On the Study of Accessibility and Interactivity for the Visually Impaired Computer Musician

Vilson Zattera
Music Department – IA – UNICAMP
vilson.zattera@gmail.com
José Fornari
NICS – UNICAMP
tutifornari@gmail.com

Abstract: With the advancement of computing technology, graphical representations of language and music were naturally incorporated to the digital media. Even though in modern personal computer there is an audio output, it is undeniable that the most relevant data output is visual (shown on the computer screen). In these circumstances, the visually impaired user has a major disadvantage to operate computers and retrieve its information in the same level of the sighted ones. As for music made with computers, blind musicians may face the challenge of being forced to handle graphical symbols, via aiding softwares, as the only possible way of operating these tools. However, these same musicians may present aural advantages, as compared to the sighted ones, to the perception of certain acoustics aspects, such as timbre recognition and sound localization. This theoretical work presents 4 themes of further development that will promote accessibility and explore the aural enhanced capabilities of blind musicians by means of computational resources. They are: 1) Stenographic musical notation, 2) Global reader; 3) Haptic-controlled soundscape synthesis; 4) Adaptive-synesthetic generative music model. Themes 1 and 2 are predominantly related to the accessibility, while themes 3 and 4 explore the aural advantages of the blind musicians. These themes are here introduced, described and discussed.
Keywords: Accessibility, Enhanced Hearing, Computer Music

Resumo: Com o avanço da tecnologia de computação, representações gráficas de linguagem e música foram naturalmente incorporadas às mídias digitais. Mesmo que o computador pessoal contemporâneo tenha uma saída de áudio digital, é inegável que a saída de dados mais relevante é visual (mostrada na tela do computador). Nestas circunstâncias, o usuário com deficiência visual tem uma grande desvantagem para operar computadores e recuperar a sua informação processada, no mesmo nível dos músicos com visão normal. Quanto à música feita com recursos computacionais, músicos cegos podem enfrentar o desafio de serem obrigados a lidar com símbolos gráficos, mesmo auxiliados por softwares especiais, como a única forma possível de utilizarem essas ferramentas computacionais. No entanto, estes mesmos músicos podem apresentar vantagens aurais, em comparação com os não-deficientes visuais, para a percepção de certos aspectos acústicos, tais como o reconhecimento de nuances do timbres e a localização espacial do som. Este trabalho teórico apresenta quatro temas de futuro desenvolvimento que promoverão tanto a acessibilidade quanto a exploração das habilidades auditivas privilegiadas dos músicos cegos, por meio de recursos computacionais. São eles: 1) notação musical estenográfica , 2) leitor global; 3) controle-gestual de síntese de paisagens sonoras ; 4) modelo computacional adaptativo-sinestésico de música generativa. Os temas 1 e 2 são predominantemente relacionadas com a acessibilidade, enquanto que os temas 3 e 4 visam explorar as vantagens auditivas dos músicos cegos. Estes temas são aqui introduzidas, descritos e discutidos.
Palavras-chave: Acessibilidade, Audição Expandida, Musica Computacional.

1. Background

Music is frequently defined as the art of organized sounds. As such, music is an immaterial form of art strongly dependent of time. In music, invisible entities known as sound objects, are structured by the composer in a time line, thus conveying in it meaning and aesthetic relevance. However, as music became increasingly more complex, a notation system was required to register and store musical compositions. Humans are the most visually orientated and dependent of all mammals[1]. Although our sight is not the most accurate – when compared to other animals, such as birds – human cognition heavily depends on visual structures to conceive subjective and objective reality into cognitive maps (Portugali, 1996). Even considering that a musical structure is in fact invisible, intangible and innocuous, its traditional symbolic notation is primordially graphic. As an example, the early music notation system, in the records, the cuneiform tablet created at Nippur, developed in about 2000 BC, is considered to be the first graphic representation of music, which was used to register music composed in harmonies of thirds, in a diatonic scale (Kilmer, 1986). The two most important sonic communication systems developed by humankind – language and music – are normally represented and structured by graphic symbols (respectively, text and musical notation).

With the advent of computing technology, these graphical representations of language and music were naturally passed to the digital media. Computers are programmed in "programming languages", that are based on the structures of natural (human) language, and organized in texts, with semantic and syntactic rules and meaning. In the contemporary personal computer, although there is an audio output, it seems undeniable that the most relevant data output is visual, delivered through the computer monitor. The input of data from the user is normally given by the traditional gestural interfaces (i.e. mouse and keyboard), who, in a way, also depend on the sense vision to be properly operated.

In this scenario, visually impaired user have a major disadvantage to access computer data information in the traditional manner. In terms of music made by computers, blind musicians may face the challenge of being forced to handle with graphical symbols, via aiding softwares (i.e. text-to-speech applications), as the only possible way of having access to the analysis, transformation and synthesis of sounds and music. These same musicians, as people deprived from vision, may as well present aural advantages, as compared to sighted musicians, to certain acoustic aspects, such as the perception of timbre subtle variations and enhanced sound source spatial localization. The amount of data that one human retina transmits to the brain is near the rate of 10 million bits per second, which is comparable to an Ethernet computer connection[2]. Without vision, the area of the brain supposed to process visual data (the brain cortex) is rewired to process information from other senses, such as touch, olfaction and hearing. This phenomenon is called Cross-modal Neuroplasticity (Lotfi, 2010). This has been studied on visually impaired individuals that present enhanced hearing.

[1]Szaflarski, Diane M. How We See:The First Steps of Human Vision. http://www.accessexcellence.org/AE/AEC/CC/vision_background.php (accessed in February 2015)

[2] http://www.eurekalert.org/pub_releases/2006-07/uops-prc072606.php (accessed in February 2015)

For instance, “in a previous study, Franco Lepore and colleagues showed that people who lost their sight at an early age could localize sound, particularly from monaural cues, better than those who could see” (Gougoux, 2005). This theoretical work aims to present four themes of further development that will promote accessibility and explore the aural excellence of blind musicians by means of computational resources. It is intended here to introduce the 4 themes that are planned to be further developed in the postdoctoral research of the first author. The perceptual, cognitive and aesthetic implications of these computational models of analysis, processing and synthesis, of symbolic and acoustic data, will be further analyzed and used in contemporary music compositions and performances. The next section presents these themes of development.

2. The road ahead

This project will explore four main themes: 1) Stenographic Musical Notational (the creation of a symbolic system of musical patterns for quick registration of formal music notational), 2) Global Reader (automatic recognition of text and musical symbols displayed in the computer screen and speech by means of optical character recognition and real-time voice synthesis models). 3) Haptic-Controlled Soundscapes Synthesis (gestural data retrieval for the intuitive parametrization of synthesis models of artificial soundscapes generation); 4) Adaptive-synesthetic generative music model (evolutionary computation models applied for meta-composition of music).

The themes 1 and 2 are predominantly related to the issue of accessibility, as they aim to promote easy and feasible access to the visually impaired musician. The themes 3 and 4 primarily explore the interactivity issue of blind musicians as they intend to explore their aural excellence, specially in terms of their enhanced hearing abilities, such as the fine spatial sound location, as described in (Gougoux, 2005). Once that there is still no formal study (as far as we know) about the enhanced timbre recognition ability among blind people, as well as the existence of an enhanced sense of pitch, loudness and time delay; the theme 4 of this project plans to also study this subject, searching for these evidences and eventually exploring these abilities to guide an evolutionary system of automatic music generation. These four themes are here introduced and described below, as intrinsic parts of the theoretical work here presented. Further development and the implementation of such models will follow in the forthcoming practical development of this project.

2.1 Stenographic Musical Notational

There are several softwares developed to provide access to symbolic musical data. Some examples are: ABC notation (http://abc.sourceforge.net), Guido Musical Notation (http://science.jkilian.de/salieri/GUIDO/index.html) and Lilypond (www.lilypond.org). They are based on customized languages that aim to facilitate the creation and edition of musical notation. A more recent example is VexTab (http://www.vexflow.com/vextab/), that enables to process musical notation online (in a web-browser). This and other computational tools aim to ease the access and lower the required time for the preparation of musical scores. However the vast majority of these solutions are not tackled to the visually impaired musician. The major obstacles are the typical data input interfaces (i.e. keyboard, mouse) and data output interface (the computer monitor and so their graphical user interfaces). Besides that, there is also the learning curve for each software as they tend to not follow the same philosophy of programming and user access.

This research theme aims to minimize this problem, accelerating the process of creating musical scores by using a shorthand language that is intuitive and also capable of textually representing some recurring musical patterns, thereby reducing the amount of symbolic information to be inserted by the user to represent a musical score. This project plans to collect and catalog repeating musical patterns that will be associated with fundamental symbols that are easy to be referred and fast to be notated. One of the main objectives of this development is to create a computer model able to collect symbolic data in text format and convert them straight to a format known as MusicXML (http://www.musicxml.com). According to its website "MusicXML was designed from the ground up for sharing sheet music files between applications, and for storing sheet music files for use in the future". This theme plans to use MusicXML in order to guarantee the portability of these files so they can be used in conjunction with the software category. Moreover, the choice of a system in text mode will directly benefit portability, accessibility and digital inclusion, favoring its use by not only the blind musician but also people with other disabilities or impairments to properly handle a computer mouse and keyboard, such as the physically disable, the very young or the elders.

2.2 Global Reader

As computers became ubiquitous tools in contemporary society, increasingly expanding to all areas of human activity, from science to arts, music has also being affected by its presence. Computers had made easier and practical some musical tasks involving the analysis, processing, and synthesis of acoustic and symbolic data. Through computers, musical notation can be copied, modified and retrieved (almost as easy as text files), through the use musical software editing tool. Similarly, musical performances aided by computers gained access to several intrinsic acoustic aspects and the retrieval of gestrual data by electronic sensors that provide information to make possible the interactive creation of music in real time.

However, the primordial access to the benefits of computation in music still comes mainly through the exercise of the sense of sight. As mentioned before, computing data are customarily arranged and expressed in the form of visual characters (i.e. texts, graphics, animations, etc.) and displayed in the computer screen. Much of this visual information is placed anywhere in the screen, in a dynamic and unexpected manner (for the user's perspective), in different sizes, colors, fonts, formats, positions and directions. Often the characters displayed in GUIs (graphical user interfaces) of softwares, such as the musical tools, are arranged in different directions and presented in non-text format (i.e. digital images, bitmaps, etc.) which makes it impossible to aiding tools (such as the text-to-speech applications, also known as "screen readers") retrieve this information for the blind user.

About the aiding tools, to exemplify, some important commercials ones are: JAWS (Job Access With Speech, by Freedom Scientific), Window-Eyes (by GW Micro), Dolphin Supernova (by Dolphin), System Access (by Serotek), ZoomText Magnifier and / Reader (by AiSquared). Some open-source ones are: ORCA (http://projects.gnome.org/orca/) and NVDA (NonVisual Desktop Access). These are prominent examples of applications for the Anglophone market. However the resources that such softwares offer are limited to specific operating systems and the restricted access of text-only files and applications that have compatibility with such automatic readers. For this reason, it is intended, in this theme of study, the development and implementation of a computational model capable of reading texts in a random-access manner, even if arranged in different formats (not only texts characters but also digital images), direction (horizontal, vertical, oblique) types, colors and sizes. This envisioned model will grant access to the visually impaired to the automatic reading of any blob of characters forming words. It will be done by the simple movement of the mouse pointer on the screen. When moving the pointer on the computer screen over a word and hitting any character, the Global Reader will automatically identify the other characters nearby that constitute a word. By the use of an OCR (Optical Character Recognition) algorithm, this word will be automatically identified and read out aloud, by means a Speech Synthesizer. Thus, by handling the Global Reader, the blind user will also create of a cognitive map of the graphic layout of the computer screen and the spatial location of each graphical object with characters (menus, commands, controls, text fields, etc.) of the Operating System and the GUIs of any software tool. In particular, this work deals with the study of a computer vision strategy of automatic character identification, word grouping and sequential text-to-speech conversion, all performed in a dynamic random access fashion. Thus, the Global Reader will allow the visually impaired to have access and operate any graphical interface of generic computer tools such as the ones used for sound processing and music editing.

2.3 Haptic-controlled soundscapes synthesis

As described in (Shellard, 2011), soundscape is a self-organizing acoustic phenomenon from where it is possible to emerge aesthetic meaning. Soundscape is a term coined by Murray Schafer, which refers to sonic immersive environment perceived by listeners who may interact with it, either passively listening and acknowledging or actively taking the role of one of the agents of their dynamic composition (Schaefer, 1957). Thus, a soundscape is in itself also the outcome of the sound perception and cognition of the listener. The term soundscape makes a direct reference to landscape; in this case, a landscape of sounds, that can be recognized by its cognitive aspects, such as: 1) Close-up; 2) Background; 3) Outline; 4) Pace; 5) Area; 6) Density; 7) Volume; and 8) Silence. According to Schafer, soundscape may be formed by 5 sonic categories of analytical concepts. They are: 1) Tonic; 2) Signs; 3) Sound Marks; 4) Sound Objects and 5) Sound Symbols. Tonic is formed by live and ubiquitous sounds, usually perceived by the listener as sonic background. Signs are the sounds in front plan, which draws the listener's conscious attention (as they may contain important information). Sound Marks are the sounds exclusively found in each soundscape (which makes each soundscape unique). Sound Objects are the atomic components of a soundscape. Pierre Schaeffer, who coined its term, defines Sound Object as an acoustic event consisting of aspects that lead the listener on a sonic perception that is special and unique. Sound Symbols are sounds that evoke cognitive and affective responses, based on the socio-cultural and personal context of each listener.

As such, a soundscape is also a product of the listener aural ability to identify its acoustics regularities. Assuming that the mind understands, acknowledges and communicates with external reality through a dynamic process of constant mental modeling, this process can be divided in 3 levels: 1) Sensorial, where the brain receives external information through the bodily senses (hearing, vision, touch, olfaction; the senses of equilibrium, heat, acceleration, pressure, hunger, thirst, pain, etc). In music the study of acoustical sensations is handled by psychoacoustics. 2) Cognitive, where the reasoning models are generated, stored and compared with the accumulated information in memory. This is the level where the acquired information is identified, processed and appraised. 3) Affection, that deals with the emotions evoked. Emotions seem to occur as an evolutionary strategy to urge the individual to take an action (the word emotion comes from the Latin word *emovere*, that means "to move"), thus inducing this individual to ratify or reject a predefined cognitive model created from sensed stimuli. This theme aims to study strategies and computational models that will enable the visually impaired musician to exercise their enhanced aural abilities to build soundscapes, by means of haptic controllers. These musicians will map, explore and control the spatial location of sound sources, as well as other intrinsic subtle sounds aspects (pitch, loudness, timbre, tempo, reverb, etc.) in order to artistically interact and create a complex acoustic systems of sound objects required to compose truly artificial soundscapes.

2.4 Adaptive-synesthetic generative music model

Adaptive algorithms have the ability of changing their inner structure according to their input data in order to optimize (evolve) its output. In music, adaptive sound synthesis models consider the process more relevant than a final result. The dynamic path of evolution of an adaptive sound synthesis is here aesthetically more meaningful than any specific reached end. Opposing to a consequentialism approach, where "the ends justify the means", for adaptive models it might be said that "the means justify the ends".

Such models are sometimes compared with the theory of evolution, as proposed by Charles Darwin. These evolutionary computer model simulates the processes of reproduction and selection of individuals (solutions) dwelling into an artificial population. Normally such evolutionary models are used in the automatic and semi-supervised resolution of complex and open problems (Fornari, 2005). An evolutionary sound synthesis method is based on principles of Evolutionary Computation to generate sounds, as initially proposed by (Manzolli, 2001). This theme will use the methodology of evolutionary sound synthesis, as described above, not to generate new audio segments, but symbolic segments (represented by music notation or MIDI protocol). In this project the control of such evolutionary process will be handled by blind musicians, so they can exercise their enhanced aural abilities to control the parameters of an evolutionary model of music generation. This will result in the process of Musical Meta-Composition, where these symbolic segments will evolve as individuals in a population set and gather together into a complex musical structure resulting into a final musical score, which its designer will be its Meta-Composer; which is here considered as the composer of a music composition process.

3. Conclusions

This theoretical study is mainly focused on music technology developments to promote accessibility for visually impaired musicians as well as to explore their aural excellence. The intention here is to study methods and tools, as well as to analyze the perceptual, cognitive and aesthetic implications of computational models of analysis, processing and synthesis of symbolic and acoustic data, which can be used by the blind musician in contemporary music composition and performance. As described above, this work conveyed four themes of further development. This research intends, above and beyond, to provide the right of use of computer resources by the blind musicians.

The implementation of the insertion of such individuals, in academic and professional fields, will help to narrow the gap of accessibility to informational technology resources, thus promoting its social equality.

4. References

KILMER, A. D. & M. Civil (1986). *Musical Instruction Relating to Hymno*dy, Journal of Cuneiform Studies 38, pp. 94-98.

PORTUGALI, Juval (1996). *The Construction of Cognitive Maps*. Springer Science & Business Media, pp 158.

LOTFI B. Merabet, Alvaro Pascual-Leone (2010). *Neural reorganization following sensory loss: the opportunity of change*. Nature Reviews Neuroscience 11, 44-52

GOUGOUX F, Zatorre R. J., Lassonde M., Voss P., Lepore F. (2005). *A functional neuroimaging study of sound localization in early-blind individuals*. DOI: 10.1371/journal.pbio.0030027

SHELARD, Mariana, Oliveira, Luis Felipe, Fornari, José, Manzolli, Jônatas, (2005) *Abdução e significado em paisagens sonoras* Revista Kínesis, Vol. III, n° 05, Julho-2011, p. 43-67

FORNARI, José, Manzolli, Jônatas, Maia Jr.., Adolfo, (2005). *Síntese Evolutiva de Segmentos Sonoros - SESS*. In: II SEMINARIO DE MUSICA, CIENCIA E TECNOLOGIA, 1., São Paulo. Proceedings online.

SCHAFER, M., R. (1957). *The Soundscape*. Destiny Books (October, 1993). ISBN-10: 0892814551

MANZOLLI, J., Fornari, J., Maia Jr., A., Damiani F., (2001). *The Evolutionary Sound SynthesisMethod*. Short-paper. ACMmultimedia,ISBN:1-58113-394-4.USA.

BRAILLE Authority of North America (1997). *Music Braille Code*. Louisville: American Printing House for the Blind.