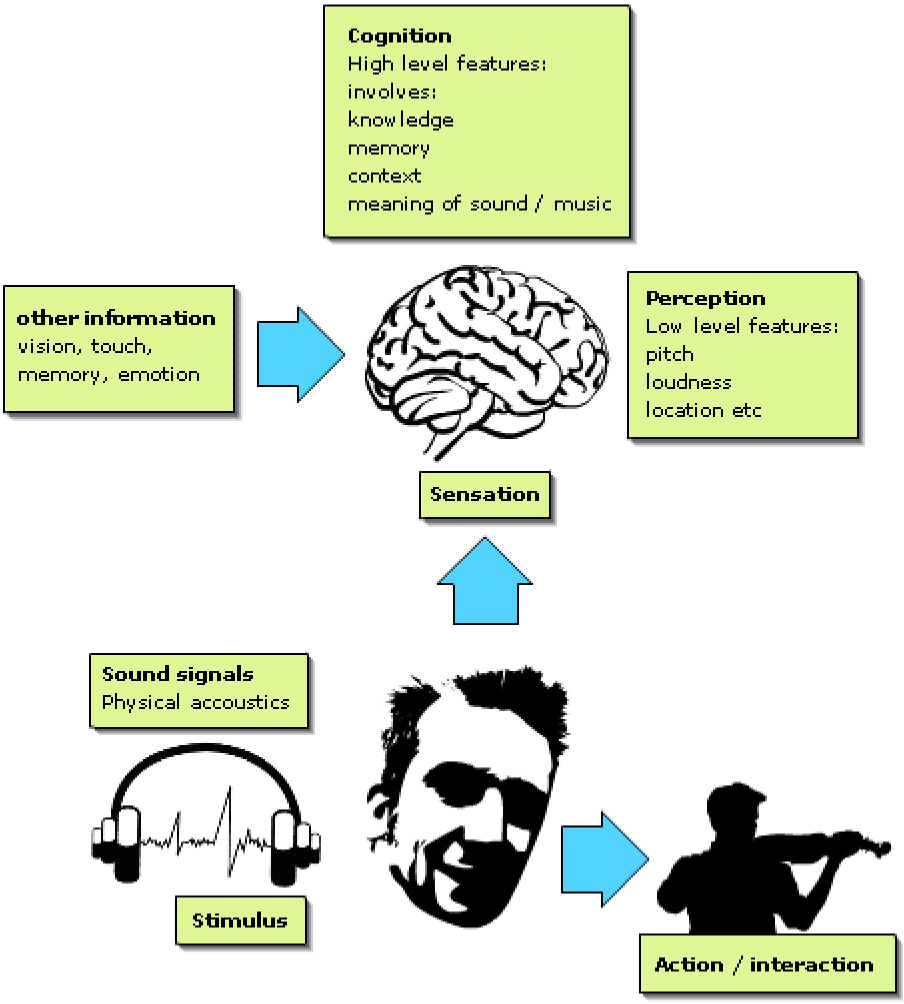
Research Goals in Digital Music: A Ten Year View

: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.

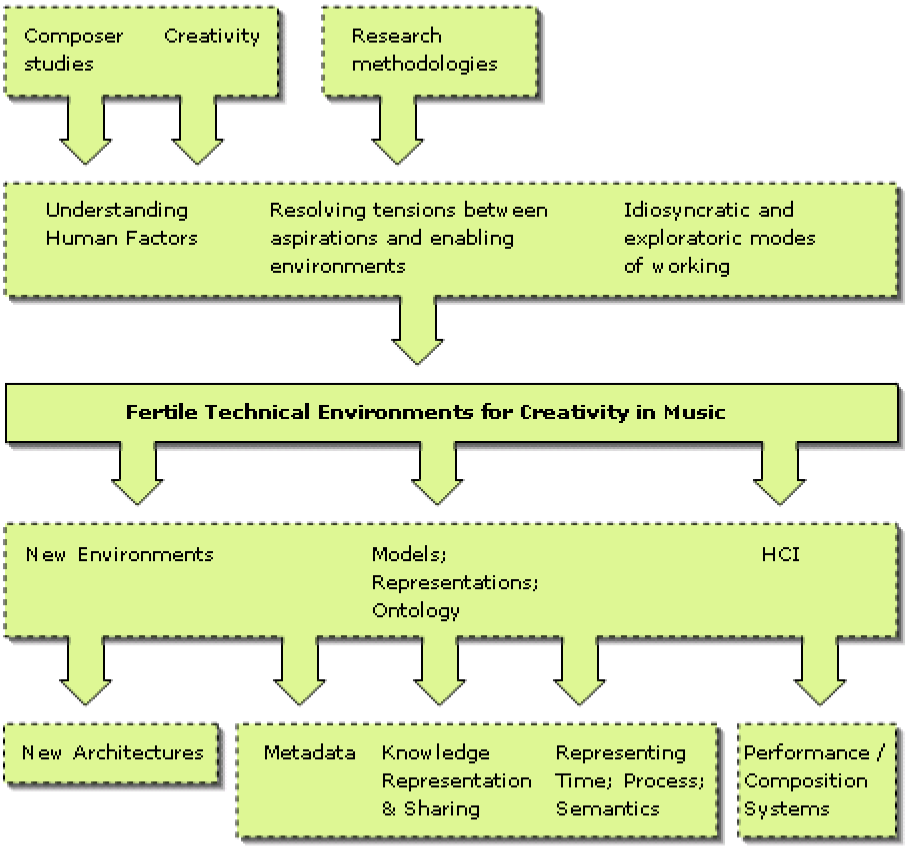
Machine listening can also involve a need to study emotional and context-dependent responses to musical sounds. Such factors are likely to be culturally and socially conditioned and finding a means of accounting for them is a key problem to be solved. This research presumes a relationship between a number 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.

Research Goals in Digital Music: A Ten Year View

Fertile Environments for Creativity

Experimental and commercial computer systems have been designed to allow composers and musicians to create musical compositions and performances using sound synthesis, digital signal processing algorithms and other musical tools. However, currently there is insufficient understanding of relevant human factors associated with the creative process. In particular, there is a need for research into the nature of creativity in digital music, such that the causes of tensions known to exist between composers and performers and the technological systems they use, can be better understood. Such knowledge will provide a sound theoretical basis for the engineering of technological systems which address these potential tensions, and thus provide more fertile environments for creativity in digital music.



Research into human factors in digital music is likely to have wider implication, since it will enhance understanding of many creative activities, and contribute to the research base to underpin the design of systems in other domains. For example, creativity is a critical driving force in science, commerce and manufacturing.

Currently, we lack research methodologies to elicit this knowledge, but over the next ten years inductive qualitative methodologies need to be developed to study professional composers and musicians at work. These methods are needed to complement those used in experimental psychology and the naturalistic study of creativity. They will better capture and analyse data to provide rich and holistic descriptions of creative activities in digital music.

An important research challenge is to establish a research base relating to **human factors** of composition and performance of digital music, as a basis for the engineering systems for creativity in digital music. This area is intrinsically transdisciplinary requiring expertise from perspectives of **music, information science, psychological** and **cognitive studies, computer science** and signal **processing**. Further, the problem domain is complex and diverse, involving idiosyncratic users with roles ranging through music production, creation, distribution and consumption.

The research must address two broad challenges, (i) to gain a richer understanding of human factors pertaining to digital music systems (represented in the top half of the diagram), and (ii) to utilize that knowledge to synthesise engineering principles for the development of improved systems (the bottom half of the diagram). In (i), objectives must be:

- To develop **research methodologies** to investigate human factors relating to creativity in digital music;
- through the application of these methodologies, to achieve a better understanding of such human factors. In particular, this will involve studies of composers and musicians "at work";
- better understanding of the causes of potential **tensions**, known to exist, between technological enabling environments (composition and performance systems) and their users (composers and musicians);
- to research the means of removing or reducing those tensions in order to make the **enabling environments** more fertile for creativity.

The second of the challenges concerns the production of technical solutions grounded in new knowledge about the creative process. Accordingly, the objectives in (ii) must include:

- The development of new **technological environments** to support creativity in digital music. This investigation must go beyond consideration of novel tool sets, so as to address the architectural contexts within which those tools are made available and the associated services provided to the users;
- to **develop models**, meta-models and ontologies for the representation of digital musical information and knowledge. Considerations should include the representation of

metadata such that information is machine understandable, and the representation of individual and community know-how. The models should also address challenging aspects of musical information, such as the representation and manipulation of time- and process-related information at the different levels of abstraction relevant to creative processes, ranging from the physical to the perceptual and aesthetic;

- to develop **human-computer interfaces** better suited to supporting creative activities in composition and performance.

Measures of success in this area of work will be the extent to which it brings about the following: **new and diverse musical forms** and **new products**, a greater **take up** of technologies within creative music genres; the development of **knowledge engineering technology** and repositories for accumulation and sharing of both individual and community know-how, thus establishing a "joined up" research base; and evidence-based theories and **explanations of creativity** in digital music, which can be applied to engineer technological enabling environments such that they provide fertile environments for creativity in digital music and other domains.

Research Goals in Digital Music: A Ten Year View

∶ Music Exploitation and Distribution

Recent advances in Digital Music Research point to many new and enhanced ways that music will be stored, distributed and used over the next 10 years.

This will include:

- Pervasive access to music
- Search and access tools
- Archiving and preservation
- High-resolution multichannel musical audio
- Coding and compression
- Musical content in research

The aim of pervasive access could be summarised as "Any music, anywhere, any time". The success of devices such as portable hard disk and MP3 players have demonstrated consumer appetite for portable music, and are in part a solution to pervasive access. Currently these devices require prior download and installation of music onto the devices. Recent mobile devices have provided downloadable access to commercial music, but even with bandwidths available via 3G, high quality real-time audio streaming is currently not cost effective. Full and flexible access to music from any location would require appropriate cost structures and efficient encoding, and new methods to search music catalogues effectively, manage rights, and interrogate self-published, non-commercial music. Research is also required to inform developments in compression technologies and improvements in audio quality at high compression ratios.

Advanced search and access tools are needed to allow users to find the music that they want. Textual labels and descriptions (metadata) are likely to remain important. Recent research in the visualization of metadata from musical collections points to new ways users can search through their own collection or remote collections of music. The goal of allowing the user to find new music they might like to hear, through methods and interfaces designed to extend the consumer's "senses" into the "space" of available music, will require research on many aspects of audio presentation and interaction.

Archiving and preservation of musical audio recordings is important to preserve recorded material for current and future generations. The National Sound Archive collects recordings issued in the UK, many donated by publishers, as well as selected overseas recordings. Efficient and effective use of such archives will need research into efficient compression and storage methods, and content analysis for metadata generation. New tools to allow visualization, searching and browsing through such archives will also be needed. Some of the requirements for these tools will be similar to tools which allow access to commercial music collections. Professional music researchers and musicologists will have special needs for search, visualization and analysis tools beyond those required for the casual music listener.

Increased media storage capacity will make high-resolution multichannel audio a practical commercial proposition in future. Such systems will require new research into recording and reproduction methods [see Musical Innovation], but will require further research into encoding and compression technology.

Efficient coding and compression of audio signals will continue to be required in the future, despite increased storage and internet bandwidths. New research towards object-based coding could provide high quality, low bit-rate musical audio coding, for example for cost-effective real-time streaming of music to mobile devices. Auditory scene analysis [see Machine Listening] could provide similar encoding methods for efficient transmission of high-resolution multichannel audio. Effective scalable audio coding systems will also be required to allow seamless transmission of musical audio content to automatically adapt to heterogeneous networks with varying capacities.

Digital rights management (DRM) is currently a major issue for music publishers, consumers and researchers, and involves complex and controversial discussions about fair and unfair use of artistic material. While current DRM systems may successfully restrict access to use by an individual person, they may also restrict use to a particular device, contrary to the expectation of many users. Interoperability is also an issue, with different music publishers and device manufacturers operating incompatible DRM systems. Research and development into DRM systems will require cooperation on a global scale, and may also need to reflect future changes in the structure and heterogeneity of the music distribution industry.

The use of musical content in research is also a difficult issue for researchers. In many research fields, experimental data is made available to aid comparison between results. This is not possible where the "data" used for experiments includes copyright musical recordings. New research methodologies are being explored to tackle this issue. These include the use of specially commissioned artistic material released under Creative Commons licences which allow free redistribution of the recorded material. For larger corpuses of existing copyright material, the submission of experimental algorithms to closed systems that perform evaluation without releasing the controlled content is already being explored. Closer links with music content rights owners may allow "research use" rights to be developed in this area.

Research Goals in Digital Music: A Ten Year View

Musical Innovation

Music created using computer technology has become a key expression of 21st century culture. Many people now have access to creative digital music tools through the widespread dissemination of computing and portable computing technology, the internet and software applications which enable musical creation (including software produced by the academic community over the past 30-40 years which is now in the public domain). These developments have resulted in a burgeoning contemporary music scene and new approaches to musical materials and musical composition.

Over the next ten years engineering research in the field of digital music should embrace evolving musical forms and new opportunities for musical performance and dissemination (including sound art installation and new forms of music distribution informed by developments in rights management, such as Creative Commons). This will ensure relevance to new aesthetics and cultural developments that have taken place in recent years - enabling researchers to seek solutions to contemporary musical problems and applications.

Many digital music research questions, which were hindered in the past by poor computational performance for audio applications, poor digital audio storage and non-real-time interactive audio applications, have now been solved. Current technologies can provide good digital audio storage capabilities and good performance for audio computation.

New research questions are now emerging related to sophisticated **human/computer interfaces**, techniques and strategies for the **control of information** about music and **creative musical processes**. Contemporary techniques to extract complex descriptive information from audio signals will lead to new approaches to the **organisation of musical materials** and their control within processes of musical creation and production.

The public performance of digital musics will also provide significant opportunities for research. Contemporary developments in human computer interaction, machine listening and mechanisms to integrate technology with human performers will be fertile ground for research. The **integration and application of existing technologies** in the context of human musical performance and **intelligent musical interaction** will be challenging problems. There are also significant opportunities to investigate new approaches to the **control of computer music,**

algorithmic composition and the application of computer algorithms for **live performance and improvisation**.

New approaches to musical performance will require developments in **music reproduction systems**, such as surround sound. This field seems likely to evolve to include the use of spatial encoding strategies which require very large numbers of audio channels and large numbers of loudspeakers for use in a variety of contexts (including the cinema and concert hall). Technological developments and research in **loudspeaker systems, spatial sound encoding strategies** and the **human perception of sound** will inform developments in this area.

Music performed with computers, or generated by them, has been shown to provide little opportunity for audiences to understand or engage with live performance or improvisation. Technical developments in these areas will benefit from close links to aesthetic and artistic research. New **methods of presentation**, presentation contexts and presentation sites (e.g. internet, public spaces or personal audio systems) could be enabled by further technological research and development.

Research in **human/computer interface** (HCI) technologies for music is an established field which finds application in music and other areas of HCI research. Research into computer mediated communication is also an emerging field. Developments in these areas will continue, related to advances in the electronic hardware capable of sensing human actions and to the mapping of sense-data to digital music processes.

It is appropriate for technological innovation to be applied in **many musical genres**. New technologies and musical innovations are not the preserve of 20th Century Avant Garde or experimental genres. New forms of digital and post-digital music (whose operating cultures are outside academia), contemporary jazz, pop, rock, fusion, ethnic musics and many other genres are fertile ground for research into the application of new digital music technologies. Engaging with new and emerging forms of music will allow technology researchers to contribute to a living and developing musical culture.

Communication and **collaboration between artists and engineers** can produce valuable research outcomes. A greater integration of disciplines and less rigid definitions of subject areas is required to move from the notions of multi-disciplinary work towards the concept of research that is inherently **transdisciplinary**. Problems can occur in communication between artistic, creative and scientific collaborators which has, to date, resulted in a slow acceptance of some technical innovations by musical communities. It will be important to establish dialogues and understanding between creative and technical innovators in music. Research and **education** in this area will be important to sustain research in the future.

Research in **human perception** of sound, and other human science disciplines, has a potential to impact significantly on creative digital music research and related engineering research. Digital music research should incorporate researchers working in these fields. Notably, sound has been shown to be significant in the development of key models of cognitive operation and to have provided insight in research into brain function (Kohler et al 2002). Important information related to physiology, cognition and brain science may result from transdisciplinary research activity in this area, in addition to human science data that will inform music technology innovations.

New methods of sound creation, generation and control have been at the heart of developments in digital music for the past fifty years. New aesthetics based entirely upon synthetic or artificial sound materials and their control, are now significant international creative cultures. **New musics** and new methods of **sound generation** will remain key areas of research and development.

Research is required to enable creators to **exploit musical data** in the complex, multi-parametric control medium presented by computer technology. Research is required in areas ranging from sound file management and the application of automatic music classification systems, to processes which allow the integration of information to inform, say, signal processing or creative processes.

Technology will continue to facilitate new approaches to the **composition of music**. Fields such as **algorithmic composition** and **performer interaction** with compositional rules will be fruitful areas of research.

Developments are also expected in the area of **live computer algorithms for performance**. These computer systems, which will interact strongly with musicians in both a supportive and a creative capacity, will be a fusion of ideas from algorithmic music, live electronics, cognitive science, artificial intelligence and performance activities. The challenge is to endow live algorithms with properties typical of human performance: strong interactivity, autonomy, innovation, idiosyncrasy and comprehensibility.

New forms of **musical notation**, its representation and the display of information about sound will create new tools for musical analysis and interaction with sound materials during the creation of music. Techniques for user interaction with notations, scores and visual displays will provide benefits to composers, performers, sound engineers, producers, musicologists and musical analysts.

It will be extremely important to research new engineering developments, technologies and methods within a **professional and contemporary musical context**. New technologies for composition and audio processing should integrate musical expertise at a professional level. A challenge for digital music research will be to engage with professional musical creators in a variety of innovative musical genres. It is important to ground research in this field through a focus on relevant problems. This will integrate knowledge accumulated by musicians about the nature, approach and methods of music, its creation, dissemination, performance and interpretation.

Creative outcomes can be developed to illustrate many engineering research developments in this field. This has many potential benefits and should be encouraged: it can illustrate research contexts, contribute to future engineering research developments, public awareness, dissemination and the industrial take-up of research.

Digital music research should attempt to reach out to **wider musical communities** where musical innovation is welcomed. Research in digital music should actively support the UK community of innovative professional music creators and performers.

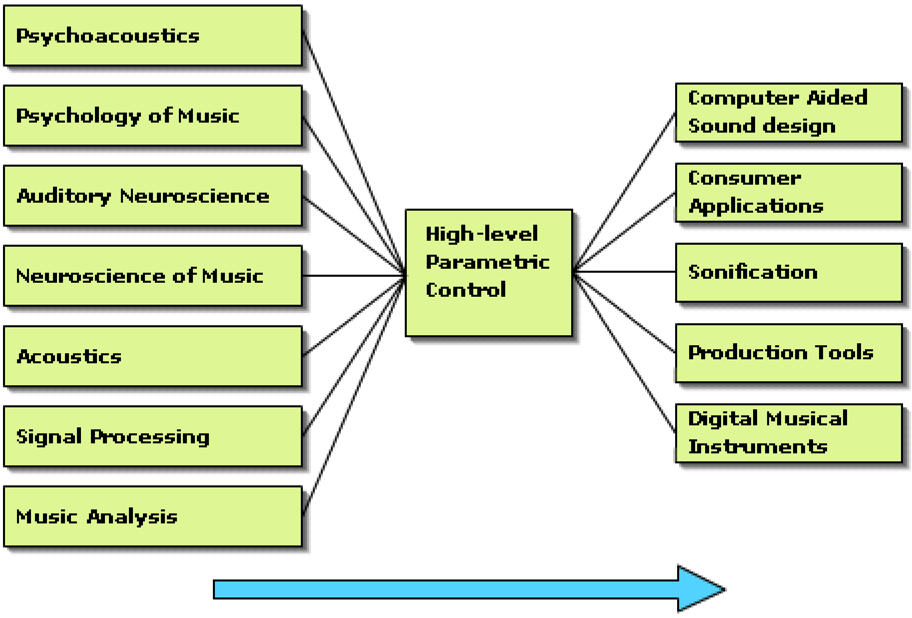
Enabling **musical innovation** is a key goal in digital music research. The creation of new musics from new technologies will allow digital musical research to make a significant contribution to 21st century creative culture and creative industries in the UK.

Reference

Kohler E., Keysers C., Umiltà M.A., Fogassi L., Gallese V., Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. Science, 297: 846-848, 2002

Research Goals in Digital Music: A Ten Year View

∴ High-Level Parametric Control



A number of powerful algorithms and systems for sound synthesis, digital signal processing and composition have been designed. However, in general these algorithms are either too complex (i.e., they have an overwhelming number of simultaneous control variables) or non-intuitive (i.e., their variables are not perceptually meaningful), or both. This limits the usability of such algorithms and systems.

As an illustrative example, consider a typical audio equalizer. This provides a large set of faders for modifying the spectrum of a sound.

Normally, an equalizer is used to modify the auditory quality of the music; for example to achieve a specific tone quality. The larger the number of faders, the more precise the equalizer control, but its operation becomes more complex, i.e. the number of possible combinations of fader positions increases. The current state of audio technologies could produce a high-level parametric equalizer with much fewer controllers, each for a particular aural response; e.g. a degree of "weightiness", a degree of "sharpness" of a sound, etc. These are not the equivalent of presets of fader positions, but proper controllers with meaningful variations within given perceptual frameworks.

The provision of high-level parametric control of sound and music is an important research challenge.

The key to progress is to develop effective methods which integrate cutting edge developments in audio and music processing technology with the ongoing research on auditory and music perception. This can only be achieved by integrating research and developments from:

- Psychoacoustics
- Psychology of Music
- Auditory Neuroscience
- Neuroscience of Music
- Acoustics
- Audio Signal Processing
- Music Analysis

In many applications of **Acoustics** and **Audio Signal Processing** it is necessary to know what humans actually hear and how. Sound, which consists of air pressure waves, can be accurately measured with sophisticated equipment. However, understanding how these waves are received and mapped into thoughts in the brain is not trivial. The field of **Psychoacoustics** studies the subjective human perception of a sound; that is the psychology of acoustical perception. But music is not a single isolated sound. Rather, music is generally defined as a composition of sounds. It requires the study of the subjective perception of organised sound: the **Psychology of Music**.

Although Psychoacoustics and Psychology of Music have contributed enormously to our understanding of subjective human musical perception, we still do not know what is happening in the brain when we listen to sound and music. The field of Auditory Neuroscience and the emerging field of **Neuroscience of Music** have started to address this question. Within the next 10 years, advances in these emerging areas will contribute to the development of new technology for Digital Music, most notably new models of machine listening, new techniques for coding audio and new approaches to the organization of musical material. Recent advances in **Auditory Neuroscience** research have already begun to inform the development of DSP chips modelled after the functioning of the auditory cortex.

It is important to emphasize, however, that such developments cannot happen in isolation, but rather as the result of **transdisciplinary** research integrating psychologists, neuroscientists, engineers and musicians.

This research will lead to the development of a new generation of sophisticated technology for:

- Computer Aided Sound Design
- Music Production Tools
- Sonification Methods
- New Digital Musical Instruments and performance tools
- Consumer Applications

Systems for **Computer Aided Sound Design** and **Music Production Tools** integrate a number of tools for manipulating sound samples (recorded sounds) or synthesizing sounds, or both. There are many techniques available for implementing such tools, most notably for sound synthesis. However, such systems can be difficult and non-intuitive to operate because they lack good means for high-level parametric control.

Sonification is the use of audio to convey information or perceive data aurally. Due to the nature of auditory perception, such as temporal and pressure resolution, it forms an interesting alternative to visualization techniques and provides good understanding of temporal information. Higher-level parameters for sound design informed by our cognitive ability will make a great impact on **Sonification Methods**, especially for auditory monitoring of complex simultaneous events.

Digital Musical Instruments are composed of a gestural controller and a sound generation unit. The gestural controller is the device that transduces the human inputs to the instrument. It is where physical interactions between the performer and the instrument take place. The sound generation unit involves methods of sound production and its controls. The relationship between the gestural variables and synthesis parameters is far from obvious and can be fruitfully varied in each musical application. A gesture may control a number of synthesis variables simultaneously. Research into high-level parametric integration methods is paramount for the design of **New Digital Musical Instruments** and augmented and extended instruments.

Finally, the example of the audio equalizer cited earlier is a good illustration of the impact of new developments in high-level control to **Consumer Applications**. Most current domestic Hi-Fi systems provide only three controllers for modifying the sound quality: volume, treble and bass. A new generation of "active" Hi-Fi systems will emerge from this research, where users will be able to control not only the audio quality of the music, but also the way in which is it interpreted; for example, the choice of playing back a piano piece in the style of different interpreters, and sophisticated search mechanism based on the musical attributes of audio recordings.

Research Goals in Digital Music: A Ten Year View

∴ Sound Reproduction and Recording

Among the most critical processes in the chain of digital music production and its reception are those of sound recording and reproduction. These are well advanced but there are still many potential developments in transducer technology, and in the capture and reproduction of sound fields. While these areas of research may seem to lie outside the digital musical domain, many sophisticated approaches to sound reproduction require digital control, processing or encoding systems.

Loudspeaker developments have occurred in many areas such as efficient loudspeaker/amplifier configurations, **directivity-controlled arrays, digital loudspeakers**, and the application of **ultrasonic parametric arrays**(e.g. HSS). Other developments in this area and new digitally controlled sound reproduction systems are likely to evolve over the next ten years. There is a requirement, by both professional and domestic users, for **loudspeaker systems** which are portable, light, capable of sustaining high sound pressure levels, multiple audio channels and high quality audio output.

Domestic sound reproduction systems have evolved in recent years. Surround sound systems are widespread, and DVD systems have led to surround being considered as a new hifi and audio standard. Existing methods of **surround encoding** vary. There are a number of approaches ranging from matrixed surround sound capture methods, to systems which provide a surround sound effect, often driven by a requirement for efficiency in the use of multiple loudspeakers and channels.

Future surround systems may use many more channels of audio and more loudspeakers or more efficient sound encoding strategies.

Several contemporary technologies exist which may be capable of improving **surround sound reproduction** so that it is capable of delivering more convincing three-dimensional immersive experiences. Wavefield synthesis, binaural head-related-transfer-function approaches, and Ambisonics are technologies which can be developed over the next ten years, but research into audio encoding strategies will rely upon complementary research in **human perception of sound, psychoacoustics, loudspeakers** and **listening tests**.

It will also be necessary to progress research into methods to **capture surround sound fields** efficiently in a variety of locations.

Audio recording systems are well developed. High sample rates, larger audio sample word lengths and good digital audio storage facilities are provided by contemporary systems. While it would seem that older research questions about computation and data storage are solved, multiple audio channels for surround sound and complex new control systems will continue to stretch **computational performance** and storage/delivery bandwidth (see Music Exploitation and Distribution).

The application of **high level parameters** in professional software **recording environments** and the integration of audio content tracking, through the application of MIR research outcomes and search technologies (as exemplified in the Semantic HiFi project), will be possible from the current research base.