EXTENDED ABSTRACT:
RELATIONSHIPS BETWEEN ACOUSTIC FEATURES AND PERCEPTUAL SEGMENTATION OF MUSIC AUDIO

Roger T. Dean

MARCS Institute, University of Western Sydney, NSW 2751 and austraLYSIS, Sydney, Australia
roger.dean@ uws.edu.au

ABSTRACT

For a successful practical sonification, boundaries where the information state changes need to be made readily apparent perceptually. I describe time series analysis techniques for the detection of segments perceived implicitly during continuous music audio. The detected segments are compared with those measured computationally, or determined musicologically, on the input audio. The degree to which perceptual segmentation can be predicted is discussed, together with some of the factors apparently responsible. This may give useful cues as to how best to structure sonifications for informational purposes.

1. INTRODUCTION

It is well known that non-specialists' perceptions of musical structure do not necessarily coincide with those of musicologists. Thus in recent work we have analysed continuous perceptions of change in musical sound by non-specialists, and shown them sometimes to be predictive of continuous perceptions of affect in models made by techniques of univariate and multivariate time series analysis. Here we discuss the relationship between perceived change and perceived structural segmentation, and the factors that may influence the perceived segmentation. The purpose here is to identify acoustic factors which may be useful in creating perceived segmentation, and hence in information display by sonification. Note that the work does not concern segregation of simultaneous or overlapping streams within continuous audio, as pioneered by Bregman [1], but rather perception of successive major structural boundaries.

2. PERCEPTION OF CHANGE IN MUSICAL SOUND

In our experiments, untrained listeners hear a sound file unfold and continuously represent the degree of 'change in the sound' by scrubbing with a computer mouse [2]. The degree of scrubbing represents the degree of change they perceive, and they do not move the mouse at all when they perceive no ongoing change. Data are recorded at 20Hz, and according to the particular purpose, analysed after resampling to 2Hz or 5Hz.

The task does not overtly concern the identification of major structural boundaries, such as have sometimes been studied [3], because this latter approach inevitably brings direct 'demand' characteristics into play. Our approach is somewhat more implicit, from an experimental perspective; and hence may reveal features that are strongly penetrative perceptually.

We find that in many situations the continuous change time series is amongst the predictors of the listeners' perceptions of affect expressed by music, notably arousal and valence measured continuously using a 2D-interface, around which the participant navigates continuously. The change predictor normally models in conjunction with acoustic series such as those of intensity and spectral flatness, and we have also shown that acoustic intensity at least may be causal and not just predictive in a correlative manner [4-6], operating through loudness perception.

The perceptual change time series can be segmented by so-called 'changepoint' analysis, which detects whether the overall series is best described by a single model, or by several segments with distinct model predictors and/or parameters. We primarily use the R changepoint package, written by Killick and Eckley [7], which is currently v 1.04 in R 3.02. Changepoints can be determined by a principled statistical probability or by a penalty system akin to the application of information criteria. The boundaries are usually most effectively detected on the basis of changes in the variance and/or mean of the perceptual response time series. In the case of musical series, we generally find that musicologically defined structural boundaries are also perceptual boundaries or very close to them; but we find additional perceptual boundaries, and some disparities [8, 9].

3. FACTORS INFLUENCING PERCEPION OF SEGMENTATION

Some experimental work has directly assessed the ability of listeners to detect segments in electroacoustic sounds, and it indicates that some acoustic factors may be asymmetric in their effects on this process [10, 11]. The features involved require further study, and so a focus of this poster will be to illustrate factors that predict the perception of segmentation. These include major peaks in acoustic intensity, changes in spectral range, entries of specific kinds of instrumental or vocal timbres, and both action-driven changepoints (e.g. the entry and exit of a concerto soloist) and some driven by compositional structure (e.g. melodic change).

So we consider first the factors (such as acoustic features) that predict the continuous change in a time series analysis model. Such a model assumes that the whole of a sound file represents a process that can be considered as one. The data are stationarised and the analysis deals with the whole. On the other hand, the changepoint segmentation then distinguishes models representing the segments. So second, we ask whether the model of a continuous change perception series contains the same segmentation points that the audio/musical data series does? And in either case, what factors seem best to predict the audio segmentation points, a widely studied issue potentially distinct from the less studied question, which best predict the perceived change series? The identification of such implicit perceptual segmentation predictors might provide useful information for the design of sonifications intended to facilitate data display and comprehension [12], since this commonly would require the identification of successive segments in the ongoing data stream.

4. REFERENCES

[1] Bregman, A.S., *Auditory Scene Analysis*. 1994, Cambridge, USA: MIT Press. 773.

[2] Dean, R.T. and F. Bailes, *Time Series Analysis as a Method to Examine Acoustical Influences on Real-time Perception of Music.* Empirical Musicology Review, 2010. **5**: p. 152-175.

[3] Noble, K., et al., *Event Segmentation and Biological Motion Perception in Watching Dance.* Art and Perception, 2014. **2**(1-2): p. 59-74.

[4] Bailes, F. and R.T. Dean, *Comparative time series analysis of perceptual responses to electroacoustic music.* Music Perception, 2012. **29**: p. 359-375.

[5] Dean, R.T. and F. Bailes, *Modelling Perception of Structure and Affect in Music: Spectral Centroid and Wishart's Red Bird.* Empirical Musicology Review, 2011. **6**(2): p. 1-7.

[6] Dean, R.T., F. Bailes, and E. Schubert, *Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation.* PLoS one, 2011. **6**(4): p. e18591.

[7] Killick, R. and I.A. Eckley, *changepoint: An R package for changepoint analysis.* R package version 0.6, 2012.

[8] Dean, R.T. and F. Bailes, *Influences of structure and agency on the perception of musical change.* Psychomusicology: Music, Mind, and Brain, 2014. **24**(1): p. 103.

[9] Dean, R.T., F. Bailes, and J. Drummond, *Generative Structures in Improvisation: Computational Segmentation of Keyboard Performances.* Journal of New Music Research, 2014(ahead-of-print): p. 1-13.

[10] Bailes, F. and R.T. Dean, *Listener Detection of Segmentation in Computer-Generated Sonic Texture: An Exploratory Experimental Study.* Journal of New Music Research, 2007. **36**(2): p. 83-93.

[11] Bailes, F. and R.T. Dean, *Facilitation and Coherence between the dynamic and retrospective perception of segmentation in computer-generated music.* Emprical Musicology Review, 2007. **2**(3): p. 74-80.

[12] Worrall, D., et al. *SoniPy: the design of an extendable software framework for sonification research and auditory display*. in *13th International Conference on Auditory Display*. 2007. Montreal: ICAD, available online at http://www.icad.org.