AUDITORY DISPLAY FROM THE MUSIC TECHNOLOGY PERSPECTIVE

Bożena Kostek
Audio Acoustics Laboratory,
Faculty of Electronics, Telecommunications and Informatics
Narutowicza 11/12, 80-233 Gdansk, Poland
bokostek@audioakustyka.org

ABSTRACT

This paper presents some applications of Auditory Displays (AD) in the domain of music technology. First, the scope of music technology and auditory display areas are shortly outlined. Then, the research trends and system solutions within the fields of music technology, music information retrieval and music recommendation are discussed. Finally, an example of an auditory display that facilities music annotation process based on gaze tracking is shown.

1. INTRODUCTION

Music Technology (MT) 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. Auditory Display, that in the past referred mostly to systems which conveyed messages in the form of sounds [1] or used navigation controlled by means of sound communication [2], has largely broadened its scope over the years. Nowadays, much of the AD research focuses on spatial auditory display (e.g. on fields 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.

It is worth mentioning that a special issue of Journal of the Audio Engineering Society devoted to AD, has been recently published [3]. 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 (SMC), Haptic Audio Interaction Design (HAID), Audio Mostly (AM), New Instruments for Musical Expression (NIME), and Interactive Sonification (IS). All thirteen papers that belong to this issue present a very broad scope of AD issues and applications [3].

The paper recalls some system solutions within music technology, music information retrieval and music recommendation. Then, an example of an auditory display that facilities music annotation process based on gaze tracking technology is shown. Finally, some comments referring to main research challenges within music technology area that need assistance from other domains, also from auditory display are given.

2. MUSIC TECHNOLOGY – RESEARCH TRENDS

As formulated by Hermann, auditory display term refers to 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 [4]. He also pointed out that the context of the user (user, task, background sound, constraints) and the designed application are essential for auditory display. Furthermore, sonification is an integral component within an auditory display system which addresses the actual rendering of sound signals [4]. 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 [5]. Direct sonification refers to the case when information displayed does not require an intermediary mapping of data to an audio signal (sound) or a set of audio parameters. Contrarily, indirect sonification occurs when the information sought is rendered from an abstract representation of objects into an audio signal (sound).

In this paper examples of sonification applied to music technology are recalled.

2.1. 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 [6] [7]. AlloSphere [8] 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 are an important domain which often use sound synthesis as the means for musicians to compose and perform [9] [10] [11]. It should be noted that the Theremin is regarded as the first successful electronic musical instrument and the first known example of a touchless musical interface [12].

SoundCatcher [13] designed by Vigliensoni and Wanderley 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 [13].

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. to evaluate and analyze their efficiency, learning curve, latency, lack of tactile feedback, and system features [14]. Among these instruments one may found virtual drum, air guitar, xylophone, and membrane, an example of such an instrument is shown in Fig. 1. [15].

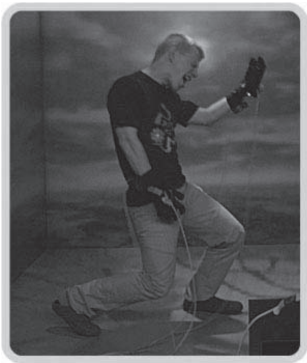

Fig. 1. Vitual air guitar [15]

As suggested by Winters et al. [16], 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 [16] [17]. They implemented sonification strategies for conveying arousal and valence, two important theoretical dimensions of emotion [7] [16]. Another example of sonification is the sound mixing system designed by Lech and Kostek [18] [19]. It is materialized in a system that enables to mix sound with hand gestures recognized in a video stream (see Fig. 2) [19].The system has been developed in such a way that mixing operations can be performed both with or without visual support, i.e. auditory display only.

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 [20]. The role of the system was to record the subject's gaze fixation points referring to his or her visual attention (see Fig. 3). 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 [21]. A prototype device, known as the Cyber-eye, was constructed in the Faculty of Electronics, Telecommunications and Informatics of the Gdansk University of Technology [21]. 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 [21] [22]. 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).

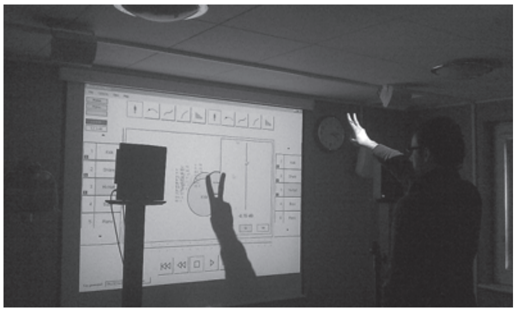

Fig. 2. Manipulation and sonification of musical objects through hand gesture recognition [19]

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 [22]. 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. 3).

Fig. 3. The heatmap generated by WWW Cyber-eye on the designed interface

2.2. 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 [23-36]. 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 [37-42].

As pointed out by Stewart and Sandler MIR applications may be intended for an expert user with a highly specific task or for a general consumer [43]. 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 [43].

Brazil et al. proposed Sonic Browser, a tool for accessing sounds or collections of sounds using sound spatialization and context-overview visualization techniques (see Fig. 4) [17].

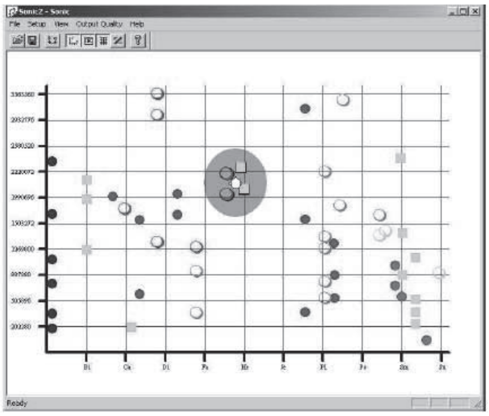

Fig. 4. Sonic Browser user interface [17]

Among the most innovative systems created to navigate through music collections one can name the nepTune [44]. 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. 5) in which the user can hear the closest sounds with respect to his/her current position via a surround sound system.

Fig. 5. nepTune navigation interface [44]

Music FINder is a music recommendation search engine [41]. The service allows to graphically present musical pieces collected on the server in the form of a three-dimensional map (see Fig. 6). 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) [41].

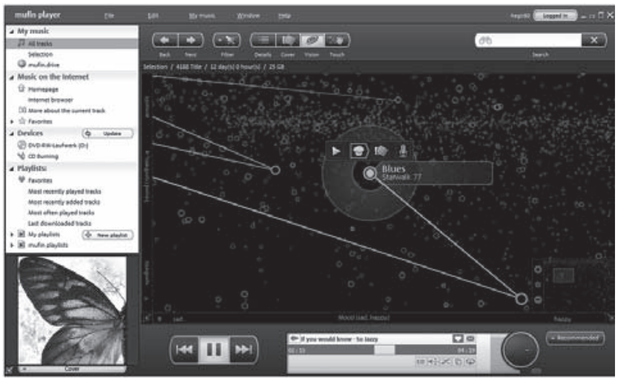

Fig. 6. Mufin – graphical presentation of music collection [41]

Musicovery is an interactive Internet radio station created by Castaignet and Vavrille in 2006 in France [42]. 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. 7) that shows the relationships among musical pieces [42].

Fig. 7. Musicovery – the user interface [42]

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 [43] or use search engines to find information on the auditory display – music technology keywords.

2.3. Music annotation controlled by gaze-tracking

In the MIR literature, there are three main approaches in terms of automatic music annotation [37-42]. 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 [23] [28] [29] [33] [38]. 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. This 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 the heat map generated for the page constructed is shown in Figure 8 (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). The page shown refer 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. The user’s task in the presented example is to annotate an appropriate tempo to a given musical excerpt.

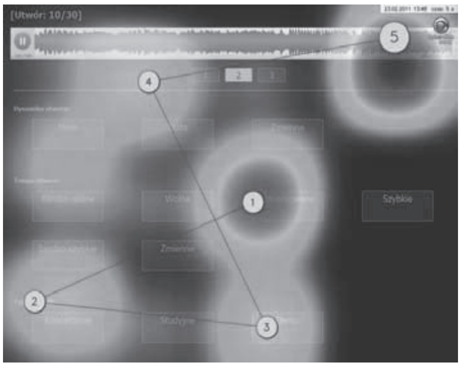

Fig. 8. The heatmap generated by WWW Cyber-eye on the designed interface

3. CONCLUDING REMARKS

In this paper, a short review of some auditory applications to music technology domain was presented. 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 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. 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.

4. ACKNOWLEDGMENTS

The research was 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).

5. REFERENCES

[1] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff and co-authors, “Sonification report: Status of the field and research agenda”, Tech. Rep., International Community for Auditory Display (1999). http://www.icad.org/websiteV2.0/References/nsf.html.

[2] A. Dobrucki, P. Plaskota, P. Pruchnicki, M. Pec, M. Bujacz, P. Strumillo, “Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localization Trials with Visually Impaired and Sighted Individuals,” J. Audio Engineering Society, vol. 58, no. 9, pp. 724-738 (2010).

[3] T. Stockman, A. Roginska, B. Walker, O. Metatla, “Guest Editors’ Note: Special Issue on Auditory Display”, 60 Number 7/8, p. 496, (2012).

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

[5] M. Fernström, C. McNamara, “After Direct Manipulation - Direct Sonification”, ACM Transaction on Applied Perception 2, no. 4, 495-499 (2005).

[6] M. Marshall, J. Malloch, M. Wanderley, “Gesture Control of Sound Spatialization for Live Musical Performance,” in Gesture Based Human Computer Interaction and Simulation, M. Sales Dias (ed.), Berlin, Springer, pp. 227-238 (2009).

[7] Winters, R. Michael, Wanderley, Marcelo M., “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

[8] Allosphere: http://www.allosphere.ucsb.edu/research.php

[9] L. Valbom, A. Marcos, “WAVE: Sound and music in an immersive environment”, Computers & Graphics, vol. 29, no. 6, pp. 871-881 (2005).

[10] F. Berthaut, M. Desainte-Catherine, M. Hachet, “Interacting with 3D Reactive Widgets for Musical Performance,” J. of New Music Research, vol. 40, no. 3, pp. 253-263 (2011).

[11] R. Selfridge, J. Reiss, “Interactive Mixing Using Wii Controller,” AES 130th Convention, London (2011).

[12] A.G. Vigliensoni Martin, ”Touchless Gestural Control of Concatenative Sound Synthesis”, Schulich School of Music, McGill University, (MoA), Montreal, CA (2011).

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

[14] T. Mäki-Patola, J. Laitinen, A. Kanerva, T. Takala, "Experiments with virtual reality instruments. In Proc. Conf. on New Interfaces for Musical Expression", Vancouver, BC, CA, 11–6 (2005).
[15] http://airguitar.tml.hut.fi/whatis.html
[16] R.M. Winters, I. Hattwick, M.M. Wanderley, "Integrating Emotional Data into Music Performance: Two Audio Environments for the Emotional Imaging Composer," The International Conference on Music and Emotion, Jyväskylä, Finland (2013).
[17] E. Brazil, M. Fernström, G. Tzanetakis, P. Cook, "Enhancing Sonic Browsing Using Audio Information Retrieval", Proc. International Conf. on Auditory Display, Kyoto, Japan, (July 2-5, 2002).
[18] M. Lech, B. Kostek, "Evaluation of the influence of ergonomics and multimodal perception on sound mixing while employing a novel gesture-based mixing interface", J. Audio Eng. Society (2013) (*in press*).
[19] M. Lech, B. Kostek, "Gesture-Controlled Sound Mixing System", ICAD, Łódź, PL (2013).
[20] B. Kunka, B. Kostek, "Objectivization of audio-video correlation assessment experiments," Archives of Acoustics, vol. 37, no. 1, pp. 63-72 (2012).
[21] B. Kunka, B. Kostek, M. Kulesza, P. Szczuko, A. Czyzewski, "Gaze-Tracking-Based Audio-Visual Correlation Analysis Employing Quality of Experience Methodology", Intelligent Decision Technologies, IOS Press, vol. 32, 217-227 (2010).
[22] B. Kunka, B. Kostek, "New Aspects of Virtual Sound Source Localization Research – impact of visual angle and 3D video content on sound perception", J. Audio Eng. Soc. (2013) (*in press*).
[23] J.-J. Aucouturier, F. Pachet, "Representing musical genre: A state of art", J. New Music Research, vol. 32(1), 83-93 (2003).
[24] E. Benetos, C. Kotropoulos, "A tensor-based approach for automatic music genre classification", Proc. European Signal Processing Conference, Lausanne, Switzerland (2008).
[25] E. Bisesi, R. Parncutt, "An accent-based approach to automatic rendering of piano performance: preliminary auditory evaluation", Archives of Acoustics, vol. 36, 2, 283-296 (2011).
[26] A. Holzapfel, Y. Stylianou, "Musical genre classification using nonnegative matrix factorization-based features", IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, 2, 424-434 (2008).
[27] J. Głaczyński, E. Łukasik, "Automatic music summarization. A "thumbnail" approach", Archives of Acoustics, vol. 36, 2, 297-309 (2011).
[28] B. Kostek, "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 (1999).
[29] Kostek B., Czyzewski A., "Representing Musical Instrument Sounds for their Automatic Classification", J. Audio Eng. Soc., vol. 49, 768-785 (2001).
[30] B. Kostek, "Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval and Psychophysiology of Hearing", Springer Verlag, Berlin, Heidelberg, New York (2005).
[31] T. Li, M. Ogihara, Q. Li, "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, CA (2003).
[32] M. Mandel, D. Ellis, "LABROSA's audio music similarity and classification submissions", Music Information Retrieval Information Exchange (MIREX) (2007).
[33] E. Pampalk, A. Flexer, G. Widmer, "Improvements of audio-based music similarity and genre classification", Proc. Int. Symp. Music Information Retrieval (ISMIR), London, UK (2005).
[34] F. Pachet, D. Cazaly, A classification of musical genre, Proc. RIAO Content-Based Multimedia Information Access Conf. (2003).
[35] G. Tzanetakis, P. Cook, "Musical genre classification of audio signal", IEEE Transactions on Speech and Audio Processing, vol. 10, 3, 293-302 (July 2002).
[36] The International Society for Music Information Retrieval /Intern. Conf. on Music Information Retrieval website http://www.ismir.net/
[37] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, E. Uziel, Social media recommendation based on people and tags, ACM, 194-201 (2010).
[38] K. Hyoung-Gook, N., Moreau, T. Sikora, "MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval", Wiley & Sons (2005).
[39] S. Ness, A. Theocharis, G. Tzanetakis, L.G. Martins, "Improving automatic music tag annotation using stacked generalization of probabilistic SVM outputs", 17 ACM International Conf. on Multimedia, New York, NY (2009).
[40] P. Symeonidis, M.M. Ruxanda, A. Nanopoulos, Y. Manolopoulos, "Ternary semantic analysis of social tags for personalized music recommendation", Proc. 9th Int. Symp. Music Information Retrieval (ISMIR), 219-224 (2008).
[41] http://www.mufin.com/us/
[42] http://musicovery.com/
[43] R. Stewart, M. Sandler, "Spatial Auditory Display", J. Audio Eng. Soc., Vol. 60, No. 11, 936-946 (2012).
[44] http://www.cp.jku.at/projects/nepTune/