Mood Mapping Technologies Within Hybrid Audio Design

Lars Graugaard and Jens Arnspang

lag@cs.aaue.dk, ja@cs.aaue.dk
Department of Software and Media Technology
Aalborg University Esbjerg

Abstract

Contemporary audio design and performance systems aim at creating a world by using technology for immersion beyond the state of art of virtual reality. Such new technology might use multi sensory audio scene capture, interactive rendering of audio scenes, agent emotion algorithms, mood mapping and other system components, aiming at new design and performance tools. We would like to report research of this type. New technological development allows the audio scene designer to play much more freely when creating the narrative for a storyline.

New technology makes it possibility to evoke inner moods and inner worlds to a much larger degree than before because the imagination is free of physical constrains. Mood mapping becomes a new and challenging tool allowing the audio designer to immerse the audience into a perceptive situation using a palette of new technical possibilities. Examples are hybrid designs, involving "fabulating" spaces, "electronic soundscapes", and "interactive and adaptive audio scenes". This becomes possible by using audio augmentations, sensor technology, tracking algorithms, and software technology for hybrid design.

Based on our ongoing European projects and our new interactive designs, we would like to present such considerations and technology for mood mapping, interactive performance and audio design.

1 Attaching Musical Expressivity to A Soundscape

In this section tools for musical expressivity are discussed

1.1 Introduction

One of interactive music's strongest assets is its ability to extract musically significant elements in real-time from a performance, and inject a modeling of that essence into an accompanying digital soundscape. An approach for correlating such data to an implicit chord structure for pitch cueing and for presence augmentation of the performer is discussed, in the particular case of the interactive opera *La Quintrala* (Graugaard 2004). Further possibilities for usage in interactive instrumental music is discussed.

1.2 Particular Accompaniment Needs in The Case of The Voice

An important task for any composer of vocal music is the cueing of the voice. Singers, much more than instrumentalists, depend on certain pitches in the accompaniment in order to maintain their relative pitch. Some of the pitches in a melody will coincide with the underlying chordal structures, while other pitches would be passing-notes between pitches. The pitches that will most readily support their need for tonal orientation are those that coincide with their own pitch. Other pitches that will be supportive are those that are in a harmonic relationship, and within a reasonable proximity in terms of octave displacement. When the pitch is in some harmonic interval, it becomes an issue for the singer to be able to anticipate their own pitch, or to be able to maintain a consistent feeling of tonality (albeit not necessarily one based on functional harmony or 'correct' voice-leading), in order to maintain intonation and tonal orientation.

1.3 Expressive and Structural Data Handling

La Quintrala is an opera for five singers and interactive, computer generated sounds, with a duration of 120 minutes excluding intermission. The interactive accompaniment is generated in real-time, with sound synthesis algorithms dynamically being affected through analysis of the singing. The sound synthesis methods were chosen beforehand and change throughout the opera, whereas pitch content for the synthezised sounds partly is generated by algorithms, partly extracted from a continuous comparison of the voice to a chord structure stored in memory. This chord structure defines the primary supportive pitches, and links the notated and electronic score together, addressing the needs of the singers for tonal 'indicators' at any given moment.

The chord sequence was derived from the melodic material it supports, which in turn was developed from the requirements of the libretto in terms of dramatic and musical content. The sequence was made with an informal technique of floating pitch connection, with no functional harmony.

The harmonic structure becomes 'associative' as pitch centers attract and repel, and thereby is perceived to shape the melodic contours. The chords are independent of octave placement and inversion, and each chord usually delimit a time segment corresponding to a melodic segment that can be sustained by a set of up to five pitches.

The analysis of the voice could have been done by means of a pitch estimation algorithm, as this would have given an indication as to the fundamental of the sound, i.e. note sung. This representation would be directly applicable to the score notation of the voice and the underlying chord structure, but such a pure pitch estimation would not contain any expressive data, since this is embedded in the spectrum of the sound, and is discarded in the process of pitch estimation. A musical score is, on the same token, very limited with respect to the actual sound of music, that is, the auditory information that arrives at the listener's ears, as described by Rowe (Rowe 2001) and others. A combination of pitch estimation with some other form of further audio analysis was needed, in order to retain expressive performance data.

Exactly what kind of expressive data required could be very different things throughout the opera. The vague definition of 'expressive data' centers on flexible timing and flexible use of dynamics and articulation, and we often refer to this highly important part of music performance as 'phrasing'. A vast variety of imprecise musical terms is accounted for in score notation concerning loudness, pitch connection, articulation, timbral shading, and time contraction and expansion etc., such as mezzo-forte, legato, sul ponticello, rallentando, and so forth. Since these notations are not as easily measurable as pitch and time, they are considered to be in the domain of performers, and a performance is judged by how well the performer expressively alters the given pitches and rhythms within the nature of the composition. Paradoxically, this expressive layer carries much of the audio content of the composition, because the composition defines and delimits its own 'expressive space', as an implicit consequence of the composer's decisions of the more precise notations of pitch, time and rhythm. The 'expressive layer' is therefore a hidden yet integral part of any musical composition, and is defined at the moment of composing even though it only comes into existence at the moment of performance.

1.4 The System

The aim for the interactive relationship between the singers and the composition in *La Quintrala* consequently became twofold: comparing pitches for chordal verification and support, and projecting the singer's subjective and relative interpretation of expressive parameters onto aspects of the accompanying soundscape.

The analysis had to give justice to the perceived totality of the sound, in order to retain something close to, or at least representative of the emotional triggers which the singers embed in the sound. Such expressive information manifests itself in low-level data such as pitch and spectral fluctuations at the onset of a note, between notes, and in changes of dynamics, and this information is readily appreciated by expert and non-expert listeners alike, as the continuous, complex sound reaches their ears. Creating algorithms for extracting such expressive content of an audio signal is not a straight-forward matter. In fact, they are in some respects easily out-performed by non-expert music listeners as described by (Martin, Scheirer, and Vercoe 1998). Fortunately, *La Quintrala* is a musical composition and not an analysis tool, and handling the exceptions, that is, the sung pitches (or significant spectrum content) that didn't match the stored chords became a compositional issue.

In order to circumvent this problem a mechanism for subjective, dynamic mapping according to the evolving musical needs was devised. Expressive data is passed through the pitch estimation algorithm unaffected and parameterized separately, rather than attempting at a classification of data in some generalized way.

The pitch estimation algorithm used is an algorithm developed by Miller Puckette (Puckette 1998). It extracts a limited number of sinusoids representing the most prominent harmonics of the sound, and the advantage of the estimation algorithm is that further representation of the spectral content was readily available by following and interpreting the sinusoids as the sound evolves. The parameterization of the expressive data is done by such registration, and its subsequent mapping onto parameters of sound generative processes. High-level parameters was devised for generalized control, for instance was duration information submitted to a 'sensitivity value', where higher amplitudes in the input signal would result in either shorter (negative sensitivity values) or longer note durations (positive sensitivity values), and the continuous amplitude envelope could affect the amplitude of the individual spectral components independently, according to their spectral weight and frequency placement. Further explorations were made by using performance data for dynamically setting even simple parameters, such as the minimum-maximum pitch limits for the accompaniment. As a result of this combined approach, the accompanying electronic soundscape in *La Quintrala* became one which could combine formal, musical needs for structure and tonal support, with the sonic manifestation of immediate, expressive performance parameters, and thereby significantly enhancing the presence of the performer in the electronic accompaniment. In fact, the accompaniment was perceived by the audience as an aural manifestation of the psychological disposition of the characters, and of the emotional charge of a scene.

The reason for this is that the use of pitch estimation for chordal verification with other spectral data can be understood as a continuum between a context-dependant, strict accordance with the stored chords, and a contextindependent modeling approach of re-synthesis of the voice. This implies that the performer's expressivity and musicality is projected into the electronics, where the projection is one of interplay rather than actual control, because the underlying algorithms driving the sound synthesis are not part of the analysis process, nor of the decisions made by the composer as to how to handle exceptions encountered in the comparison between analysis and the stored chord structure. Indeed, interaction implies some degree of lack of control on part of the performer, as referred to by Lippe in (Lippe 1998). The purpose of giving presence to a performer does not coincide with giving the performer control, because the drama being developed is not determined by such local action-reaction mechanisms, but by the largerscale dealings and consequences. The singer's psychological disposition is manifested and evolves in the electronics, and the resulting expansion – or intensification – of the dramatic content multiplies the emotional substance in a way very appropriate to opera. The accompaniment is shaped by some combination of stored information pertaining to the needs of the composition and the needs of the singer for vocal cueing, and the dramatic content of the moment, as it develops on stage in interaction between the characters.

A 5.1 multi-channel sound delivery method set-up in a rectangle around the audience was chosen for *La Quintrala*. This further enhances the appreciation of the psychology of the characters, and the sensation of urgency of the emotional content. The technique suffers from a limited sweet-spot, but the absence of specific sound sources in the wave field corresponding to physical objects on stage, made the requirement for uniform sound diffusion in all listening spots less pertinent. Each audience seat of the performance space became an appropriate listening-spot, because it didn't have to be consistent with a viewpoint. Some masking of loudspeakers could be allowed, without significantly disrupting the intended musical effect (Funkhauser, Jot, and Tsingos 2002).

1.5 Other Applications

We tend to hear a harmonic in a sound as part of the total sound, with some fundamental (Moore 1990). The electronic sounds in *La Quintrala* does not make any reference to acoustic instruments, and a singer can only apply her aural experience from acoustic music to a limited degree. This means that the presence in of a particular reference pitch in a complex sound isn't necessarily very obvious to the singer, even if it is sonically emphasized. The technique of emotion-projection may be developed further with instrumental performers, even though the voice is the one instrument which offers the widest range of possible variation in timbre. Interactive instrumental music without dramatic action still has a high emotional import which isn't referred to any object or objective, but this presumes that we accept that music really is a language of emotion, primarily expressing the composer's knowledge of human feeling, as expressed by Langer (Langer 1951). I therefore intend to apply the technique of expressive projection in interactive, instrumental music, and expect to find further development possibilities in this area, which would not be possible to the same extent with vocal music.

1.6 Conclusion

Opera is a musical drama, and the feelings of love and longing, and of hope and fear are the essence of tragedy and drama. Relating the pitch of a singer to a chordal progression for sound synthesis helps the singer to get the tonal orientation necessary, but attaching the singer's expressivity enhances the digital soundscape supporting the voice. It makes good sense to project these feelings into the electronic score as they are exposed by the singers on stage, because such exaggeration of presence and mood captures the audience's attention. It furthermore becomes an effective tool for shaping the formal development of the dramatic content. In interactive instrumental music without dramatic action there may nevertheless be a high emotional import which is not referred to any subject. The technique is readily adaptable to interactive instrumental music, where attaching performance expressivity to the electronic soundscape can be used to great advantage in parallel with the independent evolution of the soundscape.

1.7 Further Development

A content-based retrieval system for analyzing the audio well be considered instead of a sinusoidal representation, as it could prove to be more effective than conventional audio representations, which only use statistical characteristics (Martin, Scheirer, and Vercoe 1998). One such system has been described in (Cai et al. 2004). They reported experiments which show that the singular value of the 'first principle component' usually is greatly higher than others for the purpose of general feature extraction. This is comparable to the pitch estimation in terms of precision of perception, yet accommodating for fluctuations resulting from the expressive musicality of the performance. It seems obvious to extend the approach to be able to handle pitch as a container of microtonal pitch information 0.5 semitones above and below a central pitch. Applying gradual controls to these containers would make it possible to gradually 'pull in' the generated frequencies towards the stored chord, and would work well in conjunction with the running pitch output. It would probably not help singers in need of a fairly exact accompaniment, but if used with an instrumentalist – where the problem of tonal orientation isn't so acute – it could prove an interesting device to explore.

2 Concerning Emotion Technology

Inspired by for example eastern cultures, where ways to dwell into human 'aura', 'karma', etc. are known, computer assisted art disciplines such as computer music have started studying ways to have a digital computer represent emotions and assist in rendering emotions; or manipulate aural stimuli, so that certain emotions may appear; or provide composers or performers with tools, so they can ‘ask’ computers to assist a performance with certain emotional cues. For an overview of such attempts and considerations, see for example (Wanderley 1997).

2.1 Emotion and Artificial Intelligence

An alternative effort for producing human behavior, or animate behavior, is found within more traditional artificial intelligence, where modeling and rendering of biological behavior is the direct goal, see for example (Pfeifer and Scheier 1999) for a comprehensive overview. This discipline could easily be taken into the domain of interactive performance, where such ‘animates’ are not necessarily self contained artificial animals or beings, but digital sparring partners for a composer, a designer or a performer. Although successful attempts of using ‘genetic algorithms’ for this purpose in computer music are reported, see for example (Schloss et al. 2001), a true agent based or even animate based approach, inspired by the community of artificial intelligence / animate behavior, is yet to be seen in full scale performance situations.

The technology should however not be problematic to include in such music connections, per se. Computer game industry has already started using the techniques, where several plays include adaptive characters, that learn to behave after their owner. The robot industry (of all) have started emulating beings, that obviously try to emulate human behavior and human emotions, see (Breazeal 2002) and (Bar-Cohen and Breazel 2003) for overviews. However, evidence is already available, where problems are being identified. For example ‘the scary valley’, identified by Masashiro Mori, shows this. People seem to like robots that mimic humans, the more the better; up to a point, where they start looking and behaving like humans too much; then people get scared and get rapidly increasing negative feelings against the robot.

2.2 Emergence of New Tools

Examples of invoking digital sparring partners for performers are already known, at least attempts have been reported.

During the Brussels MOSART project (Musical Orchestration Systems in Algorithmic Research and Technology), see (Arnspang 2003), one partner provided an interactive opera for the Salzburg Festival, where the main character was schizophrenic; at some points dark and sinister and should sing in a low pitched and sonore ways, and at other points bright and febrile and should sing in high pitched and vivid ways. Cameras watched the action of the human singer, detected his state of personality from movements, and modified the singing voice according to both composition, libretto, and actual performers choices. More detailed examples of this emerging technology, cultivated at European connected scale first in MOSART, is given in section 1 in this paper.

During the Brussels BENOGO project (Be there, no need to go there), see (Granum 2002), a connection between digital worlds and human feeling of being present is attempted. One tool, which is being discussed is mood mapping. From a designers point of view this is of paramount importance for rendering presence in a digital world. Such attempts are found at the triple point between the disciplines of human computer interaction, the discipline of scene rendering, and the vast knowledge existing in design on cues and parameters essential for scene definition and scene building. Visual and aural mood mapping is an aspect in the BENOGO project; the visual part is commented in more detail in section 3 in this paper.

Acknowledgements

We thank the Benogo project, Future and Emerging Technology (FET) programme funded by the European Union.

References

Arnspang, J., 2003. *MOSART Brussels Project*, September 2000-2003.

Bar-Cohen & Breazeal, C., 2003. *Biologically Inspired Intelligent Robots*, SPIE Press.

Breazeal, C., 2002. *Designing Sociable Robots*, The MIT Press.

Cai, R., Lu, L., Zhang, H-J., and Cai, L-H., 2004. Improve Audio Representation by Using Feature Structure Patterns; *ICASSP04*.

Funkhauser, T., Jot, J-M, Tsingos, N., 2002. Sounds Good to Me, *SIGGRAPH 2002 Course Notes*.

Granum, A., 2002. *Benogo Brussels Project*, September 2002-August 2005.

Graugaard, L., 2004. *La Quintrala*, interactive opera for five singers and computer generated accompaniment, dur. 120 minutes, premiered September 2nd 2004.

Langer, S. K., 1951. *Philosophy in a New Key*; Harvard University Press. New American Library, New York, New York.

Lippe, C., 1998. Real-Time Interaction Among Composers, Performers, and Computer Systems; *Proceedings of the 1998 International Computer Music Conference*.

Martin, K. D., Scheirer, E. D., and Vercoe, B. L., 1998. Music Content Analysis through Models of Audition; *ACMMM98*.

Moore, F. R., 1990. *Elements of Computer Music*, p. 469. Prentice Hall.

Pfeifer, R. and Sheier, C., 1999. *Understanding Intelligence*, MIT Press.

Puckette, M., 1998. Real-time audio analysis tools for Pd and MSP. *Proceedings 1998 International Computer Music Conference*, San Francisco, pp. 109-112.

Rowe, R., 2001. *Machine Musicianship*, pp. 191-193, MIT Press, ISBN 0-262-68149-8).

Schloss, A. (ed), 2001. *Proceedings on International Computer Music Conference*, Havanna, Cuba.

Wanderley, M. (ed.), 1997. *Proceedings on KANSEI Workshop, The Technology of Emotion*, Genova.