

MODELLING SIMILARITY PERCEPTION OF SHORT MUSIC EXCERPTS

Daniel Müllensiefen

Goldsmiths, University of London, United Kingdom
d.mullensiefen@gold.ac.uk

ABSTRACT

There is growing evidence that human listeners are able to extract considerable amounts of information from short music audio clips containing complex mixtures of timbres and sounds. The information contained in clips as short as a few hundred milliseconds seems to be sufficient to perform tasks such as genre classification (Gjerdingen & Perrott, 2008; Mace, Wagoner, Teachout, & Hodges, 2012) or artist and song recognition (Krumhansl, 2010). The ability to extract useful, task-related information from short audio clips also has been shown to vary between individuals and this variability has been the basis for the construction of a sound similarity sorting test (Musil, El-Nusairi, & Müllensiefen, 2013) as part of the Goldsmiths Musical Sophistication test battery (Müllensiefen, Gingras, Musil & Stewart, 2014) where participants are asked to sort 16 800ms clips into 4 groups by perceived similarity. In this talk we will present data to explain the individual differences in the ability to extract meaningful information from short audio clips and to compare audio extracts on the basis of sound information alone. In addition, we will present two approaches to identify audio features of the short sound clips that drive listeners judgements. The first approach (Musil, El-Nusairi, & Müllensiefen, 2013) makes use of timbre features in combination with powerful statistical prediction methods to approximate listener judgements. In contrast, the second approach (Müllensiefen, Siedenburg & McAdams, in prep.) relies on Tversky's theoretically motivated model of human similarity perception (Tversky, 1977) to explain listener judgements and makes use of 22 spectro-temporal audio descriptors from the clips using the Timbre Toolbox (Peeters et al., 2011). Non-negative matrix factorization was employed to decompose the clips-descriptor matrix into a matrix of binary features which were then fed into Tversky's ratio models of similarity perception. Results show a superiority of the second approach using Tversky's similarity model that explains a higher proportion of the variance in the listener judgements and requires considerably less parameter tuning. The results are discussed in the context of psychological approaches to similarity perception which seem to apply well to the perception of musical sound.