Digital Music Research UK Roadmap
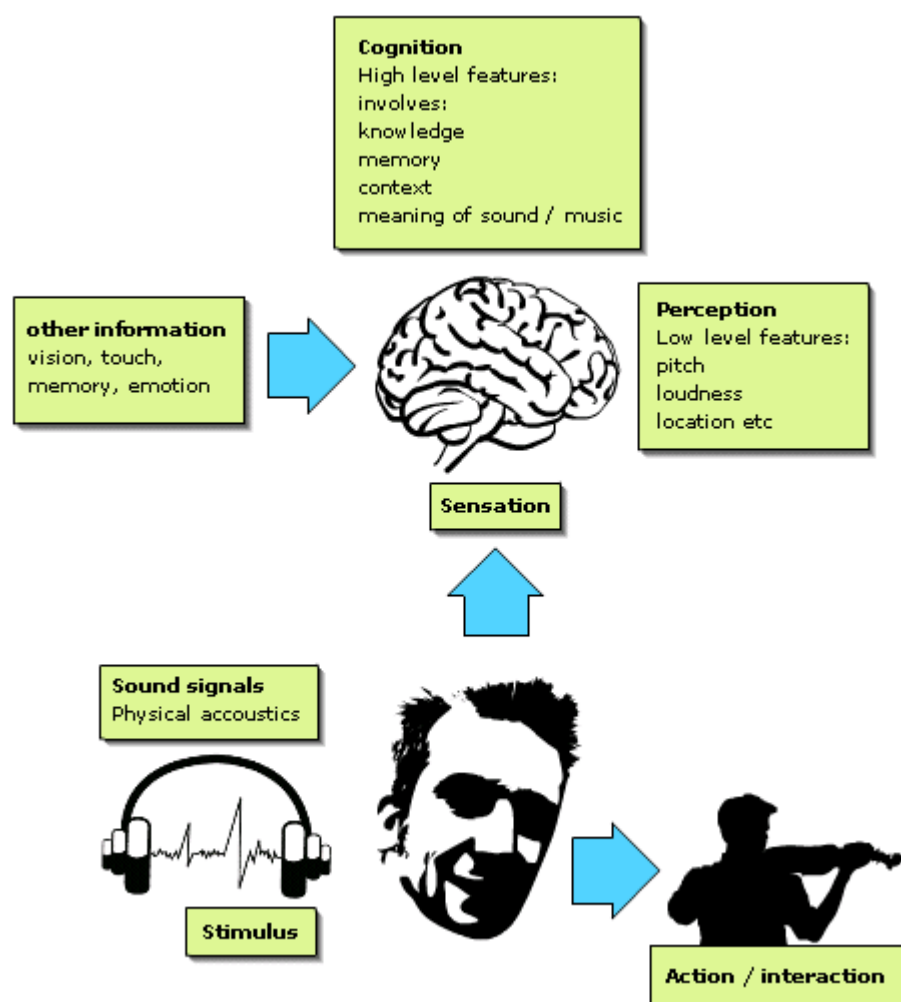Developed by the Digital Music Research Network

Research Goals in Digital Music: A Ten Year View

Machine Listening

Machine listening in digital music will form one of the primary themes of research over the next ten years. Essentially this involves the development of artificial devices and algorithms capable of analysing audio signals in such a way as to extract meaningful information about the music content and sound quality. It

includes the modelling of human perceptual and cognitive processes using computational approaches. This requires a trans-disciplinary approach, combining elements of acoustics, signal processing, audio engineering, musicology, practice-based research, expert systems and psychology/cognitive science.

Machine listening in digital music can be divided into a number of related themes, as shown in the diagram. Broadly, it can also be considered to incorporate both an interest in the musical objects, themes, gestures and genres implied by an auditory stream, as well as with the quality and attributes of the audio signals themselves (e.g. spatial and timbral attributes of auditory objects and scenes). Related themes in machine listening include audio and music **evaluation, analysis, understanding, representation, performance interaction,** and **description**.

Computational auditory scene analysis is a relatively young field that involves computational models in the process of analysing auditory scenes for their salient components. In musical terms this will involve the recognition of individual instruments, parts, themes and structures within the audio information representing a musical performance. This can be linked with forms of automatic music transcription and analysis, as well as with automatic performance-following and performance interaction systems designed to integrate machine performers with human ones. Automatic composing algorithms can also be integrated with machine music listening tools, to enable interactive real-time composition.

One long term aim in this field is to be able to develop machine listening applications capable of evaluating a number of discrete attributes of sound quality and predicting human responses accurately. Machine listening applications of this type can be used for automatic evaluation of audio products, processing algorithms, recording systems, acoustic spaces, sound reproduction systems and the like. The key issues to be addressed include developing satisfactory auditory scene analysis algorithms, cognitive models, test signals and statistical procedures that can be used to model the various human response processes. Research is likely to progress according to a so-called 'brown box' model - influenced by an understanding of human physiological and psychological mechanisms, but not necessarily attempting to model neurological processes accurately. A successful system might be regarded as one that produces results similar to those given by human listeners within a clearly defined context.

Machine listening can also involve a need to study emotional and context-dependent responses to musical sounds. Such factors are likely to be culturally and socially conditioned and finding a means of accounting for them is a key problem to be solved. This research presumes a relationship between a number of sub-attributes of musical sound and overall quality or preference evaluations. A key problem is that it is not yet known what attributes are most relevant in different evaluation scenarios, or what their relative weightings are. Neither is it

known to what extent human preference for music or sound quality can be modeled or generalised.

In relation to music and sound description, machine listening tools need to be developed with the aim of extracting salient descriptive information about the content of audio signals. This will be of considerable importance for the generation of metadata for classification and searching systems in the era of distributed on-line resources. Advanced music cataloguing and searching applications will then be possible, based on novel forms of querying that could be based on sounds rather than text.