The metal scene has undergone an evolution in sonic characteristics that is partially rooted in music technology developments and associated production techniques that have given rise to a set of identifiably distinct ‘metal’ timbres. The science of psychoacoustics provides a measurable way of assessing these perceptual changes. This article proposes a methodology for psychoacoustic analysis in order to track the evolution of metal-specific timbres, and to speculate on the development of, and applications for, timbral metering in metal production. Sampling, synthesis and digital editing techniques have given rise to entirely new genres of music, yet metal as a genre has remained ostensibly with much the same musical format. Nevertheless, the sonic fingerprint of modern metal production often bears the hallmark of these computer-aided developments. In particular, the metal audience is now, consciously or not, used to the sound of triggered drum samples and sound replacement, ‘re-amping’ guitar and bass signals, and the restricted dynamic range of the ‘loudness war’. As a genre that is known for taking joy in all things loud, both anecdotally and in academic research, the issue of manipulating dynamic range compression in order to increase perceptual loudness in modern music production is perhaps even more pertinent to the metal genre than in other popular music forms.
INTRODUCTION

It is not uncommon for metal audiences to be critically invested in the production value of new releases. For example, a search of Internet review archive Encyclopaedia Metallum of Carcass’s 1993 release, Heartwork, finds that six out of ten amateur reviews contributed to the site explicitly detail the production values of the release, amongst the more genre-universal critiques of songwriting, style and performance (Various 2013). The commercial metal press often present ‘in the studio’ reports with bands working on new material, and traditionally, liner notes include credits for producer, recording engineer and studio details (Berelian and Dickinson 2005).

Perhaps the prevalence of Digital Audio Workstation (DAW) software means that, as well as being a largely musically literate audience (many metal music fans can play one or more instruments), the audience’s understanding of production techniques has also increased. Ultimate-metal, a popular Internet forum, dedicates a section to the subject, and the popular music technology press now often feature metal-specific production tips and features for the enthusiastic amateur audience (Mynett 2009a; Vento 2009).

This article extracts various acoustic measurements from example recordings of individual instruments and full mixes on commercial releases. The common sonic characteristics revealed by an analysis of these measurements suggest that the development of digital recording technology and particular production methodologies has been accompanied by a noticeable timbral feature-set in modern metal productions.

This article proposes the application of acoustic measurement techniques to an analysis of the developments that have gradually evolved the timbral characteristics of metal. The article will first introduce existing timbral attributes and descriptors before examining specific production techniques for metal in a selection of small acoustic analyses. Though the musical selections used here are fairly mainstream, a certain amount of bias in the analyses presented is inherent in the authors’ choices – the examples are all taken from bands that have a heritage in the late 1980s and early 1990s, a period that the author is personally nostalgic towards. This time frame is also useful as it corresponds to the major changes in digital recording and editing that are now commonplace in the recording world. Nevertheless, the framework used in these analyses is transferrable to other examples.

DEFINING PRODUCTION ANALYSIS

Changes in music technology have often influenced the creation of entirely or relatively new musical genres (for example, the emergence of hip hop, house music, or more recently, dub-step). Yet, metal has, somewhat uniquely, both endured and evolved to utilize new music technology within its existing sonic blueprint whilst the core of distorted or overdriven guitar, bass, drums and vocals has remained. Psychoacoustic techniques allow for a unique window to the changing production styles that heavy metal has undergone in the relatively recent period of computerization. This article hopes to illustrate the application of these techniques to metal production analysis, in other words: ‘what can the study of music technology and psychoacoustics reveal about state-of-the-art metal production techniques?’

Sonic evaluations can yield acoustic cues for production analysis. This article focuses on a small selection of releases from the last two decades. Acoustic measurement is certainly less common than work dedicated to the cultural,
TIMBRE AND ACOUSTIC CORRELATES

When addressing sound perception, psychoacoustic research generally regards there to be three perceptual attributes of sound, loudness, pitch, and timbre. Both loudness (Stevens 1955) and pitch (Stevens 1935; Verschuure and Van Meeteren 1975; Hartmann 1997) now have acoustic correlates determined by listener testing, and are well understood by acousticians, musicians and most casual listeners alike. This list of perceptual attributes has been further enlarged by some research to include perceived duration, spatial location and reverberant environment (Levitin 1999). Spatial location and reverberant environment are particularly relevant to modern production with spatial panorama, phase and reverberation used as day-to-day ‘effects’ in a DAW, though of all these attributes, it could be argued that perceptual loudness in particular is of significant interest to the heavy metal audience (see ‘Loudness and specific loudness’, this article).

Regardless of this expanded collection of perceptual attributes, timbre is still considered to be the attribute that listeners use to distinguish sounds when they are of otherwise equal pitch and loudness. The American Standards Association (ASA), defines timbre as that attribute of sensation, in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar (ASA 1960; Definition 12.9). This definition is a not an easy starting point in that it defines timbre by what it is not, rather than what it actually is. One example illustrating this considers an unpitched sound such as that of rustling leaves. According to the ASA definition, this sound would have no timbre (Sethares et al. 2009). A more robust definition is given in response by Pratt and Doak, as ‘the sensation […] whereby a listener can judge that two sounds are dissimilar using other criteria than pitch, loudness or duration’ (Pratt and Doak 1976: 317). Prior to the ASA ‘non-definition’ of timbre, a number of acoustic correlations for timbre were suggested, including Harmonicity (Seashore and Rothschild 1934) and other spectral characteristics – specifically the number, distribution and relative intensity of partials (Fletcher 1934).

Schouten (1968) subsequently expanded upon these proposed timbral acoustic correlates to include:

- The range between tonal and noise-like character
- Spectral envelope
- Time envelope in terms of rise, duration and decay
- Changes of spectral envelope (formant-glide) and fundamental frequency (micro-intonation)

Music Information Retrieval is an interdisciplinary field mainly focused on techniques for extracting meaningful data from music automatically. The data targeted varies widely but includes both acoustic measurements and musical descriptions of form and structure.
The prefix: an onset of a sound that is quite dissimilar to the ensuing lasting vibration. ‘Tone colour’ and ‘sound quality’ have occasionally been used synonymously with timbre (Erickson 1975; Helmholtz and Ellis 1954; Scholes 1986). However, ‘tone colour’ can imply that the spectral properties of the sound are solely responsible for its timbre, contradictory to research specifying the importance of temporal acoustic correlates with relation to the perception of timbre (Krumhansl 1989). Moreover, sound quality is most often found associated with playback, stereophony, multichannel sound and loudspeaker performance measurement. Therefore, work focusing on these instances of sound quality (Choisel and Wickelmaier 2007; Flanagan and Moore 2000) are avoided here.

When discussing the contribution of harmonic overtones, attack and noise content to timbre as a whole, Pellman (1994) provides a summary of various descriptors which categorize timbre by relating stimuli to listeners’ previous experience of musical instruments: bright, dark, mellow, hollow, pure, raspy, breathy, horse, smooth, abrupt, sharp, gentle and easing. In summary, timbre is usually defined as one or more of the following:

- Everything (sonically) that is not loudness or pitch or reverberant quality or duration
- A perceptual attribute by which a listener can judge two sounds, with the same loudness and pitch, as dissimilar to one another
- A multidimensional perceptual attribute, with various acoustic correlates (spectral, temporal or spectrotemporal)
- The type of sound an instrument makes, be it emotional or instrument specific
- A high level descriptor which uses learned cues to identify non-musical sounds

For the purposes of this article, the term ‘timbral attribute’ shall be taken to mean a verbal descriptor which existing research has shown contributes in whole or in part to listener perception of timbre. The term ‘acoustic correlate’ shall be taken as any acoustic property, or combination of properties, shown to correspond to listener perception of a specific timbral attribute, group of attributes or timbre as a whole.

ACOUSTIC OVERLAP IN TIMBRAL ATTRIBUTES

When determining acoustic correlates, perceptual scaling techniques have allowed researchers to use listener judgments to evaluate the cognitive distances between sounds, moving towards a model of acoustic correlates within a ‘timbre space’ (Grey and Moorer 1977). The procedure is typically as follows:

- A stimulus set is generated (either by synthesis, or acoustic modification of recorded sound), with equal pitch, loudness and duration
- A descriptive scale is selected, sometimes from prior verbal elicitation, using subjective terms (for example, sharp/dull, compact/scattered, bright/dark, full/empty, colourful/colourless (Von Bismarck 1974a)
- Paired sounds (synthesized or manipulated samples) are rated numerically according to similarity by a listening panel

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52.
A multidimensional scaling analysis is used to find the best-fit solution to the listener responses. The distance between stimuli in the solution illustrates their perceptual similarity. Dimensions in the solution are interpreted as timbral descriptors across the timbre space.

Some attributes have been found to share a degree of overlap in their acoustic correlates. One example can be found in the timbral descriptors bright and sharp. Anecdotally it is quite plausible to imagine these terms being used with a degree of overlap in reference to electric guitar timbres. In such a case, sharper timbres might be found at the far end of an imagined brightness scale. Brightness has been shown to be acoustically correlated to spectral centroid and fundamental frequency (Schubert and Wolfe 2006). Sharpness has been shown to be correlated to spectral centroid as a function of critical band rate, measured in ‘Acum’ (Von Bismarck 1974a; Zwicker and Fastl 1999). This acoustic overlap is somewhat congruent with the descriptive overlap found in the example above. The difference between the metric for each is shown in Equation 1 for brightness, and Equation 2 for sharpness. The difference is that the spectral centroid calculation for brightness has no factoring by human frequency sensitivity, perceived loudness or final perceptual unit. Rather, it is a solely acoustic parameter, found to be highly correlated to the sensation of brightness through subsequent listener testing.

**Equation 1:**

$$SC = \frac{\sum \frac{m_i f_i}{N}}{\sum \frac{m_i f_i}{N}}$$

**Equation 2:**

$$S = 0.11 \int_{0}^{24 \text{Bark}} N^1 g(z)dz \int_{0}^{24 \text{Bark}} N^1 dz \text{acum}$$

Equation 1: Brightness as a function of spectral centroid SC, where $N$ = number of partials, $m_i$ = magnitude of partial $i$ and $f_i$ = frequency of partial $i$.

Equation 2: Sharpness in acum from Zwicker and Fastl (1999). The scale, Bark, represents the 24 critical bands of human hearing. $g(z)$ is a weighting factor emphasizing the upper critical bands. $N^1$ is the ‘specific loudness’, or level as a function of critical band rate, weighted by $g(z)$, which increases in the upper critical bands (16 to 24 Bark).

Such metrics are important for signal-processing engineers developing tools for psychoacoustic metering, for example, the Sonic Visualizer toolkit (Cannam, Landone and Sandler 2010). Eventually such tools might allow for ‘musician friendly’ signal processing devices with timbral controls. Returning to the guitar recording example, a full set of timbral metrics could ultimately lead to a guitar signal processor with controls marked ‘bright, heavy, clear, warm’ instead of ‘high, drive, mid, low’. The next section examines timbral attributes and acoustic measurements in examples from metal recordings.
DEVELOPMENTS IN DRUM PROCESSING

‘Modern metal drums are doctored, unauthentic and over-processed, yet this style of engineering is the standard’ (Neely 2012).

Close microphone techniques, whereby individual drum shells are each captured discretely in the recording studio by dedicated microphones, combined with DAW recording, offer a huge range of processing options to the metal producer, including:

- Basic editing. This usually comprises the selection (by ear) of subjectively ‘well performed’ passages and the subsequent duplication of these sections throughout a recorded performance.
- Audio quantization (making discrete time/phase adjustments so that all drum hits fall on a precise timing grid, usually a strict division of the overall tempo). Beat Detective is one of the best-known audio quantization tools, developed by Avid as part of their popular Pro Tools DAW.
- Sample reinforcement, where close-mic’d drum recordings are augmented with pre-recorded drum samples. ‘Close-mic’d’ refers to a recording technique whereby each instrument has an individual microphone placed extremely close to the sound source, in this case, a microphone placed inside the shell of a kick drum, usually facing the beater of the kick pedal.
- Total drum replacement – similar to sample reinforcement, yet without utilizing any of the original recorded drum parts.

The range of tools involved in such work raises questions as to why so much adjustment is required. Mynett surmises that ‘due to the fast kick-drum patterns involved (double kick-drums or pedals are a prerequisite) and the intricate nature of the drum parts, it is normal that the drum tracks heard on a finished production are not entirely as performed’ (Mynett 2009a).

It is fair to say that these techniques are hardly exclusive to the world of metal production, in fact sample replacement and layering are common across many other genres, including pop, rock and some types of dance music. However, these genres tend to use such processes solely for sonic adjustment, rather than for the additional performance enhancements that they also afford metal drum productions. As such, it is true that virtuosic drum performances can often define metal releases, in some cases as much as their lead guitar counterparts (Christe 2010; Gross 1990). Many of metal’s subgenres have ‘celebrity’ drummers. Names like Dave Lombardo, Mike Portnoy, Chris Adler, Lars Ulrich, Joey Jordison, Tomas Haake, etc. will often feature in the metal press, including readers’ polls (Rosenberg 2012; Shaw 2013). The question of whether the metal audience is unconcerned by, or even oblivious to, the authenticity of such artificially created performances remains difficult to answer, especially as the subject is given so little coverage in the mainstream metal press.

Nevertheless, these ‘superhuman’ drum performances, and the production techniques used to derive them, suggest that either the audience is unaware of these processes being used, or that the issue of authenticity is less significant to them than the overall production standard and the complexity of performance that such techniques afford the genre.

Figures 1–3 show the differing spectral and temporal features a sampled kick-drum trigger gives in comparison to a traditionally recorded, close mic’d kick drum in Figures 4–6, taken from a multitrack session by producer Andy
Sneap (whose production credits include Machine Head, Testament and Megadeth). The sampled kick drum was provided by Roadrunner metalcore act DevilDriver, and is used as a complete drum replacement for real kick-drum timbres. Spectrally, the sample is considerably sparser than the recorded kick drum, with a large peak in the 100 Hz region, a significant dip, with the remaining harmonic content from approximately 2.9 kHz through to 5 kHz. The recorded kick drum, in contrast, has a less pronounced peak at the first harmonic, and whilst it maintains a similar overall spectral curve, there is still a prominent harmonic at the ~1 kHz region, as well as further spectral content, before the main slope of upper harmonic content, which continues to show some content in a much higher region than the sampled kick-drum sound, including a prominent peak at 7 kHz.

One possible reason for the use of kick-drum samples with this type of frequency content (i.e. a reduced high end, a boosted first harmonic and a significant gap between the first harmonic and the ~3 kHz region) might be to allow ‘space’ in the spectrum of the finished mix for distorted bass guitar, and rhythm guitar textures to occupy (see the next section, ‘Timbral analysis of guitar tracking’).

Keith Andrews, a recording engineer, summarized the drum tracking and triggering process used on Heartwork (Carcass, 1993) as follows:

In the end, triggers were used, with samples coming from an Akai S1100. There were a bunch of options for the kick sound, but in the end...
Figure 3: A waterfall plot (FFT) of a drum trigger sample. Frequency is on the horizontal axis with amplitude on the vertical axis.

Figure 4: A waveform of a single ‘real’ kick-drum hit recorded in the traditional close-mic fashion.
Figure 5: A spectrogram of a single ‘real’ kick-drum hit recorded in the traditional close-mic fashion. Frequency is on the vertical axis, with amplitude indicated by ‘warmer’ colour.

Figure 6: A waterfall plot (FFT) of a ‘real’ kick-drum hit recorded in the traditional close-mic fashion. Frequency is on the horizontal axis with magnitude (amplitude) on the vertical axis. Time is shown on the depth axis.
we used one with not much ‘boom’ in it, since the original kicks on tape were pretty ‘strong’ and ‘deep’.

(Andrews 2012)

The timbral descriptors Andrews utilizes are worthy of note. Despite being a trained electrical engineer, rather than referencing particular spectral or temporal acoustic characteristics in his description of the kick-drum sounds, Andrews instead refers to cognitive analogies – ‘deep’, a visual metaphor, and ‘strong’, a tactile metaphor. This is common across the body of timbral research. When asked to describe a sound, a musician is likely to use a combination of direct and metaphorical language (Darke 2005), for example, ‘The bass is wooly’, ‘That passage is very baroque sounding’, ‘The kick drum is hard’, ‘The guitar is very harsh’ (Johnson and Gounaropoulos 2006). This kind of descriptive language is often a far cry from the quantitative approach used by acousticians (Grey and Gordon 1978; Grey and Moorer 1977; Schouten 1968), and this presents a shared obstacle for sound designers and signal processing engineers who seek to measure, and perhaps manipulate such perceptual characteristics (Schomer and Jehan 2001; Wessel 1979). Here, the recording engineer rests somewhere between these two worlds, yet seems to prefer the language used by musicians – perhaps this is an implicit requirement of the recording engineer, the ability to work with musicians on a day-to-day basis, translating their timbral language into technical actions.

TIMBRAL ANALYSIS OF GUITAR TRACKING

The guitar, in heavy metal, is used... as a rhythm instrument (technically, the guitar is a percussion instrument), as a solo instrument (often referred to as playing ‘lead guitar’, a misnomer in that heavy metal guitar solos generally follow, and therefore do not ‘lead’ [Friesen and Epstein 1994]). As forceful and popular as the modern metal drum performances captured on record can be, the enduring sonic characteristic most associated with the genre is probably the electric guitar. It is no surprise that the press that regularly accompanies lead guitarists is commensurate, with books and memoirs dedicated to the subject of the most popular metal guitarists (McIver 2008; Mustaine 2011). Modern recording technology allows for a great range of timbral variations to be captured. The most distinctive timbral characteristic of all metal guitar sounds is the use of distortion (Marui and Martens 2001).

A NOTE ABOUT GUITAR DISTORTION AND OVERDRIVE

Distortion (or natural overdrive) introduces harmonic content to the frequency spectrum of guitar or bass signals. Distortion and overdrive are often mistakenly used interchangeably. Overdrive results when an increase in input voltage gain means that the output capacity of a device is exceeded and is in itself a method for producing distortion. Distortion introduced by vacuum tube overdrive tends to introduce 1st and 2nd order harmonic distortion, whereas distortion introduced to the electrical signal by means of transistor overdrive often introduces 3rd order harmonic distortion. Hence a differing timbre is created spectrally by means of the newly introduced harmonic content at higher frequencies. Traditional equalization (amplification or attenuation of existing spectral content) cannot replicate these timbres, though there are various other ways to generate this effect. When
A guitar amplifier is fed more input signal than its tubes can take, the signal becomes compressed and additional harmonics are generated in the output (soft overdrive). If this process is taken too far, the whole waveform might be 'clipped', creating both even and odd harmonics in the amplified output. Transistor amplifiers do not tend to create the softer type of overdrive, as they are usually more efficient than tube designs across the operating range – though they will clip hard if driven beyond this. This signal chain is illustrated in Figure 7.

There are also other parts of the signal chain that a guitarist or producer could attack in order to create distortion – higher output pickups, effects pedals to increase voltage gain or emulate transistor clipping, preamplifier tube stages or even power amplifier tube distortion – but the most common of all is speaker distortion. Early distortion pioneers began deliberately damaging their loudspeaker cones to create new types of distorted guitar tone.

Figure 7: Approximation of signal flow from electric guitar to backline amplification. Blocks on the left of the figure have low voltage outputs, blocks on the right have high voltage outputs. Signal from the electric guitar is fed by unbalanced jack to any effects processors, such as distortion, delay, chorus pedals, etc. These are usually wired in series before the pre/power amplifier (although sometimes effects units might be hooked up via the effects send/return of the preamplifier). Preamplifier then feeds power amplifier, which boosts overall level to loudspeakers.
Ray Davies and the Kinks were famous for this, some time before the early heavy metal pioneers adopted distortion (Chappell 1999).

Recording practice has determined that the engineer can vary the amount of distortion achieved across the speaker cone by careful microphone placement – when microphones are placed by the edge of the cone, or off-axis to the centre of the cone, they pick up a much higher degree of speaker breakup (the physical process whereby a signal which cannot be reproduced completely by a fixed drive unit causes a ripple effect at the end of the cone), resulting in greater perceived distortion. The ‘classic’ way to record distorted guitar is often to use a capacitor microphone a few feet back, capturing the ambience of the room, along with one or two dynamic microphones closer to the speaker cabinet, before phase-aligning and balancing these channels. DAW recording has a practically infinite number of tracks, meaning that guitar parts can be overdubbed (or ‘double-tracked’) again and again if desired. If there is not a lot of double-tracking involved, a thick stereo sound can be achieved by hard panning the close microphones left and right (Chappell 1999; Fales and Berger 2005; Mynett 2009a).

It could be argued that the distorted tones favoured by metal guitarists have, to some extent, also shaped the musical content of metal insofar as the additional harmonics created by the distortion effects tend to cause timbral ‘mud’ if third notes are included; hence most metal riffs are based on the ubiquitous ‘power chord’, a chord with no major or minor thirds, comprised instead of a root note, fifth and octave.

AMPLIFICATION AND SPECTRAL ARTEFACTS

Guitar amplifiers use speaker sizes which force a de facto spectral filtering effect; 12” speaker cones, which are common in guitar speaker cabinets, do not allow for much sound replication above 3 kHz. Distorted bass often adds harmonics in the region of 800 Hz, meaning that careful spectral filtering is often used by producers and engineers in order to create space in the frequency domain of a finished metal mix (Mynett 2009b).

This idea can be extended to include the spectral profile of the drum triggers illustrated in Figures 1–3, by using equalization to attenuate the target frequency range from the sample set being used for drum replacement. Fast-Fourier Transformation (FFT) is a sonic analysis technique which uses a series of overlapping windows in the time domain to capture the spectral and harmonic content of a frame. Using an FFT analysis allows the studio engineer to pinpoint such harmonic content both quickly and accurately.

Advances in signal processing have led to a trend in amplifier modelling that makes use of convolution technology to impose cabinet or preamplifier spectral characteristics to cleanly recorded guitar signals (Gallo and Adviser-Selesnick 2010). Convolution provides a method of combining two signals to generate a new signal that contains features of both source signals. Recording a transient sound (a starter pistol, or a popped balloon) captures a room’s reverberant characteristics as an impulse response. The impulse response is then convolved with another audio signal to apply the reverberation of the first recording to the second. Impulse responses can be used to create a digital filter in a similar manner. Convolution can then be used to examine the periodicity of signals through auto-correlation, a process that finds similarities between a signal and delayed versions of itself. Different signals can then be compared...
through cross-correlation to find similarities. Thus, convolution-based signal processing techniques can be used to apply amplifier settings from a variety of guitar signal chains to a cleanly recorded input signal (Farina and Armelloni 2005). It is now not uncommon for guitarists in the metal community to share impulse responses on Internet forums such as guitarampmodelling.com, recabi.net, etc. With a huge variety of virtual amplifiers accessible quickly and at little or no cost, giving the ability to accurately replicate existing guitar timbres, it is perhaps unsurprising to see similar spectral characteristics becoming commonplace in many modern metal guitar sounds.

This amplifier modelling technology can be combined with another now common practice in modern metal production: re-amping, whereby clean guitar sounds are recorded, and then at the mix stage, sent out to real amplifiers to be recorded back into the DAW. Any latency introduced by this process can be eliminated in the DAW by simply time and phase aligning the re-amped signal with the original source, a routine process that would be almost impossible if recording onto analogue tape. Again, this means that as well as keeping expensive studio time and amplifier hire costs to a minimum, producers can replicate complicated signal chains and their associated timbres. Spectral analysis, shown in Figures 8–12, confirms these similarities in the overall spectral content in several modern metal productions.

Figure 8: The spectral plot of an isolated distorted guitar recording. Prominent harmonics at integer multiples suggest a distorted timbre. Note lack of significant content at >15 kHz – this analysis was carried out on a compressed audio sample (no CD quality version of the separate multitrack files was available).
Figure 9: A spectral plot of the full mix, using traditional recording (no kick triggering or convolution guitar amplifier modelling), including the guitar part shown in Figure 8. Overall there is an approximately 1/f shape, with strong spectral content in the 1.2 and 3 kHz regions. Note the lack of significant content at >15 kHz – this analysis was carried out on a compressed audio sample (no CD quality version of the separate multitrack files was available).

Further to these basic acoustic analysis techniques, there are various methods of perceptual analysis which have been adapted to the analysis of guitar timbres, distortion-based effects, and the codification of guitar timbre descriptors such as sharp or heavy (Atsushi and Martens 2005; Martens and Marui 2005). ‘Heavy’ as a descriptor can function in a number of ways, and is not necessarily related to guitar timbre alone, yet it is documented as the most common descriptor for guitar timbre in relation to metal (Fales and Berger 2005) and has been correlated to a number of acoustic features, most commonly the ratio of noise and harmonic content to fundamental frequency, and partially to the dynamic envelope. Any discussion of perceptual characteristics in metal guitar performance would be incomplete without considering pitch. Pitch is by its definition distinct from timbre, yet there has been some research which considers the two as intrinsically related (Houtsma 1997; Krumhansl and Iverson 1992).

Alternative tunings, including the use of 7-string guitars, have become common in some metal subgenres. The spectral characteristics that accompany such tunings include a correspondingly lower range of fundamental
frequencies. At lower frequencies, pitch becomes more difficult for listeners to accurately discriminate (Schneider 2001), meaning that the additional harmonics generated as a product of distortion can help to provide the main ‘cue’ as to the actual pitch content in a passage. This principle is somewhat analogous to an ‘aural exciter’ signal processing effect, which generates additional (or phase aligns existing) upper harmonics in order to both increase perceived brightness and clarity of pitch (White 1997). In other words, producers can utilize distortion characteristics that might help listeners to differentiate the otherwise unclear pitch characteristics that using particularly low tunings cause. In the metal world, this is perhaps most pertinent to low-tuned 5-string bass parts, where adding distortion to the bass guitar has now become a common practice in modern metal production to help the bass ‘cut through’ the mix (Mynett 2009a). There is also some evidence that an idealized spectral ratio of 1/f exists across musical genres (Levitin, Chordia and Menon 2012), confirming informal observations in Figures 8–12 which show marked similarity in the overall spectrum of the various different sound files, albeit with a presence

Figure 10: Showing a spectral plot of a discrete production, full mix, which utilized both kick triggering and convolution guitar amplifier modelling techniques. There is a similar overall shape but with a significantly increased frequency boost in the lower/sub-bass region of 25–100 Hz. Similar to Figure 9 there is a prominent spectral content at the 3 kHz region, with a slightly more pronounced roll off in the 5–15 kHz region. Note the lack of significant content at >15 kHz – this analysis was carried out on a compressed audio sample (no CD quality version of the separate multitrack files was available).
peak in the very low frequency range, and around the 3 kHz spectral centroid peak that we can attribute to distorted guitar processing (whether ‘real’, i.e. recorded using a real amplifier, or modelled by convolution).

LOUDNESS AND SPECIFIC LOUDNESS

Loudness, both as an aesthetic issue within the genre (Christe 2010), and a technical issue amongst recording engineers and psychoacousticians (Deruty 2011; Genuit, Sottek and Fiebig 2009), is a well-developed concept, with several formulas for its calculation, and practical instruments for measurement in existence. Loudness is a familiar and fondly thought of characteristic for the heavy metal audience and its scholars, as Weinstein points out ‘Heavy metal’s loudness is not deafening, irritating, or painful […] but empowering’ (Weinstein 2009). Therefore it is reasonable to suggest that loudness is a perceptual measure that requires less by way of introduction than timbre.
Metallica’s 2008 release, *Death Magnetic*, brought the issue of the ‘loudness war’ – already a popular topic amongst the hi-fi and audio engineering community (Deruty 2011; Vickers 2010, 2011) – into greater perspective for the metal audience (Michaels 2008; Paul 2011). In the mastering process usually undertaken before pressing and release, dynamic range compression is used to reduce the range between the loudest peaks and the quieter passages of music, using automated gain reduction algorithms, either in hardware processing, or more recently, in multiband DAW-based signal processing algorithms, allowing for different amounts of dynamic range reduction in specific frequency ranges. This reduction in dynamic range then allows for additional gain to be applied, increasing the perceived volume of the finished product, regardless of peak amplitude. This drive to be perceptually louder has been well documented by audio mastering engineers (Katz 2007; Vickers 2011). There appears to have been some response from the metal community as a backlash against these increasingly compressed releases. Earache Records founder, Digby Pearson, recently blogged about their policy stating that the ‘label had succumbed to the loudness war’ in the 1990s, before resolving to master music with the preservation of dynamic range in mind with more...
recent releases (Pearson 2010). Nevertheless, loudness and metal remain inextricably linked for many listeners: The loudness of metal music is arguably one of its strongest defining characteristics, recognizable both to metalheads and those outside the metal scene (McKinnon 2009).

Psychoacoustically, loudness presents an interesting perceptual characteristic in that it is dependent on more than amplitude alone – a phenomenon that might come as a surprise to the casual listener. Loudness is in fact a product of amplitude and frequency. A series of equal loudness contours were developed (Fletcher 1934) and subsequently revised (Robinson and Dadson 1994).

Figure 13: Spectrogram of Mouth for War. Time is shown on the horizontal axis, with frequency on the vertical axis from 0–20 kHz. The ‘intensity’ of colour indicates amplitude (brighter colours indicate louder signal at a given frequency/time).

Figure 14: Waveform of Mouth for War.
Figure 15: Waterfall FFT plot of Mouth for War excerpt. Frequency is on the horizontal axis with magnitude (amplitude) on the vertical axis. Time is shown on the depth axis.

1956; ISO 2003) and have since been refined to give a good understanding of perceptual loudness, though loudness metres based on these theories rely on test signals which are not so readily adapted to the measurement of loudness in musical stimuli. Specific loudness is an adaptation of loudness in octave bands (Stevens 1955) or to the critical-bands in the human hearing range (Zwicker 1961; Zwicker and Fastl 1999), which are more sensitive at upper frequency bands to perceived sharpness. This allows for more sophisticated metering of perceived loudness that can incorporate the masking effects of noise in nearby critical bands, and incidentally forms part of the perceptual encoding routine utilized in MP3-type lossy compression algorithms (in lossy compression, some data is irrevocably lost by the compression process, unlike the approach taken by lossless audio compression such as the FLAC codec). This method of measuring specific loudness (DIN 45631/A1) is capable of measuring time-varying perceived loudness in noisy signals and signals which incorporate noisy and harmonic timbres, and is therefore the most applicable method to the systematic analysis of loudness in metal (and other musical genres), though there is still much work to be done in the field of standardization of perceptual loudness measures for music (Fastla, VölK and Straubinger 2009; Genuit 2010).

Figures 13–24 show spectrograms, waveforms and spectral plots of four releases illustrating these changes in spectral content and loudness: Mouth
for War (Pantera, 1992), and from the same group’s later release, Hellbound (2000); Heal My Wounds (Corrosion of Conformity, 1994); and a more recent release, Captive Bolt Pistol (Carcass, 2013). The comparative differences between the 2000 and 2013 releases are less marked, both having been recorded in the DAW era. The entirely tape-recorded 1992 release by Pantera shows a significant decrease in average and peak loudness, clearly illustrated in the waveform. The overall spectral shape of the 1994 release which was likely recorded on tape but with digital editing and mastering involved in the chain, is somewhere in between that of the more recent recordings – the loudness is more comparable to the newer releases than

Figure 16: Spectrogram of Clean My Wounds. Time is shown on the horizontal axis, with frequency on the vertical axis from 0 to 20 kHz. The ‘intensity’ of colour indicates amplitude (brighter colours indicate louder signal at a given frequency/time).

Figure 17: Waveform of Clean My Wounds.
Figure 18: Waterfall plot of Clean My Wounds excerpt. Frequency is on the horizontal axis with magnitude (amplitude) on the vertical axis. Time is shown on the depth axis.

Figure 19: Spectrogram of Hellbound. Time is shown on the horizontal axis, with frequency on the vertical axis from 0 to 20 kHz. The ‘intensity’ of colour indicates amplitude (brighter colours indicate louder signal at a given frequency/time).
the 1992 release, but the boost in the sub-bass frequencies and the presence peaks at 3 kHz are absent.

Even a rudimentary glance at the waveforms of these (relatively main-
stream) metal releases from the past two decades reveals the trend: a decrease in dynamic range in the waveform, with spectral characteristics also showing an increase in both high frequency content, and sub-bass content. The acous-
tic fingerprints left by these new production technologies are clear enough to merit the consideration of a timbral toolset for dedicated analysis of the production techniques and the genre.

**Figure 20:** Waveform of Hellbound.

**Figure 21:** Waterfall plot of Hellbound excerpt. Frequency is on the horizontal axis with magnitude (amplitude) on the vertical axis. Time is shown on the depth axis.
Tracking timbral changes in metal productions...

Figure 22: Spectrogram of Captive Bolt Pistol. Time is shown on the horizontal axis, with frequency on the vertical axis from 0 to 20 kHz. The ‘intensity’ of colour indicates amplitude (brighter colours indicate louder signal at a given frequency/time).

Figure 23: Waveform of Captive Bolt Pistol.

TOWARDS A TIMBRAL TOOLSET

Production is a creative process, and sonic analysis of production techniques cannot easily be reduced to an exact science (Berger 2011; Bigerelle and Iost 2000). Nevertheless, the short analyses of acoustic characteristics presented here, combined with technical literature that suggests the artistic application of digital production techniques for the deliberate manipulation of perceptual sonic characteristics is common in current metal production.
Figure 24: Waterfall plot of Captive Bolt Pistol excerpt. Frequency is on the horizontal axis with magnitude (amplitude) on the vertical axis. Time is shown on the depth axis.

The tools to shape these perceptual characteristics are well documented in popular press dedicated to music production and technology (dynamic range manipulation, spectral enhancement, convolution modelling, sound replacement and so on), though the link between these tools and a robust perceptual analysis method has not yet been recognized or exploited. A ‘timbrometer’, using acoustic metrics to quantify perceptual attributes, would be a useful tool for the musicologist carrying out sonic analysis of production style, as well as being a useful tool for the current generation of home recording enthusiasts, of which there are undoubtedly many working in metal and its subgenres. In order for a ‘metal timbrometer’ to be effective it would need to be calibrated by an approach combining a verbal elicitation protocol with a multidimensional scaling analysis that allowed various, potentially overlapping, acoustic correlates to be extrapolated from a waveform, spectrogram or other acoustic representation. This would give a tool not dissimilar to existing loudness metres, with applications in automated or computer-aided mixing and mastering, as well the aforementioned musical production analysis.

CONCLUSIONS
Music production has evolved radically with the advent of the DAW, and despite little change in the basic instrumentation employed by the genre, metal is no exception. Technical innovations in digital signal processing...
techniques have been exploited to increase perceived loudness in the well-documented ‘loudness war’, with commercial artists falling victim to some level of consumer backlash. Yet the subtly changing timbral characteristics, particularly related to the use of distortion on bass guitar, spectral brightness and ‘heaviness’ in guitar, and drum triggered sub-bass frequencies in kick-drum sound replacement, are less often addressed. Nonetheless, a small selection of acoustic analyses shows that these techniques are both commonly used and identifiable by their acoustic ‘fingerprints’ in current metal productions.

In psychoacoustic literature, timbral attributes can be correlated to acoustic fingerprints to give a fuller descriptive language for measurement and manipulation of musical timbres, and some work has been undertaken in the timbral measurement of specific guitar-related timbres for metal. Further work still exists in bridging this gap, expanding on existing tools for sonic and musicological analysis to develop signal processing applications for practical music production techniques, similarly to the existing loudness manipulation tools which are now commonplace in mixing and mastering studios.

REFERENCES


ASA (American Standards Association) (1960), American standard acoustical terminology: (Including mechanical shock and vibration), New York, NY: ASA.


Tracking timbral changes in metal productions...


SUGGESTED CITATION


CONTRIBUTOR DETAILS

Duncan is Research Fellow in Music with Artificial Intelligence, working on an EPSRC-funded project entitled Brain-Computer Interface for Monitoring and Inducing Affective States at the Interdisciplinary Centre for Computer Music Research (ICCMR) in Plymouth, UK. He holds a Ph.D. in Psychoacoustics and Signal Processing from the Institute of Sound Recording (IoSR), University of Surrey, and was previously a postdoctoral researcher (Wolfson College Oxford) and Music Technology Officer for the Faculty of Music, University of Oxford. Research interests include sound recording, electroacoustic composition, and perceptually motivated digital signal processing. Commercial productions include work for EMI, Polydor, DeWolfe, Naxos, OUP and NMC. Metal he has a soft spot for includes early 1990s death/grind, ‘1994-era’ commercial metal, doom, sludge and noise.