PAN
ARCHIVES OF ACOUSTICS
Vol. **39**, No. 2, pp. 203–214 (2014)

DOI: 10.2478/aoa-2014-0025

Auditory Display Applied to Research in Music and Acoustics

Bożena KOSTEK

Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics
Narutowicza 11/12, 80-233 Gdansk, Poland; e-mail: bokostek@audioacoustics.org

(*received November 14, 2013; accepted April 24, 2014*)

This paper presents a relationship between Auditory Display (AD) and the domains of music and acoustics. First, some basic notions of the Auditory Display area are shortly outlined. Then, the research trends and system solutions within the fields of music technology, music information retrieval and music recommendation and acoustics that are within the scope of AD are discussed. Finally, an example of AD solution based on gaze tracking that may facilitate music annotation process is shown. The paper concludes with a few remarks about directions for further research in the domains discussed.

Keywords: Auditory Display, Music, Acoustics, Music Technology, Music Information Retrieval, Sonification, Music Annotation.

1. Introduction

Music and acoustics are two closely related fields. It may be said that music builds on acoustical bases. To name just a few – acoustics describes physics of music, it offers physical bases of musical instrument behavior. Musical instruments are developed according to unique instrument acoustical features based on excitation and radiation characteristics, resonance systems. Acoustics reproduces music by means of loudspeaker systems radiating sound, but it depends on human perception, another common feature linking both fields. Instrument performance depends on musical training, and music listening depends on perception (Beauchamp, 2011). Music Technology (MT), an important part of music, covers many areas related not only to Music Information Retrieval (MIR) but also to fields conceptually closer to Auditory Display (AD) such as musical interaction, performance modeling, music cognition, synthesis and digital audio effects. Blauert (2012) proposes to apply a perceptionist's view to problems and tasks in acoustics. This holds in particular, when the primary goal as for example in AD is not a physically authentic, but rather a perceptually plausible synthetic sound.

Auditory Display in the past referred mostly to systems which conveyed messages in the form of sounds or audio signals being the sole interaction mode (Kramer *et al.*, 1999) or it was used as navigation controlled by means of sound communication (Dobrucki *et al.*, 2010). Over the years, it has largely broadened its scope. It may reinforce or complement other modes of interaction such as visual display (Stewart, Sandler, 2012). It may also be said that AD may replace visual display, and such an example will be shown later on. Currently, much of the AD research focuses on spatial auditory display (e.g. such as externalized sound through headphones, ambisonics, wavefield synthesis, surround systems, 3D audio systems), at the same time putting great emphasis on understanding the functions of the human auditory system. On the other hand, enormous technological progress fostered rapid development of mobile communication and catalyzed the need for novel ways of interaction with devices such as smartphones, tablets, etc., thus creating new fields of application for AD. One may also look into papers of International Conference on New Interfaces for Musical Expression (NIME). There are examples in terms of common interests for both Music Technology and AD. Vamvakousis and Ramirez (2012) showed the Eye-Harp, an eye-tracking musical interface for controlling melodic, harmonic and expressive aspects of musical instruments in real time, have similar expressive potentials to a traditional musical instrument. In addition, many AD papers related to music could be found within the scope of International Conferences on Auditory Display (ICADs).

It is worth mentioning that a special issue of Journal of the Audio Engineering Society devoted to AD, has been recently published (Stockman *et al.*, 2012). Guest Editors of this issue, i.e. Stockman, Roginska, Walker and Metatla pointed out that apart from using sound to display data, monitor processes or support human interactions with systems and devices including augmented and virtual reality systems, AD encompasses also Sound and Music Computing, Haptic Audio Interaction Design, Audio Mostly, New Instruments for Musical Expression, and Interactive Sonification. All papers that belong to this special issue present a very broad scope of AD issues and applications (Stockman *et al.*, 2012).

The paper recalls a variety of applications of AD to music technology, music information retrieval and music recommendation. Also, some examples of the research performed or supervised by the author in the AD area are shown. Then, an example of AD that may facilitate music annotation process based on gaze tracking technology is more thoroughly investigated. This is a potentially unique application of gaze tracking exploring technology potential for improving music annotation. Another example of the AD application is the audio mixing system based on hand gesture recognition. Finally, some comments referring to main research challenges within music technology area that need assistance from other domains, also from AD are given.

This paper has been presented as a plenary paper at the 19th International Conference on Auditory Display (ICAD-2013) in Łódź, Poland (Kostek, 2013), then revised and extended for publication in Archives of Acoustics to acquaint its Readers with some of the AD applications to the music domain.

2. Auditory display in music technology and acoustics – research trends

As already mentioned, recent years have seen an outgrowth of AD archetypes since the domain inception. Clearly, there are a number of ways to describe the AD term, that is why some of them are recalled here, but it should be pointed out that in the literature there is no one formal definition of what AD is. In the author's opinion there's a reason all these definitions exist. First of all, as already mentioned, some early definitions outgrowth their archetypes, also the AD usage defines its name, for example it may ascribe sound resulting from transformation data into audio, an equipment that does this or a process encompassing all those. In addition, auditory display refer to the use of sound to communicate information, whereas sound reproduction points out in the direction of the way sound is rendered through loudspeakers, headphones or bone conduction as interpreted by Hermann (2008). Hermann characterizes auditory display as conversion of sound signals into audible sound, thus it encompasses also the technical means to create sound, such as for example loudspeakers, headphones or bone conduction headphones (Hermann, 2008). He also pointed out that the context of the user (user, task, background sound, constraints) and the designed application are essential for auditory display. Moreover, AD encompasses several other notions: sonification, audification and auditory interfaces. Sonification is an integral component within an auditory display system which addresses the actual rendering of sound signals (Hermann, 2008). Further, Hermann says: "Similar to scientific visualization, sonification aims at enabling human listeners to make use of their highly-developed perceptual skills (in this case listening skills) for making sense of the data." (http://sonification.de/son). According to the definition functioning in the AD area, information projected from an auditory display can be classified as direct or indirect sonification of data (Fernström, McNamara, 2005). Within sonification one may also differentiate between auditory icons (everyday sounds designed to convey information about events by analogy to everyday sound-producing events), earcons (nonverbal audio messages, short, structured musical phrases that can be parameterized to communicate information in AD) and spearcons (defined as brief sound cues created by compressing a text-to-speech sound file). They were invented as an alternative communication channel to graphical computer icons to convey information. Audification is meant as an auditory display technique for representing a sequence of data values as sound. To differentiate between these two terms, one may say that audification is like writing data directly to a sound file, and sonification is a more general notion that refers to the technique and the process, and can help especially in areas where a visual representation of the data would be overwhelming or difficult to interpret. The last term mentioned above, i.e. auditory interfaces, refers to means realizing sonification, thus this may be translated into: "acoustically rendered interfaces".

In this paper examples of sonification along with auditory interfaces applied to music technology are briefly reviewed.

2.1. Spatial Auditory Display

Blauert and Rabenstein (2012) say that: "It is one of the goals of audio technology to present sound fields to listeners in such a way that they experience an auditory perspective, that is, perceive auditory events in various directions and distances, which may then form complex auditory scenes" (Blauert, Rabenstein, 2012). To this end one should recall studies carried out in many research centers resulting in still not perfect but acceptable spatial sound systems transforming auditory scenes into sound that envelops the listener. Perhaps the most prominent paper within the AD area is the one by Shinn-Cunningham and Streeter (2005) that shows that sound source location can easily be manipulated in the sense that spatial information can be used to represent arbitrary information in an auditory display. More recent research study by Stewart (2010) recalled theory of spatial hearing along with sound field reproduction. She showed ways of optimizing binaural auditory display to be used in interfaces for music search and discovery. The term binaural is reserved for signals that are recorded (dummy-head or binaural microphone) and reproduced at two ears. They represent perceptual cues for sound localization, which include the amplitude of sound at each ear, i.e. IID (*Interaural Intensity Difference*), the arrival time at each ear, i.e. ITD (*Interaural Time Difference*) and the spectrum difference of audio signal. Contrarily, spatial audio signals are a group of uncorrelated (or correlated) signals delivered to individual transducers of the audio display system (Tran *et al.*, 2009). They reconstruct arbitrary sound fields within the space. However both are used in the context of achieving spatial immersion of the listener. Stewart built a system, called "The amblr", which served as an art installation. The amblr is a binaural auditory display for music discovery, in which spatial audio is rendered with virtual Ambisonics (STEWART, 2010). Ambisonics is a technique especially valuable for AD. It aims at recording information on the soundfield and reproducing it over a loudspeaker array to produce the impression of hearing three dimensional sound image. The method has well-defined mathematical foundations that allow for easy manipulation of spatial elements. Also conventional multichannel diffusion can be integrated with Ambisonics systems.

There are some other spatial reproduction systems that may be used in AD installations, among them one may refer to Directional Audio Coding (DirAC), which is a method for spatial sound representation, applicable for different sound reproduction systems. As a spatial-sound processing technique, DirAC can determine the direction of arrival of a sound wave, which is the most important information in spatial hearing (Ahonen *et al.*, 2012). Also, mentioned in Introduction externalized sound through headphones, wavefield synthesis, surround systems, 3D audio systems may be used in AD. They may be applied in the context of an immersive auditory display front-end, aimed at spatial interactive sonification to creating a flexible virtual soundscape environment.

2.2. Music Technology – interactive sonification applications

In the context of music technology we can find many examples of interactive data sonification, among which spatial auditory display-based applications/venues are the most popular ones (Marshall *et al.*, 2009; Winters, Wanderley, 2012). AlloSphere (2013) is one of such venues that enable musicians working in spatial setups to explore the potential of immersive music, although its primary goal was different.

The relationship between performer, movement and music is always an interesting sonification application. Immersive interfaces for musical expression often use sound synthesis as the means for musicians to compose and perform (Valbom, Marcos, 2005; Berthaut *et al.*, 2011; Selfridge, Reiss, 2011). It should be noted that the Theremin instrument is regarded as the first successful electronic musical instrument and the first known example of a touchless musical interface (Vigliensoni, Wanderley, 2011a). The musician stands in front of the instrument and moves his or her hands in the proximity of two metal antennas. The distance from one antenna determines frequency (pitch), and the distance from the other controls amplitude (volume).

SoundCatcher designed by Vigliensoni and Wanderley (2011b) is as an open-air gestural controller for singers that allows them to sample their performance, loop and process it in real-time, creating new possibilities for performance and composition in live, rehearsal, and recording contexts. It uses ultrasonic sensors to measure the distance of the performer's hands to the device located in a microphone stand. Tactile and visual feedback employing a pair of vibrating motors and LEDs are provided to inform the performers when they are inside the sensed space (Vigliensoni, Wanderley, 2011b).

A series of gesture-controlled "virtual reality instruments" using computer vision, magnetic trackers, and data gloves for user input were designed and created by Mäki-Patola *et al.* (2005) to evaluate and analyze their efficiency, learning curve, latency, lack of tactile feedback, and system features. Among these instruments one may found virtual drum, air guitar, xylophone, and membrane, an example of such an instrument is shown in Fig. 1 (see

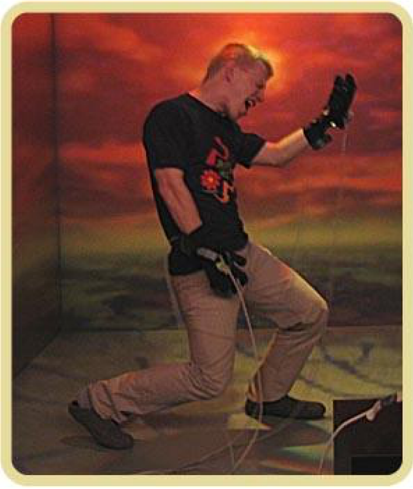


Fig. 1. Virtual air guitar (http://airguitar.tml.hut.fi/whatis.html, 2014).

the http://airguitar.tml.hut.fi/whatis.html website for more information, 2014).

As suggested by Winters *et al.* (2013), there are three levels of gestures in the study of expressive movement that need to be conveyed in sonification, the material, structural, and interpretive. Typically, high-resolution motion capture systems are used to study these movements. Sonification can be used to quickly browse through these data to facilitate the analysis process (Winters *et al.*, 2013; Brazil *et al.*, 2002). They implemented sonification strategies for conveying arousal and valence, two important theoretical dimensions of emotion (Winters *et al.*, 2012; 2013).

Another example of sonification may be the sound mixing system designed by Lech within his Ph.D. thesis and supervised by Kostek (Lech, Kostek, 2013a; 2013b). It is materialized in a system that enables to mix sound with hand gestures recognized in a video stream (Lech, Kostek, 2013b). It is an example of automatic sonification of video sequences for artistic purposes.

Audio mixing or mixdown is the process by which multiple recorded sound tracks are combined into one or more channels. This process may be considered as creating sound rather than only editing it. As a performance tool, it offers a natural method of controlling sound during the mixing process. There are two basic ways to mix sound. The first, so-called ("out of the box") uses audio mixing consoles with knobs and faders, and the second, so-called ("in the box"), employs DAW (Digital Audio Workstations) stations equipped with mixing software. However, in either case the ultimate aim is to obtain a very impressive and realistic sound. To some extent, both mentioned methods rely on visual information. This applies especially to DAW-based mixing, which may be seen as operating against the rule of "listen with your ears, not your eyes". Moreover, a user of a DAW station is limited by the ergonomics of a computer interface (i.e. mouse, keyboard, touch screen).

The engineered system is composed of a PC, a webcam, a multimedia projector and a screen for the projected image (see Fig. 2). A camera lens is directed at the projection screen. The whole projected image and the shadows of the user's hands are visible in the captured video stream. The user is situated in a sweet spot (a best listening position) located between the screen and the multimedia projector, from where they can control mixing processes by hand gestures. No infrared sensors, infrared cameras, gloves, accelerometers or markers are used. The system bases on subtracting the video stream captured by the camera from the image projected by the multimedia projector and finding hands in the processed output. Both dynamic gestures (motion trajectory) and static gestures (palm shape) are recognized by the system. Dynamic gestures are closely associated with static gestures. Thus, performing the same motion with a palm shaped differently has various meanings. Moreover, the order in which gestures are performed can represent a gesture class.

The system is divided into two sub-systems, namely the application recognizing gestures, and the application being a gesture dedicated graphical overlay for any DAW software. The communication with the DAW software is based on the MIDI protocol. The graphical overlay receives system actions generated by the gesture recognition application and sends relevant MIDI messages. In the background of the software there are two decision classifiers, i.e. SVM (Support Vector Machine) and fuzzy rule-based system. After initializing the gesture recognition application, the user can train the SVM classifiers separately for the left and right hand. This enables to assign audio parameters for each hand independently and modify two parameters simultaneously. Employing fuzzy inference in gesture recogni-

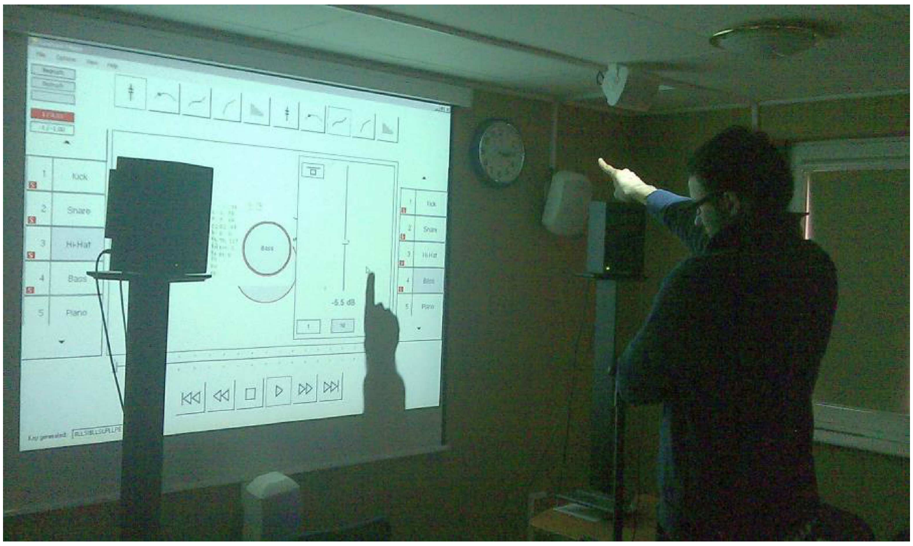
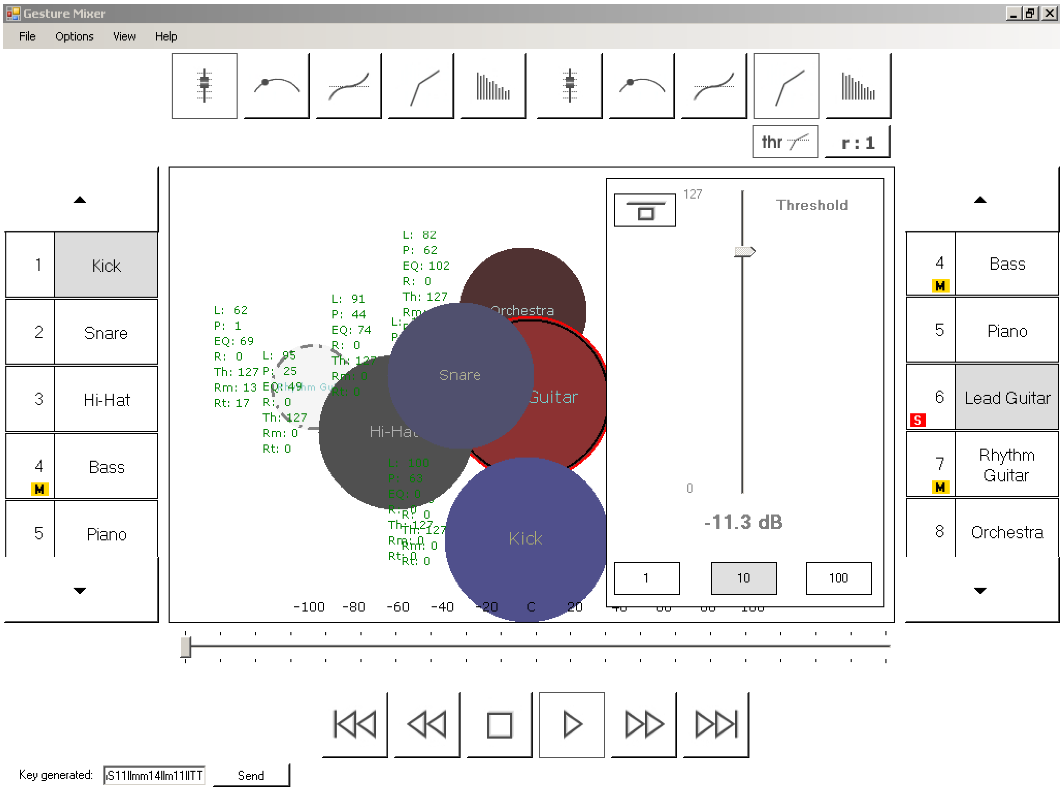


Fig. 2. Manipulation and sonification of musical objects through hand gesture recognition (Lech, Kostek, 2013b).

tion enabled to obtain average efficacy equal 96.94% for one hand gestures, and 99.30% for both hands gestures.

The system has been developed in such a way that mixing operations can be performed without visual support, i.e. auditory display only or using a graphical user interface shown in Fig. 3. In case the system is handled by gestures only it is possible to close eyes and perform operations without affecting sight also when visual stimuli are provided. The GUI contains menu with iconographic representation of all available sound mixing operations (Fig. 3). The horizontal and vertical positions represent the panorama and equalizer gain, respectively. The middle section of the GUI application (middle, left side) contains circles representing audio sources (Fig. 3). The size of the circle represents the level. Directing a hand over the circle with an index finger extended selects the particular audio source. With the audio source selected, hand movements cause respective circle position changes and thus the panorama or equalizer gain can smoothly be adjusted. A user can choose parameters and operations by directing a hand over GUI icons. Some of these functions can be chosen directly by performing a dynamic gesture with a palm appropriately shaped. In Fig. 3 there are audio source listed employed in the mixing process.

Experiments with the system were constructed in such a way that the influence of parameter visualization on sound mixing results and the ergonomics of the interface in comparison with mouse and keyboard could be verified. The sound mixing processes were carried out using the interface engineered and the Steinberg Cubase Studio 5. The experiments have been performed for various manners of system controlling. In the experiments 10 professional mixing engineers have been involved. The task of each engineer was to mix provided eight audio tracks which significantly differed from each other regarding both musical and signal features. None of the engineers had been familiar with the provided audio material before the experiments. Each mixer was asked to develop the individual idea for the final qualities of a mix. The aim was to preserve this idea in all mixing manners, obtaining, in consequence, identical mix every time. Order of the mixing manners was different for each engineer. Its aim was to eliminate effect of learning the process leading to serial correlation. When finished, each engineer was asked to fill in the questionnaire examining various aspects of the system. The examined qualities were: precision, convenience and intuitiveness. The engineers were also asked to order their own mixes from best sounding to the worst sounding. Based on statistical evidence the results obtained proved that visualization of audio signal parameters adversely affects the aesthetic value of obtained mixes. Mixing by hand gestures leads to obtaining mixes of a higher aesthetic value than mixing



Key generated: S11llmm14llm11llTT

Fig. 3. Graphical user interface supporting visual mode of sound mixing, circles represent audio sources, its horizontal and vertical positions represent the panorama and equalizer gain, and the size of the circle represents audio source level (Lech, Kostek, 2013a).

with a mouse and keyboard. Also, it turned out that when not supported by visualization and displayed parameter values the engineers seemed to devote much more attention to sound balance. Eight of the audio engineers considered the engineered system, handled by gestures in the constraint GUI mode, as enabling to focus on sound better. One of the mixing engineers considered manners employing gesture interaction as enabling to better focus on sound, regardless presence or absence of visual information. For one engineer, the engineered system involved sight to the smaller extent only when handled by mouse and keyboard.

Overall, experiments performed with the engineered system showed that it was possible to achieve a sufficient degree of ergonomics and accuracy of mix control. The mixes resulting from mixing via gestures without visual support were more vivid than mixes obtained directly using the DAW software. This was appreciated for sound clarity by mixers and audio professionals and conformed to project expectations (Lech, Kostek, 2013b).

2.3. Music Technology – cross-modal applications

The phenomenon of perceiving the world based on combined inputs from human senses resulting in interactions between two or more different sensory modalities is called multimodal (or cross-modal) perception. The eye-gaze tracking system was employed in the so-called audio-visual correlation experiments (Kunka, Kostek, 2012). The role of the system was to record the subject's gaze fixation points referring to his or her visual attention (see Fig. 4). The exploitation of an eye-gaze tracking technique in the investigation of the impact of visual stimuli on virtual sound source localization has been published by the authors previously (Kunka *et al.*, 2010). A prototype device, known as the Cyber-eye, was constructed in the Faculty of Electronics, Telecommunications and Informatics of the Gdansk University of Technology (Kunka, Kostek, 2013). The device enables to illuminate computer users' eyes by infrared light and to acquire eye view for further processing. The system is composed of hardware solutions and accompanying software that analyze a user's activity during a given task (Kunka *et al.*, 2010; Kunka, Kostek, 2013). This system serves also as an auditory display, because one of the system functionalities enables to listen to the music when the user keeps his gaze sufficiently long on the displayed object (the so-called ROI – region of interest if it corresponds to the musical instrument or a vocalist). An example of denoting ROI that may be used in the annotation process based on the gaze-tracker is shown in Fig. 5.

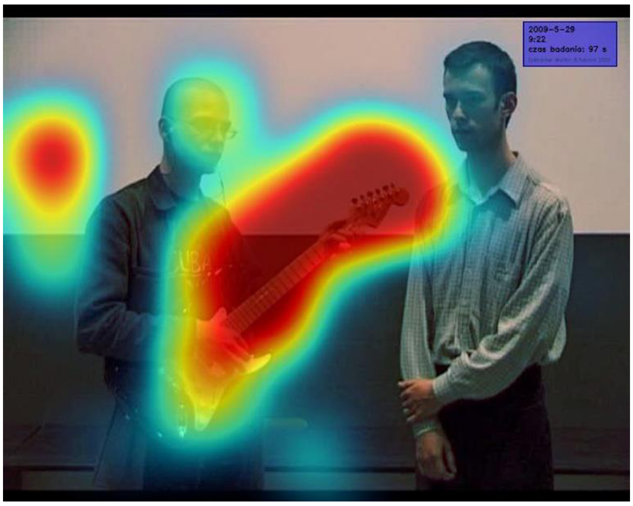


Fig. 4. The heat map generated by WWW Cyber-eye on the designed interface.

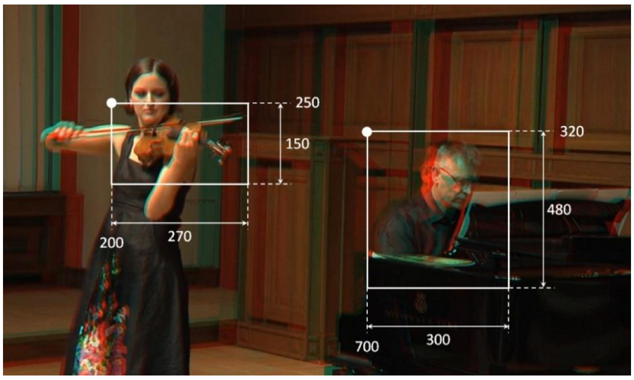


Fig. 5. Denoting ROI – region of interest on the image to be annotated by gaze tracking (3D image for stereoscopic glasses viewing).

The information about the direction of the viewer's gaze allows attractive elements of the presented visual content to be tracked. These data are useful in the objectivization of the test procedure results obtained during the subjective evaluation (Kunka, Kostek, 2013). It should be emphasized that the study of the interaction of sound and visual stimuli on human perception may contribute to the introduction of some changes to the preparation of audio-visual content, also with regard to stereoscopic video and spatial audio. Red color in the heat map generated by the gaze-tracking application denotes most intense user's gazing at the objects in the image (Fig. 4).

Another example of the cross-modal applications is a system that may provide a service for music annotation controlled by gaze-tracking. In the MIR literature, there are three main approaches in terms of automatic music annotation (Guy *et al.*, 2010; Hyoung-Gook *et al.*, 2008; Symeonidis *et al.*, 2009; Mufin system; Musicovery system). File annotation by means of automatic tag retrieval from databases such as Gracenote or FreeDB is the simplest method. The second approach uses information based on a low-level description of music (Aucouturier, Hyoung-Gook *et al.*,

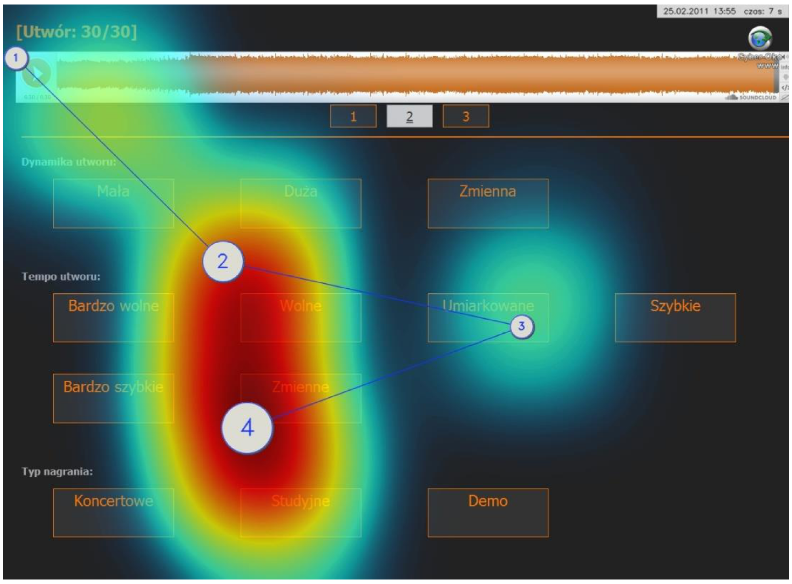


Fig. 6. The heat map generated by WWW Cyber-eye on the designed interface – page No. 2 (in Polish); denotations are as follows: Excerpt no. 30; 1st line – dynamic range – from left to right: low, high, changeable, 2nd and 3rd lines – tempo: very slow, slow, moderate, fast, very fast, changeable, 4th – type of recording: concert-live, studio, demo.

2005; Kostek, 1999; 2005; Kostek, Czyzewski, 2001; Pampalk *et al.*, 2005; Pachet, 2003). In the third approach, individuals are employed to manually add tags to music files. This method requires a large number of "experts" with musical background, and is time-consuming. The method may also be called social tagging, when a statistically significant number of people participate in the process.

Manual annotation of musical pieces may be supported by the analysis carried out with regard to computer users' reactions to music they listen to. Currently, the technological potential supports gaze-tracking, in which objectivization of annotation process is possible by means of observing the level of the user's interest in the retrieved multimedia material. The very same technology provides also a possibility to audify the musical object presented at the screen by means of gaze tracking.

An example of such experiments are shown in Fig. 6. Subjects were asked to tag individual musical excerpts. The page shown refers to tempo of the recording. To facilitate navigation (and also to listen to the song), time line of a given musical piece as well as navigation elements through the music collection are displayed in the upper part of the form on each page. A heat map generated for the page constructed for this purpose is shown in Fig. 6. As mentioned already, generated colors – from blue – the most infrequent to red – the most frequent denote frequency of looking at the objects in the image). To simplify: the larger and red (more focused) the area in the heat map is, the longer the user fixated his/her gaze on a target, thus circles of varying sizes signify "fixations" – areas where the user looked for a significant amount of time. In addition, a gaze plot represents a visualization of the path a user's eyes followed from one point to another. A line is drawn to represent this path. The user's task in the presented example is to annotate an appropriate tempo to a given musical excerpt (Kostek, 2013).

2.4. Music Information Retrieval

Music Information Retrieval (MIR) is an interdisciplinary domain that focuses on automated extraction of information from audio signals, and enables to search the indexed information (Aucouturier, Pachet, 2003; Benetos, Kotropoulos, 2008; Bisesi, Parncutt, 2011; Głaczyński, Łukasik, 2011; Holzapfel, Stylianou, 2008; Kostek, 1999; 2005; Kostek, Czyzewski, 2001; Li *et al.*, 2003; Mandel, Ellis, 2007; Pachet, Cazaly, 2003; Pampalk *et al.*, 2005; Tzanetakis, Cook, 2002; http://www.ismir.net/). The ongoing research focuses on the improvement of the efficiency and effectiveness of music recognition (e.g. in terms of performance), but also on the way to deal with data analysis retrieved from music collections. Among MIR system one should list music recommendation services that include social networking systems, Internet radio stations, and Internet music stores (Guy *et al.*, 2010; Hyoung-Gook *et al.*, 2005; Ness *et al.*, 2009; Symeonidis *et al.*, 2008; http://www.mufin.com/us/, 2013; http://musicovery.com/, 2013).

As pointed out by Stewart and Sandler (2012) MIR applications may be intended for an expert user with a highly specific task or for a general consumer. Therefore human factors always do need to be considered for music browsing or search interfaces. They showed a timeline of auditory display systems enhancing searching and browsing audio content created over the years, which shows that the concept of human computer interfaces often in the form of auditory display has been present in music collection search for more than two decades (Stewart, Sandler, 2012).

Brazil *et al.* (2002) proposed Sonic Browser, a tool for accessing sounds or collections of sounds using sound spatialization and context-overview visualization techniques (see Fig. 7) (Brazil *et al.*, 2002). The intention of the authors of the Sonic Browser was to map properties of the sonic objects to arbitrary features of the visual display. For example, file size can be denoted by size of visual symbols, horizontal and vertical location may be associated with date and time. The user's GUI (Fig. 7) shows visual display in which axes the Y-axis denotes the file size against file name on the X-axis. All sonic objects within the grey shaded circle play simultaneously, panned out in a stereo-space around the cursor.

Among the most innovative systems created to navigate through music collections one can name the nepTune (http://www.cp.jku.at/projects/nepTune/, 2014). Given an arbitrary collection of digital music files, nepTune creates a virtual landscape which allows the user to freely navigate in this collection. The clustering is used to generate a 3D island landscape (see Fig. 8) in which the user can hear the closest sounds

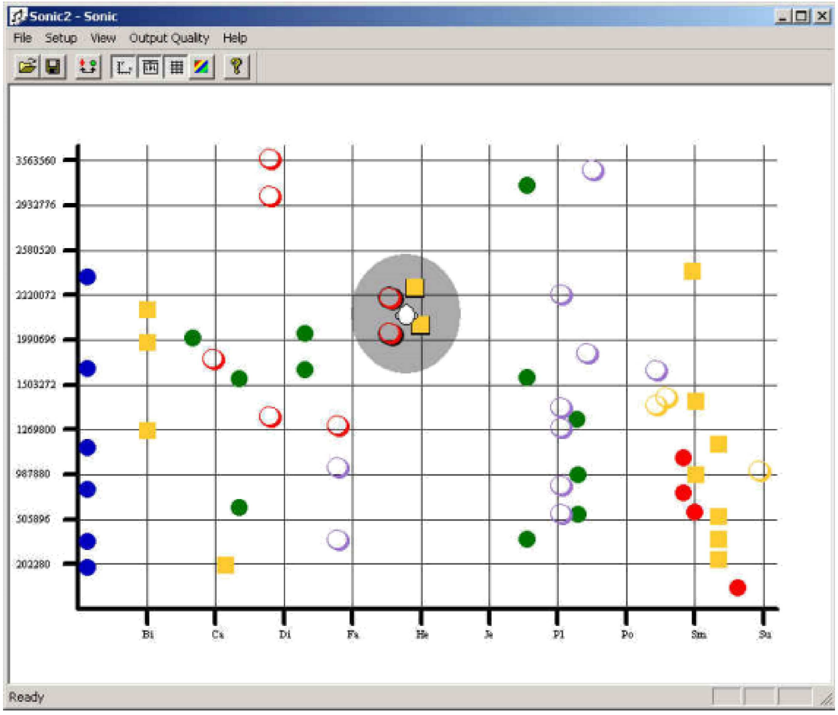


Fig. 7. Sonic Browser user interface; Y-axis denotes the file size against file name on the X-axis (Brazil *et al.*, 2002; Brazil, Fernström, 2003).

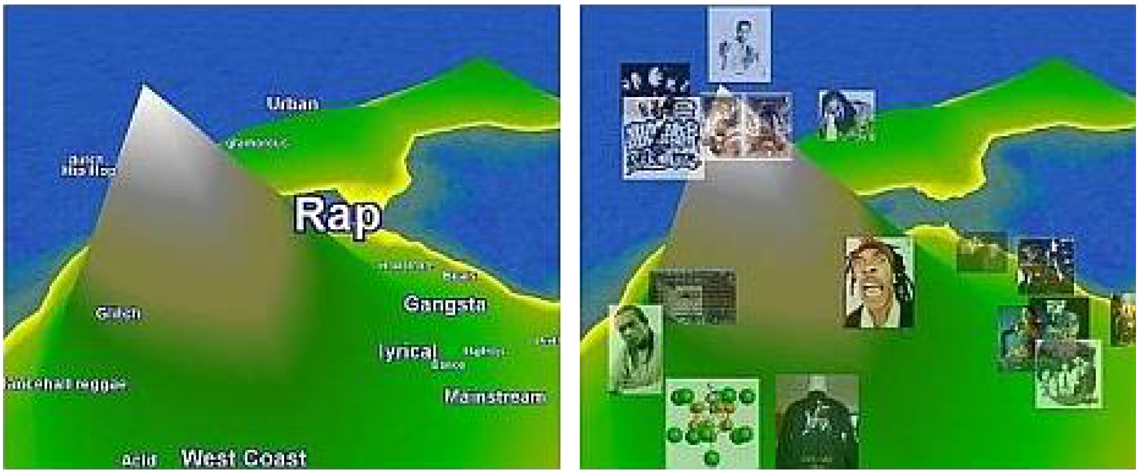


Fig. 8. nepTune navigation interface (http://www.cp.jku.at/projects/nepTune/, 2014).

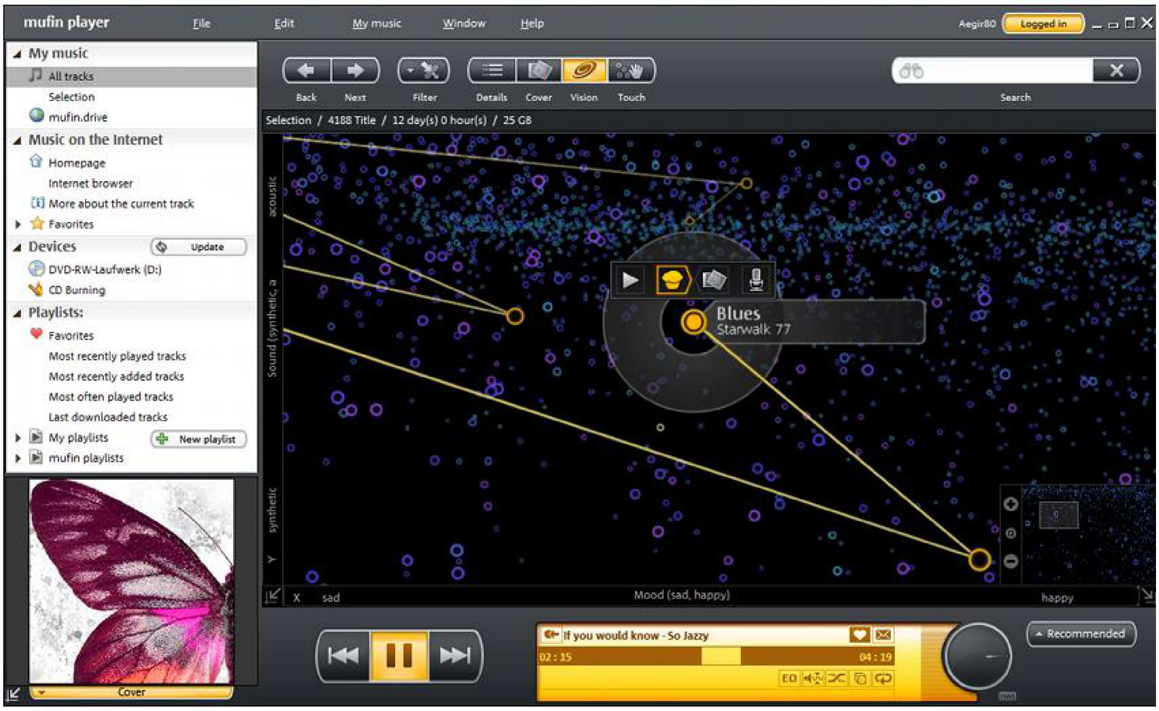


Fig. 9. Mufin – graphical presentation of music collection (http://www.mufin.com/us/, 2013).

with respect to his/her current position via a surround sound system.

Music FINder is a music recommendation search engine (http://www.mufin.com/us/, 2013). The service allows to graphically present musical pieces collected on the server in the form of a three-dimensional map (see Fig. 9). This enables one to observe the relationships among particular tracks of the database. One is able to define the mood (horizontal plane) sound choice from synthetic to acoustic (vertical plane) and when moving backwards or forwards, the user is able to select the adequate tempo of a music (from calm to aggressive) (http://www.mufin.com/us/, 2013).

Musicovery is an interactive Internet radio station created by Castaignet and Vavrille in 2006 in France (http://musicovery.com/, 2013). The system creates playlists containing recommended tracks according to the user's mood. Playlists of the user's anticipated preferences are built on the basis of the obtained classification. It is possible to select musical pieces from a specified decade. The service displays recommendation in the form of a map (see Fig. 10) that shows the relationships among musical pieces (http://musicovery.com/, 2013).

In the rich literature of this subject, many other examples of user interfaces in the domain of MIR or music recommendation in the form of auditory displays may be found. One may refer to the paper by Stewart and Sandler (2012) or use search engines to find information on the auditory display – music technology keywords.

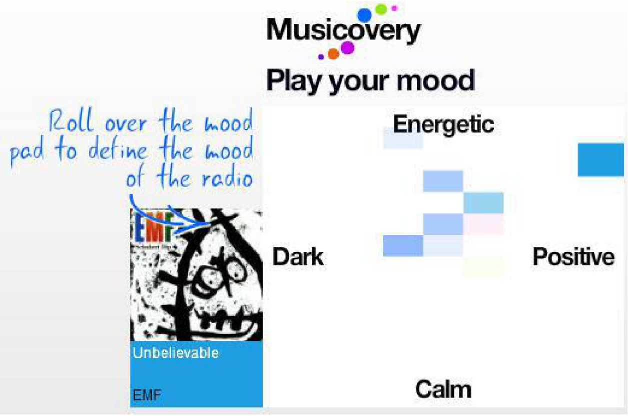


Fig. 10. Musicovery – the user interface (http://musicovery.com/, 2013).

3. Concluding remarks

In this paper, a brief review of some auditory display area notions applied to music technology domain was presented along with some examples of the research performed or supervised by the author in this area. Whereas the majority of studies within music technology concentrate on technology itself, less attention is paid to the sonification issues. That's why more research should be undertaken to identify resources and needs to work collaboratively on the development of common solutions in these two areas.

Challenges that could be identified within the joint music and acoustics technology and auditory display areas are related to the role of human factors such as for example user's personality and experience, emotions, etc. in the user's models and personalized services. Once the relevant perceptual attributes have been identified, a special focus can be put on them in the further process of auditory display synthesis, specifically in the context of sound quality (Blauert, Jekosch, 2012). Also, another important issue which merits attention from researchers and practitioners is spatial audio systems applied to mobile devices to increase user's auditory satisfaction.

Acknowledgments

The research was partially funded within the Project No. SP/I/1/77065/10 entitled: "Creation of Universal, Open, Repository Platform for Hosting and Communication of Networked Resources of Knowledge for Science, Education and Open Society of Knowledge", supported by the National Center for Research and Development (NCBiR, Poland).

References

1. Ahonen J., Del Galdo G., Kuech F., Pulkki V. (2012), *Directional Analysis with Microphone Array Mounted on Rigid Cylinder for Directional Audio Coding*, JAES, **60**, 5, 311–324.

2. Airguitar website (http://airguitar.tml.hut.fi/whatis.html, accessed March 2014).

3. Allosphere: http://www.allosphere.ucsb.edu/research.php (accessed Nov. 2013).

4. Aucouturier J.-J., Pachet F. (2003), *Representing musical genre: A state of art*, J. New Music Research, **32**, 1, 83–93.

5. Benetos E., Kotropoulos C. (2008), *A tensor-based approach for automatic music genre classification*, Proc. European Signal Processing Conference, Lausanne, Switzerland.

6. Beauchamp J.W. (2011), *Perceptually Correlated Parameters of Musical Instrument Tones*, Archives of Acoustics, **36**, 2, 225–238, DOI: 10.2478/v10168-011-0018-8.

7. Berthaut F., Desainte C.M., Hachet M. (2011), *Interacting with 3D Reactive Widgets for Musical Performance*, J. New Music Research, **40**, 3, 253–263.

8. Bisesi E., Parncutt R. (2011), *An accent-based approach to automatic rendering of piano performance: preliminary auditory evaluation*, Archives of Acoustics, **36**, 2, 283–296.

9. Blauert J. (2012), *A Perceptionist's View on Psychoacoustics*, Archives of Acoustics, **37**, 3, 365–371, DOI: 10.2478/v10168-012-0046-z.

10. Blauert J., Jekosch U. (2012), *A Layer Model of Sound Quality*, J. Audio Eng. Soc., **60**, 1/2, 4–12.

11. Blauert J., Rabenstein R. (2012), *Providing Surround Sound with Loudspeakers: A Synopsis of Current Methods*, Archives of Acoustics, **37**, 1, 5–18, DOI: 10.2478/v10168-012-0002-y.

12. Brazil E., Fernström M., Tzanetakis G., Cook P. (2002), *Enhancing Sonic Browsing Using Audio Information Retrieval*, Proc. International Conf. on Auditory Display, Kyoto, Japan.

13. Brazil E., Fernström M. (2003), *Audio Information Browsing With The Sonic Browser*, Proc. CMV'03 Proceedings of the conference on Coordinated and Multiple Views In Exploratory Visualization, IEEE Computer Society Washington, DC, USA.

14. Dobrucki A., Plaskota P., Pruchnicki P., Pec M., Bujacz M., Strumillo P. (2010), *Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals*, J. Audio Eng. Soc., **58**, 9, 724–738.

15. Fernström M., McNamara C. (2005), *After Direct Manipulation – Direct Sonification*, ACM Transaction on Applied Perception, **2**, 4, 495–499.

16. Głaczyński J., Łukasik E. (2011), *Automatic music summarization. A "thumbnail" approach*, Archives of Acoustics, **36**, 2, 297–309.

17. Guy I., Zwerdling N., Ronen I., Carmel D., Uziel E. (2010), *Social media recommendation based on people and tags*, ACM, 194–201.

18. Hermann T. (2008), *Taxonomy and Definitions for Sonification And Auditory Display*, Proc. of the 14th International Conference on Auditory Display, Paris, France June 24–27.

19. Holzapfel A., Stylianou Y. (2008), *Musical genre classification using nonnegative matrix factorization-based features*, IEEE Transactions on Audio, Speech, and Language Processing, **16**, 2, 424–434.

20. Hyoung-Gook K., Moreau N., Sikora T. (2005), *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, Wiley & Sons.

21. Kostek B. (1999), *Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics, Studies in Fuzziness and Soft Computing*, Physica Verlag, Heildelberg, New York.

22. Kostek B., Czyzewski A. (2001), *Representing Musical Instrument Sounds for their Automatic Classification*, J. Audio Eng. Soc., **49**, 768–785.

23. Kostek B. (2005), *Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval and Psychophysiology of Hearing*, Springer Verlag, Berlin, Heidelberg, New York.

24. Kostek B. (2013), *Music Information Retrieval in Music Repositories*, Intelligent Systems Reference Library, 42, Springer Verlag, Berlin, Heidelberg, Chapter 17, 464–489.

25. Kostek B. (2013), Auditory Display from the Music Technology Perspective, 19th International Conference on Auditory Display (ICAD-2013), Lodz, Poland.

26. Kramer G., Walker B., Bonebright T., Cook P., Flowers J., Miner N., Kunka B., Kostek B. (2012), *Objectivization of audio-video correlation assessment experiments*, Archives of Acoustics, **37**, 1, 63–72.

27. Kunka B., Kostek B., Kulesza M., Szczuko P., Czyzewski A. (2010), *Gaze-Tracking-Based Audio-Visual Correlation Analysis Employing Quality of Experience Methodology*, Intelligent Decision Technologies, IOS Press, **32**, 217–227.

28. Kunka B., Kostek B. (2013), *New Aspects of Virtual Sound Source Localization Research – impact of visual angle and 3D video content on sound perception*, J. Audio Eng. Soc., **61**, 5, 280–189.

29. Lech M., Kostek B. (2013), *Evaluation of the influence of ergonomics and multimodal perception on sound mixing while employing a novel gesture-based mixing interface*, J. Audio Eng. Society, **61**, 5, 301–313.

30. Lech M., Kostek B. (2013), Gesture-Controlled Sound Mixing System, 19th International Conference on Auditory Display (ICAD-2013), Lodz, Poland.

31. Li T., Ogihara M., Li Q. (2003), *A comparative study on content-based music genre classification*, Proc. 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 282–289, Toronto, Canada.

32. Mandel M., Ellis D. (2007), *LABROSA's audio music similarity and classification submissions*, Music Information Retrieval Information Exchange (MIREX).

33. Marshall M., Malloch J., Wanderley M.M. (2009), *Gesture Control of Sound Spatialization for Live Musical Performance*, in Gesture Based Human Computer Interaction and Simulation, M. Sales Dias (ed.), Berlin, Springer, 227–238.

34. Mäki-Patola T., Laitinen J., Kanerva A., Takala T. (2005), *Experiments with virtual reality instruments*, In Proc. Conf. on New Interfaces for Musical Expression, Vancouver, BC, Canada, 11–6.

35. Mufin system; http://www.mufin.com/us/ (accessed Nov. 2013).

36. Musicovery system; http://musicovery.com/ (accessed Nov. 2013).

37. NepTune system; http://www.cp.jku.at/projects/nepTune/ (accessed March 2014).

38. Neuhoff J. and co-authors (1999), *Sonification report: Status of the field and research agenda*, Tech. Rep., International Community for Auditory Display, (http://www.icad.org/websiteV2.0/References/nsf.html, accessed Nov. 2013).

39. Ness S., Theocharis A., Tzanetakis G., Martins L.G. (2009), *Improving automatic music tag annotation using stacked generalization of probabilistic SVM outputs*, 17 ACM International Conf. on Multimedia, New York, NY.

40. Pachet F., Cazaly D. (2003), *A classification of musical genre*, Proc. RIAO Content-Based Multimedia Information Access Conf.

41. Pampalk E., Flexer A., Widmer G. (2005), *Improvements of audio-based music similarity and genre classification*, Proc. Int. Symp. Music Information Retrieval (ISMIR), London, UK.

42. Selfridge R., Reiss J. (2011), *Interactive Mixing Using Wii Controller*, AES 130th Convention, London, UK.

43. Shinn-Cunningham B.G., Streeter T. (2005), *Spatial Auditory Display: Comments on Shinn-Cunningham et al.*, ICAD 2001, ACM Transactions on Applied Perception, **2**, 4, 426–429.

44. SONIFICATION – http://sonification.de/son, a website providing definitions of notions within AD, by HERMANN T, 2014.

45. Stewart R. (2010), *Spatial Auditory Display for Acoustics and Music Collections*, Ph.D. thesis, School of Electronic Engineering and Computer Science Queen Mary, University of London, UK.

46. Stewart R., Sandler M. (2012), *Spatial Auditory Dis play*, J. Audio Eng. Soc., **60**, 11, 936–946.

47. Stockman T., Roginska A., Walker B., Metatla O. (2012), *Guest Editors' Note: Special Issue on Auditory Display*, J. Audio Eng. Soc., **60**, 7/8, 496.

48. Symeonidis P., Ruxanda M.M., Nanopoulos A., Manolopoulos Y. (2008), *Ternary semantic analysis of social tags for personalized music recommendation*, Proc. 9th Int. Symp. Music Information Retrieval (ISMIR), 219–224.

49. The International Society for Music Information Retrieval /Intern. Conf. on Music Information Retrieval website http://www.ismir.net/ (accessed Nov. 2013).

50. Tran P.K., Amrein B.E., Letowski T.R. (2009), Audio Helmet-Mounted Displays, In T.R. Letowski, E. Schmeisser, D. Russo, & C.E. Rash (Eds.), Displays: Sensation, Perception and Cognition Issues, Rash, C.E., Russo, M.B., Letowski, Ft. Rucker, U.S. Army Aeromedical Research Laboratory: Ft. Rucker, AL, 175–236.

51. Tzanetakis G., Cook P. (2002), Musical genre classification of audio signal, IEEE Transactions on Speech and Audio Processing, 10, 3, 293–302.

52. Valbom L., Marcos A. (2005), *WAVE: Sound and music in an immersive environment*, Computers & Graphics, **29**, 6, 871–881.

53. Vamvakousis Z., Ramirez R. (2012), Temporal Control In the EyeHarp Gaze-Controlled Musical Interface, Inter. Conf. on New Interfaces for Musical Expression, NIME'2012, Ann Arbor, Michigan, USA.

54. Vigliensoni G., Wanderley M.M. (2011a), *Touchless Gestural Control of Concatenative Sound Syn-*

thesis, Schulich School of Music, McGill University, (MoA), Montreal, Canada.

55. Vigliensoni G., Wanderley M.M. (2011b), *Soundcatcher: Explorations In Audio-Looping And Time-Freezing Using An Open-Air Gestural Controller*, McGill University Music Technology Area, Montreal, Canada.

56. Winters R.M., Wanderley M.M. (2012), *New Directions for Sonification of Expressive Movement in Music*, 18th International Conf. on Auditory Display (ICAD2012) Atlanta, Georgia (June 18–21, 2012) (http://hdl.handle.net/1853/44450, accessed Nov. 2013).

57. Winters R.M., Hattwick I., Wanderley M.M. (2013), *Integrating Emotional Data into Music Performance: Two Audio Environments for the Emotional Imaging Composer*, International Conference on Music and Emotion, Jyväskylä, Finland.