), but also the different vibrational characteristics and timbre within the classification types, and, furthermore, the potential number of micro-timbral variations across the instruments’ performance space. When considering other factors such as computational constraints (Bilbao 2010 & 2012; Masri 1996; Macon et al. 1998), timbral and perceptual resolution (Beauchamp 2010; Aramaki et al. 2006; So 2001; Horner 1995) and the number of control parameters (Roads 1996; Sandler 1990a; Sandler 1990b; Serra and Smith 1990), it is little wonder why synthesis techniques such as finite difference time domain (FDTD), additive synthesis, spectral modelling synthesis (SMS), linear predictive coding (LPC), waveguides and wavetable synthesis have failed to gain traction in real-time computer models of the jazz drum set.

Consequently, the most efficient method for representing instrumental timbre is through pulse code modulation (PCM) sampling, owing to its high quality approximations of the original signal (Klingbeil 2009, 5; Zagaykevych and Zavada 2007, 160) as well as its convenient storage, manipulation and playback capabilities (Bongers 2000, 42). However, this format is considered unexpressive, with expression developed at the expense of implementation (labour and hardware costs) (Kahrs and Brandenburg 1998, 315-317). Practically, this approach lends itself to extracting spectral information, which is of use not only for sonically representing percussive timbres, but also for future synthesis models.

The sample database

In order to fully represent the number of possible micro-variations of a drum, an exhaustive number of samples must be taken. This is, however, impractical owing to computational constraints. Limitations of existing sample databases, when considering the sonic representation of an instrument, include the number of samples taken per instrument. Hellmer (2006) used a database totalling 98 samples across nine instruments, of which a maximum of 28 samples represented the hi-hat, a minimum of four samples represented the bass drum, and four samples represented a crash cymbal. This falls short of the multisampling approach of many commercial software drum programs, whose databases use 127 drum sounds per instrument – the maximum number of samples assignable to a MIDI note. However, the
samples in these databases are intended to produce smooth timbral changes across the MIDI note velocity, with minimal micro-timbral variation.

In capturing percussive gestures, Tindale et al. (2004) recorded 1,260 samples of a snare drum using a brush and a stick at different membrane locations, to analyse the spectral features of the samples for timbral classification. Tindale acknowledges the timbral variation caused by excitation location by specifying five different strike locations, each location struck 20 times by three different subjects. These strike positions are shown in figure 7.

Although the specification of these strike locations provides a more detailed timbral representation of the snare drum, the locations are only representative of part of the membrane. Dis-uniform tuning can create areas of higher tension in different parts of the membrane, and cause changes to the vibration speed and boundary reflection times. It is important, therefore, to take samples from a range of locations across the drumhead, and not just specific locations across one line of a single polar axis and at various strike strengths.

Tindale’s results demonstrated the successful automatic recognition of samples with differences in timbre, resulting from different strike locations (Tindale 2004). These results validate the recording of samples across the entire surface of the drum, together with an increased sample size of at least 1,000 samples per instrument. With a total of nine instruments, the sample database in this model contained a minimum of 9,000 samples.

Timbral feature extraction

Extracting pitch values from un-pitched samples is extremely difficult, owing to the autocorrelation algorithms that attempt to find stable pitches where none exist. This results in artefacts and misleading pitch values, which are a concern when analyzing idiophones such as crash and ride cymbals. Feature extraction programs such as Praat (Boersma and Weenink 2013), a specialised speech analysis and synthesis tool7, and MIRToolbox (Lartillot and Toiviainen 2007), a suite of modular music analysis functions available for use with the

7. Although Praat is mainly concerned with speech analysis, it has been successfully applied to a flute in Hindustani classical music (Ramesh and Sahasrabuddhe 2008) for pitch estimation.
commercial software Matlab, both use autocorrelation in their pitch algorithms. Furthermore, pitch analysis in Praat is difficult with sounds containing higher noise floors (Boersma and Weenink 2013), of which cymbals can be considered “noisy” due to their nonlinear and chaotic behaviour.

In MIRToolbox, however, there are a number of other algorithms for extracting other features that do not use autocorrelation methods\(^8\), and are therefore suitable for use on both pitched and unpitched percussion. Such extractable musical features include brightness, spectral centroid and RMS energy. Strike strength is a common parameter used to map samples to MIDI keyzones in commercial software, particularly for simulating dynamics where spectral changes occur (Starr 2003, 12-13). One function that computes the global energy of the sample in MIRToolbox is RMS energy. Although using loudness as a parameter for composition is not new, it was useful to include this parameter for two reasons. Firstly, this parameter allowed for the recreation of performance as a means of evaluating the model. Secondly, and more importantly, from a compositional perspective this parameter is important for critical compositional devices, for example, crescendos, decrescendos, dynamics, sound duration, and timbre. Since average values per sample are required, segmentation and framing were not necessary.

### Compositional feature extraction

When looking at timbral parameters for musical form, it is useful to consider the most perceptually important timbral dimensions. One such dimension is brightness (Ara-maki et al. 2006; Barhet et al. 2008; Darke 2005; Donnadieu 2010; Giordano and McAdams 2006; Marozeau and de Cheveigne 2007; Pressnitzer and McAdams 2000; Schubert and Wolfe 2006; Turcan and Wasson 2003; and Risset and Wessel 1999, 147-148). The \( \text{mirbrightness} \) function in the MIRToolbox measures the proportion of spectral energy over a designated cut-off frequency (Juslin 2000, as cited in Lartillot 2010)\(^9\). The nine instruments in the jazz drum kit provide a rich palette of sound for compositional exploration, ranging from the low frequencies of the bass drum, to the higher frequencies of the cymbals.

Another useful timbral parameter is spectral flatness. In MIRToolbox, \( \text{mirflatness} \) determines the flatness of the frequency distribution from a ratio between the geometric and arithmetic mean (Lartillot 2010, 148). Spectral flatness is indicative of how closely the sound resembles white noise (Black 2003, 22), which is relevant given the propensity for cymbals to have very noisy spectrums (e.g. a high flatness value) (Black 2003, 22).

### Data collection: protocol and procedure

Recordings were taken of the isolated drums inside a studio reinforced with sound absorbing curtains in order to mitigate early reflections and outside noise\(^10\). The drums were arbitrarily

---

8. Later versions of MIRToolbox contain additional features such as inharmonicity, which uses autocorrelation methods (Lartillot 2010).

9. The default cut-off frequency in mirbrightness is 1500Hz.

10. The sound absorbing curtains and tracks were made by JANDS (www.jands.com.au).
tuned by the author with the intention of ensuring that the membrane tuning was not uniform, so as to generate as much micro-timbral variation as possible. A set of 'capture' rules were devised to ensure maximum timbral variation and to ensure a random sample of strikes from the widest possible surface area of the drum. Firstly, each drum should be struck at least 1,000 times (1,000 +5%). Secondly, each strike must be: a) executed with a stick; b) devoid of expression; c) a single stroke; d) in a different location to the previous stroke; e) at a different strike strength to the previous strike; and f) commence only when the sound from the previous strike is perceived to have ended. Any mis-hits (for example, strikes that include contact with the rim) will be included in the database. Due to the excitation method of a bass drum using a bass drum pedal and beater, rules (a) and (d) do not apply to this drum.

Equipment

A full list of drum set instruments can be seen in table 1.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Size (in)</th>
<th>Depth (in)</th>
<th>Height (in)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bass drum</td>
<td>22</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Snare drum</td>
<td>14</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>Hi-hat</td>
<td>14</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>Floor tom</td>
<td>16</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>Low tom</td>
<td>12</td>
<td>11</td>
<td>36</td>
</tr>
<tr>
<td>Medium tom</td>
<td>13</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>High tom</td>
<td>10</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>Ride cymbal</td>
<td>20</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>Crash cymbal</td>
<td>16</td>
<td>3</td>
<td>49</td>
</tr>
</tbody>
</table>

Table 1 / A list of drums used, their size, depth and height (top to floor) in inches.

Each drum sound was recorded using two Neumann KM140s, due to their cardioid characteristics, non-coloured reproduction and suitability for close-miking with percussion. They were also chosen due to the quality of the reproduction against other microphones\(^{11}\). The microphones were positioned in an X-Y configuration, exactly 11” (approximately 30cm) above the centre of each drum. Other equipment used during the recording process included:

- 2 x Canon HG21 HDD video Cameras. One camera was situated directly above the drum to capture the strikes from above and the other camera was placed to the side to capture an alternative angle;
- 1 x 24” Apple iMac; and
- 1 x Millennia pre-amp.

All strikes were recorded continuously (in one take) into Adobe Audition as 44.1kHz, 16-bit, .wav files.

\(^{11}\) A pair of B&K pressure sensitive microphones and a HATS device was also used to record the individual strikes, although the audio was discarded in favour of the audio from the KM140s.
The low, medium and hi-toms were kept on their respective stands during the capture of the strikes to ensure that accurate representations of the instrumental decay through the stands were maintained. The toms were mounted on the bass drum, with the exception of the hi-tom, which was mounted on the crash cymbal.

Data preparation

Once 1,000 drum sounds were recorded, the resulting audio files were imported into Pro Tools 9, where the attack points were identified using Pro Tools Beat Detective. Each attack point (the first zero crossing before the attack) was then visually accounted for and exported as regions into separate audio files with a .wav extension. Each of the individual audio files (or samples) were manually accounted for and truncated in Wavelab 5, in order to compensate for perceived differences in the sample end point, between the strike during recording (see procedural rule [f] above) and the actual recording and representation of the recording in a digital system. As there were no changes to the audio file format, sample-rate or bit depth during the use of Pro Tools or Wavelab, no audio artefacts were added to the sample database.

Feature extraction

MIRToolbox was chosen to analyse the audio samples, owing to the large number of musical feature extraction functions it offers, particularly those related to timbre (mirrms, mirbrightness, mirroughness, mircentroid). MIRToolbox was also chosen because of its capability to conduct batch operations, as well as its capacity for a range of analysis output formats.

Analysis

Although it was noted earlier that loudness, spectral centroid and spectral flatness were the key features to be used in this study, additional features were extracted. The reason for this is two-fold: firstly, at the time of analysis, the efficacy of the approach and analysis of the instruments were untested and, in the event that analysis of these features failed, time would be saved in the re-analysis; and secondly, having completed the analysis, data exists for further work, where different features can be applied to the model. The features extracted from the audio files can be seen in table 212.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Measurement/Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudness (RMS)</td>
<td>DBFS</td>
</tr>
<tr>
<td>Bark Envelope</td>
<td>Float value (0.0=min, 1.0=max)</td>
</tr>
<tr>
<td>Brightness</td>
<td>% of spectral energy &gt;1500Hz (0.0=min, 1.0=max)</td>
</tr>
<tr>
<td>Centroid</td>
<td>Frequency (Hz)</td>
</tr>
<tr>
<td>Cepstrum</td>
<td>Frequency (Hz)</td>
</tr>
<tr>
<td>Flatness</td>
<td>Float value (0.0=min, 1.0=max)</td>
</tr>
<tr>
<td>Roughness</td>
<td>Float value (0.0=min, 1.0=max)</td>
</tr>
</tbody>
</table>

Table 2 / A list of features extracted from the audio files using MIRToolbox.

12. All values are a mean average of the total audio file.
The resultant values of the measurements for each instrumental sample were aggregated into a text file representing each instrument. These text files were then used to evaluate the variation in features between the captured strikes. Once the audio files were analysed using MIRToolbox, the data showed that a number of individual samples had failed to produce a valid measurement. However, these processing failures amounted to only 0.04% of the total samples processed, well within the ±5% tolerance that was factored into the number of strikes taken. Samples that produced failed measurements were removed from the database.

**Initial results of instrumental micro-timbre**

Results from the MIRToolbox analysis showed variations in loudness, flatness and spectral centroid for each of the instruments’ captured samples. These are shown in figures 8 to 16, and are grouped by membranophones (figures 8 to 13) and idiophones (figures 14 to 16). The results demonstrate that the proposed method for capturing variations in percussive strikes produced a sample database that adequately represented the micro-timbral variations in each instrument.

The results show adequate variation in the ratings of samples, between the loudness, flatness and spectral centroid features. A comprehensive range of values in each parameter has been captured, allowing for interesting compositional devices, such as crescendo, where the database is classified by loudness ratings.

![Figure 8](image1.png)

**Figure 8** / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the bass drum samples captured.

![Figure 9](image2.png)

**Figure 9** / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the snare drum samples captured.
Figure 10 / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the floor tom samples captured.

Figure 11 / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the low tom samples captured.

Figure 12 / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the medium tom samples captured.

Figure 13 / Variations in (a) loudness; (b) spectral flatness; and (c) spectral centroid in the high tom samples captured.
Parametric variation inherent in the sample database, within and between features, indicates that any parametric classification will feature different schematisations, resulting in differing sample selections. Furthermore, these findings also highlight the inherent timbral variation within each instruments’ dataset. The findings also suggest that there are sufficient differences between each of the parameters to produce different database structures, particularly where the database is ordered by minimum-maximum feature values (similar to most MIDI volume key assignments). This will result in different samples being played depending on the parameter, irrespective of whether the parameters are ordered in the same way. However, it is acknowledged that this is not a conclusive timbral representation of the instruments owing to the limited feature extraction and the multi-dimensional nature of timbre. This limited representation of the instruments presents problems with the database in its current form.
These figures do not represent relationships between other parameters excluded from the analysis and, therefore, do not provide a comprehensively detailed timbral description of the strikes. As a result, it is possible that two samples with similar values in a database may exhibit differences in secondary timbral features. Conversely, two samples with different database values may have similar secondary timbral features. Since one of the main objectives of the study is to simulate human percussive timbral variation, any timbral variation between two strikes of similar value in these parameters will add to the perceived variability of human performance. However, such a large dataset has the potential to create excess variation, in three contextual effects: instrumental, performance and compositional.

Instrumentally, each database is a sonic representation of the sound produced when striking across the entire surface of the instrument. In the case of membranophones, excitation location is a significant cause of the vibrational characteristics of the drum and – although locally these variations may be consistent – across the surface of a drum, the vibrational differences may produce large differences in other timbral features. This is relevant when considering large differences between minimum and maximum values in a database. One example of this relates to the vibrational differences between strikes at the centre of the membrane and those close to the rim where membrane stiffness may be higher; increasing the pitch with the reflection of transversal waves affect the decay time. Since there is little evidence to support the correlation between flatness and centroid database schemas, as well as other timbral features, it must be assumed that the database will consist of other saliently unordered timbral features.

From a performance perspective there are two implications. Firstly, the dataset consists of samples at different strengths, causing changes in the vibrational behaviour and, subsequently, changes in secondary timbral features. One example of this relates to changes in frequency observed in tom-toms with greater strike strength, in which pitch was difficult to measure given the autocorrelation function of the analysis method. The selection of a sample with a higher variation in salient timbral features may result in the model playing a strike that is out of context (e.g. the performance of an accent in a structurally atypical place). This is tightly linked to the second implication, which relates to problems in individual sample selection, where the control algorithm, as an abstraction of the performance context, must select a specific value to play a sample, representative of a particular performance context.

A 3D correlation of the spectral features, as shown in figure 17, provides further insight into the timbre model, particularly in relation to instrumental behaviour and compositional implication. Concerning instrumental behaviour, there are clear correlations between centroid and flatness, which tend to increase with loudness, particularly among the idiophones. The tom-toms also display a characteristic ‘boomerang’ or ‘hockey stick’ curve, in which there is a tendency for higher levels of flatness with lower loudness, and a higher centroid with increased loudness. Similarly, the bass drum also displays higher flatness with lower loudness, although the correlation is less distinct compared to the tom-toms, possibly owing to the bass drum pedal mechanism. The snare drum also exhibits higher flatness with lower loudness, and the centroid is fairly consistent across the loudness and flatness dimensions. With instrumental behaviour governing the distribution of points in each of the instrument’s 3D loudness/timbre space, the sample selection paradigm allows for the arbitrary selection of any point within the timbre space, thus intersecting instrumental behaviour with composition. Understanding the correlation between the spectral features of an instrument’s behaviour also allows for greater...
compositional freedom, as correlated timbre space (e.g. higher loudness and high spectral centroid) can be used in conjunction with orthogonal point selection in more uncorrelated timbre space (e.g. lower loudness and high spectral flatness).

These findings also support the compositional validity of employing database classifications based upon the timbral features of loudness, flatness and spectral centroid for the purposes of exploring the micro-timbre of percussion instruments. Since the compositional perspective consists of an exploration of percussive micro-timbre using the database schemas as compositional parameters, there is a necessity to mitigate the effect of uncorrelated, salient timbral features, which are analogous to the implications in an instrumental context. In summary, these findings support the use of these parameters for the modelling of performance variation, and for compositional use, although the database must be timbrally constrained in a way that is useful in both applications. It is therefore necessary to perform data reduction on the dataset of each instrument in order to confine the timbral variation.

Figure 17 / 3D correlations of the spectral features for a (a) bass drum; (b) snare drum; (c) hi-hat; (d) floor tom; (e) low tom; (f) medium tom; (g) high tom; (h) ride cymbal; and (i) crash cymbal.
Data reduction and classification by strike location

The dataset of each instrument lacks both instrumental and performance context, which must be addressed by way of data reduction. One way the dataset can be reduced is to classify the samples according to strike location, in which excitation location can potentially excite different modes of vibration, thus causing changes in timbre. This is particularly the case for membranophones, where the impact of dis-uniform tension is likely to produce localised timbral similarity. Furthermore, as vibrational modes operate both concentrically and diametrically, it is useful to consider such an approach to classifying the samples – in which more complex movement affects strike accuracy on different radial planes on the skin – that traverse both diametrical and concentric modes. Furthermore, a change in strike location may be deliberate in trajectory planning.

However, classifying samples based on existing vibrational modes are not conducive to inferring a performance context, owing to extremes in precision that the modes would infer regarding performance, from a precise (6,3) mode, to the less precise (1,1) mode (see figure 1). Consideration must be given to the reclassified sample sizes. Therefore, a performance-based demarcation, adapted from Fletcher and Rossing’s (2,2) mode, has been chosen to re-classify each instrument’s samples. This is shown in figure 18, with: (a) the original (2,2) mode; (b) the adapted membranophone demarcation; and (c) the adapted idiophone demarcation. The centre area in (b) relates to the tendency for a drummer to strike in the centre and the option for centre spots on the tom-tom. This area corresponds to 1/3 of the diameter. In (c) the centre area relates to the bell in the centre of the cymbal.

To reclassify the sample database, a trace of each drum from the overhead video footage was drawn onto acetate, with the demarcation points calculated from scaled measurements of the video at full screen. Each strike was visually accounted for and assigned a number based upon the location of the strike, relative to the demarcation zones for the respective instruments in (b) and (c) above. Since the strike location of the bass drum pedal is constrained to one location by the mechanism, such demarcation does not apply. The bass drum is the only instrument with such a large database. The resultant sample reclassification increased the total number of potential zones in which a performer can strike and reduced the sample size in each zone, while maintaining adequate timbral variations in each zone. Graphical representations of each parameter and zone are shown in figures 19 to 23 (membranophones) and figures 24 to 26.
(idiophones). This approach also facilitates the potential use of stochastic methods to infer performance inaccuracies, contextualising strike locations relative to other drums by possible zonal strike weighting developed from rules generated from performance analysis.

Figure 19 / Snare drum sample features after modal demarcation
(loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 20 / Floor tom sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.

Figure 21 / Low tom sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 22 / Medium tom sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 23 / High tom sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 24 / Hi-hat sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 25 / Ride cymbal sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
Figure 26 / Crash cymbal sample features after modal demarcation (loudness, flatness and centroid, left to right). Modes 1-5 are (a) to (e) respectively.
It is not the intention of this study to describe the differences or trends between each of the instruments. However, there are a few observations that should be noted from the graphs, which are useful for this investigation. The findings suggest that, in general, there are consistencies between each zone in each instrument. For example, the flatness and spectral centroid of the snare drum are generally the same across the five zones, with similar ranges between minimum and maximum values, even though each zone consists of different strikes. The data reduction method has retained a large diversity of values in some zones, most notably in the tom-toms and idiophones. Additionally, there are large differences between idiophones with the hi-hat having higher average flatness and spectral centroid values compared to the ride and crash cymbals, which display similar characteristics. This can be attributed to the interaction between the top and bottom cymbals.

A significant implication of this data reduction method relates to the allocation of samples to zones. The findings indicate that there are different numbers of samples in each demarcated zone. In some instances there are very few samples compared to other zones in the same instrument. An example of this is shown in figure 25 (d), ride cymbal mode 4. In the context of samples from this mode being played in sequence with samples from the other modes, timbral variations may be produced as values are skipped in the parametric mapping, resulting in the selection of relatively higher (or lower) parametric values for similar mapping values. This will result in an ordered database of a zone, producing variation between values in the same database and between other similar values in other zones of the same instrument. One hypothesis is that the variations produced by the different sample numbers in each mode will produce variations significant enough to be considered ‘accented strikes,’ subsequently assisting in the model by conveying a greater sense of human performance. This is in contrast to the complete dataset producing “irrational” strikes, which are strikes so timbrally and dynamically irregular that they convey a greater sense of artificial performance. This data reduction method localises secondary timbral features, thereby making the timbre of the samples in each zone more consistent, particularly for membranophones.

Thus, the demarcation of the drums based upon a combined approach to instrumental mechanics and performance lends itself to the structural implementation of the audio database for use in a performance model.

Conclusion

It should be noted that the drum set used in the experiment was a typical, and relatively inexpensive, jazz drum set. It had seen considerable use in different environments and, therefore, was not in the best condition. This resulted in poorer quality samples, which contained timbral inconsistencies caused by the use of lower quality materials in the construction of the drums. In addition, at the time the samples were collected, the performance classification concept was still in its infancy. This meant that some instruments had greater variations in amplitude and timbre across performance modes at similar parametric values. However, large timbral variations might also be attributed to the size of the strike surface area in each of the demarcated performance zones, encompassing the area close to the centre and,

13. This is a result of the random nature of the sample capture, in which data reduction via the demarcation points were neither required nor conceived.
at the other extreme, the rim. Tension in the drum is higher towards the rim, thus the timbral variation is inherently different.

One of the most important factors to consider, when attempting to sonically represent a nine-piece jazz drum set for sound synthesis in computer modelling, is the inherent micro-timbral variety in the instruments. The sampling synthesis paradigm served to identify the protocols and procedures necessary for capturing the maximum number of possible timbres for each instrument, and included a set of rules that governed the performance of strikes for data capture. The levels of micro-timbral variation, across the different parameters, attest to the robustness of the protocols and procedures. The methods used to capture the strikes, including the microphone type and the positioning of each instrument, produced clear samples with minimal tonal coloration. In addition, the neutral-reproduction characteristics of the microphones effectively facilitated the capture of micro-timbral variations, which were audible across the sample database and were visually represented in the graphical analysis of the samples.

In order to maintain experimental integrity, the sample database had to be prepared for analysis. The beat detection points were manually checked and verified using Pro Tools Beat Detective. This prevented truncation errors that could have negatively impacted the sample analysis. However, this processes was time-consuming and labour-intensive. One area for further research might be to consider alternative methods, requiring less manual intervention, for automatic beat detection. Once the truncation points were manually checked and verified, MIRToolbox was used to analyse the sample database. This process was relatively straightforward, and the sample database yielded minimal calculation errors (approximately 10% for each instrument).

It is important to note that it was not the intention of this study to appraise the applicability of the analysis functions to the dataset, owing to the potential for idiosyncratic behaviours in the construction of these functions when applied to percussion instruments, particularly cymbals, as noted with the autocorrelation function in some of the analysis techniques. In order to address this issue, one area for further research might be to repeat the sample-collection procedure using a higher quality drum kit and collect more samples so that each zonal classification contains the same number of samples. This would provide an opportunity to explore the use of MIRToolbox as an analysis technique for the compositional parameterisation of percussion by producing another, larger, dataset for each timbral parameter. It would also create greater stability in the sample selection by mitigating timbral and amplitude disparities caused by having unequal numbers of samples in each classification. Increasing the number of samples would also increase the number of performance modes. This would enhance the resolution of the instrumental sonification. Although an increase in the number of demarcated performance zones would increase the resolution of the performance context, such an approach would require more parametric datasets and more contextual rules.

Once the sample database had been analysed in MIRToolbox, it was reclassified using three feature-based parameters (loudness, spectral flatness and spectral centroid), and reordered based on these parameters (low to high). When listening to the database, it became clear that the process of sample reclassification and reordering would produce some very interesting aural and compositional results. At the same time, however, the linear ordering of samples by feature had one unexpected consequence. This was due, primarily, to the large number...
of samples in the database. The linear ordering of samples, using the three feature-based parameters, rendered other timbral features nonlinear. For example, when the sample database was manipulated to increase loudness levels, from low to high, it produced random pitch variations. This can be attributed to the instruments’ wide timbral variations. This issue was addressed by classifying the drum strikes by location, taking into account the vibrational characteristics of drums. Reducing the surface area of each group of samples ensured that each sample, within each group, was subject to similar vibrational behaviours. This resulted in a reduction in differences in timbral characteristics.

It was necessary to demarcate the surface of each instrument of the nine-piece jazz drum set in order to maintain the timbral consistency of the sample database. One unexpected consequence of arbitrarily tuning the membranes was the presence of timbral inconsistencies in the sample database. Consistent tuning would have made timbral variations less prominent and negated the need to demarcate the surface of each drum component into smaller strike zones. Nevertheless, from a performance-modelling perspective, the decision to demarcate strike zones across the surface of each drum instrument marks an entirely new approach to generating micro-timbre. Having said that, further research might assess the viability of this approach on uniformly tuned drums.

While the samples in the database displayed other, secondary, timbral characteristics, each of the feature-based parameters (loudness, spectral flatness and spectral centroid) had their own, unique characteristics. The ‘loudness’ parameter, for example, worked as expected across all of the drum instruments. When assigned linearly to MIDI velocity, the changes in amplitude were commensurate with the velocity value. The timbral variations across the loudness curve were more pronounced than expected. However, this can be attributed to the strike area of the demarcated performance zones.

Likewise, using a similar MIDI velocity-mapping process, the ‘spectral flatness’ parameter produced some interesting timbral, and dynamic, results, which differed across the instruments. A higher spectral flatness value produced a flatter sound in the snare drum. Notably, the snare drum was consistently lower in amplitude at higher spectral flatness values across the demarcated zones. Similarly, the amplitude of the hi-hat changed with increased spectral flatness. Samples with the greatest spectral flatness had a characteristic ‘closed hi-hat’ sound, indicating that the vibrational interaction with the bottom cymbal caused greater spectral flatness. The decrease in amplitude, with increased spectral flatness, indicates a peak strike-strength equivalent to the maximum amplitude of the vibrational interaction. This closely resembled white noise, compared to louder strikes. It suggests that strikes with attack transients, greater in amplitude than the vibrational interaction of the two cymbals decay, produce lower levels of spectral flatness.

The amplitude of the tom-toms behaved in much the same way as the hi-hat, with spectrally flatter samples having lower amplitudes. This is consistent with changes in spectral slope, which reduced the spectral flatness for increased strike strengths. Conversely, the ride and crash cymbals were louder at higher spectral flatness levels. Stronger strikes excite a large number of non-linear chaotic frequencies very quickly. Therefore, in the case of the cymbals, this is most likely attributed to the chosen method of calculating averages in the MIRToolbox mirflatness analysis algorithm, whereby frequencies are measured as an average over the duration of the sample. Further research might explore the suitability of this algorithm on
percussion instruments for the purposes of producing more accurate representations of spectral flatness at different strike strengths. Further research might also investigate the compositional possibilities of combining the inverse amplitude mappings (particularly of the ride and crash cymbals) and the inherent secondary timbral features.

The spectral centroid parameter, with similar MIDI velocity mappings, produced some interesting compositional results, despite producing unsurprising acoustical results. In certain instances, the spectral centroid parameter behaved in much the same way as the spectral flatness parameter, particularly in the case of the snare drum. In other instances, however, the spectral centroid parameter produced opposing results. For example, the hi-hat tended to produce more ‘open’ sounds at maximum flatness. Nevertheless, this was expected and was most likely caused by the increase in the number of frequencies excited and the longer decay times, which corresponded to an increase in strike strength, resulting in a greater proportion of spectral energy in frequencies higher than the spectral centroid cut-off. This is also true of the crash and ride cymbals. The tom-toms displayed similar amplitude characteristics to that of spectral flatness, although the samples had a brighter decay with a higher spectral centre of gravity. Moreover, as the spectral centre of gravity increased, there was a small, but noticeable, change in pitch. A further area of work might repeat the sample database analysis, using a different threshold level than the 1500Hz default setting in MIRToolbox.

References


