View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UWL Repository

Proceedings of the 2nd AES Workshop on Intelligent Music Production, London, UK, 13 September 2016

SONIC CARTOONS AND SEMANTIC AUDIO PROCESSING: USING INVARIANT PROPERTIES TO CREATE SCHEMATIC REPRESENTATIONS OF ACOUSTIC PHENOMENA

Andrew Bourbon and Simon Zagorski-Thomas

London College of Music
University of West London
{andrew.bourbon,simon.zagorski-thomas}@uwl.ac.uk

ABSTRACT

The history of recording is often characterized as a history of improving audio quality whereas the notion of *sonic cartoons* requires us to re-think it as a history of clarity and the creation of schematic representations. Using the ecological approach to perception and embodied cognition, we will consider the *invariant properties* of various acoustic phenomena as a way of developing a range of strategies for semantic audio processing. In this context, semantic audio processing refers to plug-ins that utilize semantic descriptors to control multiple parameters. An example of this might be Waves' Tony Maserati Signature Series which provide controls with labels such as 'thump' and 'snap'. The notion of *sonic cartoons* provides a framework for a much more nuanced application of this approach.

1. SONIC CARTOONS

One of the characteristics that differentiates humans from other primates is the development of representational systems, such as language and visual art, in which symbols are understood to represent some aspect of our experience. We would argue that music itself is a representational system but that is a discussion beyond the scope of this paper. What we believe is unarguable is the idea that recorded music is a schematic representation of an actual or constructed musical performance. In the very literal sense, when a microphone is used to transduce sound waves in a room into an electrical signal it does so in a schematic manner. Photography translates light into a schematic representation that can be used to recreate some features of the original perceptual experience in another context and sound recording does the same thing. In the same way, visual artists can create abstract images that suggest meaning based on their similarity to (and difference from) visual experiences of 'reality' and previous artwork. 'Abstract' sound recordings and electronic compositions are doing the same thing. We have developed the notion of *sonic cartoons* [1,2] to explore this idea of recorded sound as schematic representation.

1.1. *Invariant properties* and *affordances*

Using the ecological approach to perception [3,4], the neural theory of language and metaphor [5,6] and embodied cognition [7,8], this theoretical model utilizes Gibson's [3] notion of *invariant properties* and *affordances*. This is based on the idea that the knowledge structures of the mind are founded on the re-enactment or simulation of perceptual experience and that expectations (the potential *affordances*) of any given set of experiences become distilled down to a given feature set – the *invariant properties*. Thus, for example, my experience of sound in different types of physical space has connected a set of *invariant properties* such as reverberation time, pre-delay and the accumulation of bass frequencies with *affordances* such as potential visual perception and movement in different sized spaces. These connections will have cultural and emotional associations as well as the more universal associations that relate to the physiology of being a human in an earth-like environment. So, while we are all likely to be able to recognize the sound of a large stone-walled enclosure, some of us may make cultural connections to churches and others may not – and each of us will have a different set of emotional and experiential *affordances* that suggest meaning to us.

1.2. Recordings as schematic representations

The idea of *sonic cartoons*, then, looks at how recording practice can and has influenced our perception and interpretation of recorded sound through the manipulation of *invariant properties* in schematic representations. Thus, in addition to the history of changing dynamic and frequency range in recording, we can also examine the range of distortions to spatial audio that have been used. For example, Miles Davis' 1959 *Kind Of Blue* used screens to reduce the extensive reverberation of Columbia Records' 30th Street Studio in New York and then added chamber reverberation to the three soloists' microphone signals. In effect, the recording creates a schematic representation of a performance in a large space – adding a longer reverb tail on the three higher frequency range instruments and inhibiting the reverberation on the instruments with the lower frequency content (drums, bass and piano). The use of close microphone placement also allowed producer Teo Macero to provide some additional bass frequency level without the muddiness. Macero – and many other sound engineers and producers at that time and since – provided the *invariant property* of bass frequency build up that is usually the result of the slower decay of low frequency reverberation through the use of a different parameter – increasing the volume of the direct (un-reverberant and un-muddy) sound of bassy instruments. He created a schematic representation – the equivalent of a line drawing – that provided some features and not others and yet still manages to create the impression of an ensemble performance in a large space.

Another feature of this can be illustrated by Leonardo Da Vinci's Burlington House Cartoon – a charcoal and chalk drawing of Jesus, Mary, St. Anne and John the Baptist.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Our appreciation of this representation resides in the way that we can recognize both what is being represented and Da Vinci's skill in using the representational system. Although we don't have such a well-established cultural convention for praising sound engineers and record producers as we do for visual artists, we all recognize *Kind Of Blue* as being a good recording as well as a good performance. It has a clarity and an unreal stillness that wouldn't be possible in a live performance – and of course the audio experience is different through speakers than it would be in a concert hall. In short, just as we're very rarely 'fooled' by a film into thinking we're witnessing the reality of what is represented, we also recognize recorded music for what it is – and we can appreciate the qualities of a schematic representation of a musical performance as much as we appreciate film making, painting and photography.

2. SONIC CARTOONS & SEMANTIC AUDIO

Waves' Tony Maserati Signature Series plug-ins provide controls with labels such as 'thump' and 'snap'. This reflects a more universal problem of communication that recording and mix engineers face [9] – the use of non-technical and metaphorical language in discussions with musicians. The term semantic audio, particularly within AES circles, more commonly relates to research that works in the opposite direction – of extracting features and meaning from the raw data of audio files. However, the question of what kinds of sonic features would we expect when someone uses a semantic descriptor such as 'heavy', 'fat' or indeed 'thump' is equally interesting.

By exploring *invariant properties* and their *affordances* for interpretation, we aim to explore how controls with semantic descriptors that map onto a range of potential parameters (and multiple channels of audio) can be developed. While current examples, such as those released by Waves, are based on modeling signature processing techniques associated with well-known individuals, our aim is to explore more general processing techniques and the possibility that the *affordance* of a particular semantic descriptor might be achieved in a variety of different ways.

3. EXAMPLES

The audio examples will come from two key approaches to processing, focusing on both a channel based approach and bus based processing. In the channel based processing tools such as the Waves Maserati DRM will be explored, looking particularly at the thump and snap control. The examples will show the effect of this process and also break down the processes involved in creating the stated characteristics. Alternative methodologies for creating perceived thump and snap will be employed, with a range of elements including bass guitar processed with the same tools, looking at how thump is a different proposition on a less transient element than the drums.

The examples will also explore the idea of effort, undertaking processing both on drum overheads and also on a vocal, looking at how it is possible to increase the perceived performance effort of the artist recorded. Though a range of terms could be explored, the selected examples provide an insight into the often-complicated layering of processing required to represent the affordances of a given semantic descriptor.

The second set of examples come from bus based processing- the new Waves Scheps Parallel Particles plugin offers a set of parallel processing tools to be used on a channel, again featuring semantic descriptors such as air, bite, thick and sub. The Waves Manny Marroquin Tone Shaper similarly works though parallel processing. A multibus system has been developed inspired by the work of Michael Brauer that allows the manipulation of tone in both parallel and destination busses using descriptors such as thick, lift, punch, air and warm. Examples of the effect of these busses will be provided, demonstrating how these broad descriptors can be used on parallel busses to bring warmth, punch or air to varied materials simultaneously and without compromise to the original signal.

The audio examples mentioned in this section can be found at: http://www.uwl.ac.uk/sonic-cartoons

4. CONCLUSION

The aim of this paper has been to present some initial ideas about how the notion of *sonic cartoons*, which was developed as an analytical tool based on the ecological approach to perception and embodied cognition, can be marshaled in the development of practical production tools that utilize the notion of semantic audio.

While it is relatively simple to suggest that plug-ins should be developed that utilize 'natural language' rather than technical terms, the real challenge lies in the problems of defining how terms such as 'air', 'heavy', 'fat' or 'thick' might be translated into processing strategies. Obviously, these terms are highly context dependent and the challenge lies either in developing much more sophisticated analytical tools to guide these choices or in using these types of tool as a 'rough shaping' device that cuts down the time that experts need to spend on mixing and allows them simply to finish a mix off with any necessary refinements.

5. REFERENCES

[1] A. Bourbon and S. Zagorski-Thomas, "The ecological approach to mixing audio: Agency, activity and environment in the process of audio staging," *J. Art of Record Production*, vol. 11, 2016.

[2] S. Zagorski-Thomas, "Sonic Cartoons," *Sound As Popular Culture*, 2016.

[3] J. J. Gibson, *The ecological approach to visual perception*, Psychology Press, 1979.

[4] E. F. Clarke, *Ways of listening: an ecological approach to the perception of musical meaning*, Oxford University Press, 2005.

[5] J. A. Feldman, *From molecule to metaphor: a neural theory of language*, MIT Press, 2008.

[6] G. Lakoff and M. Johnson, *Metaphors we live by*, University of Chicago Press, 2003.

[7] G. R. Semin and E. R. Smith, *Embodied grounding: social, cognitive, affective, and neuroscientific approaches*, Cambridge University Press, 2008.

[8] M. Iacoboni et al., "Grasping the intentions of others with one's own mirror neuron system," PLoS Biol, vol. 3, no. 3, 2005.

[9] Y. J. Lee "How can actor network theory and ecological approach to the perception be used to analyze the creative audio mixing practice?" AES Convention 140, 2015.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).