Unifying Performer and Accompaniment

Lars Graugaard

Aalborg University Esbjerg,
Department of Software and Media Technology,
Niels Bohrs Vej 6,
DK-6700 Esbjerg, Denmark
lag@cs.aaue.dk

Abstract. A unique real time system for correlating a vocal, musical performance to an electronic accompaniment is presented. The system has been implemented and tested extensively in performance in the author's opera 'La Quintrala', and experience with its use in practice is presented. Furthermore, the system's functionality is outlined, it is put into current research perspective, and its possibilities for further development and other usages is discussed. The system correlates voice analysis to an underlying chord structure, stored in computer memory. This chord structure defines the primary supportive pitches, and links the notated and electronic score together, addressing the needs of the singer for tonal 'indicators' at any given moment. A computer-generated note is initiated by a combination of the singer – by the onset of a note, or by some element in the continuous spectrum of the singing – and the computer through an accompaniment algorithm. The evolution of this relationship between singer and computer is predefined in the application according to the structural intentions of the score, and is affected by the musical and expressive efforts of the singer. The combination of singer and computer influencing the execution of the accompaniment creates a dynamic, musical interplay between singer and computer, and is a very fertile musical area for a composer's combined computer programming and score writing.

1 Introduction

Alignment of score and electronics in real time became a research area in 1984 when Barry Vercoe and Roger Dannenberg independently presented their seminal papers on this concept [33, 9, 10]. Much work has been done since then [7, 27], and recent new directions have been in combination with the much larger research area of affective computing, such as correlating intended emotion to musical analysis in real time emotion tracking [13, 17].

In this paper a practical implementation of a new concept of correlating a performer to an accompanying electronic soundscape is presented, where timbre and expressivity of a performer is analyzed for subsequent use in sound synthesis. The system addressed practical needs in an artistic project. Proven and reliable algorithms and concepts of data handling from previous research was used upon due scrutiny,

R. Kronland-Martinet, T. Voinier, and S. Ystad (Eds.): CMMR 2005, LNCS 3902, pp. 169 – 184, 2006.
© Springer-Verlag Berlin Heidelberg 2006

such as spectral peak estimation [30, 20] and high-level pitch manipulation as sets of pitch-classes [11].

It was not the intention to build a new system from bottom up, and the scientific contribution of this paper reside in describing the first implementation of a generalized system for sonically unifying a performer in real time with an electronic soundscape intended to support and cue the performer, while at the same time enhancing the performer's affective presence in the accompaniment. Specifically,

- The paper is a contribution to music intelligence as a subset of machine intelligence, and it is shown that a computer programme flexibly can adapt a soundscape to a performer in real time.
- The affective content of a musical performance can be used to enhance sound synthesis, without necessarily imprinting the spectral content of the performance analysis.
- The system is based on a structure whereby a performance and its accompanying electronic soundscape can correlate with a predefined informal chordal structure, for instance notated in a performance score.

The report of the artistic project that gave rise to the system is encompassed as one example of its use, and it is seen that the system can be used as a general tool when coordination of an underlying chordal structure in a musical score with a performer's expressivity is called for.

The artistic contributions are best appreciated by listening to excerpts of the performances at www.graugaard-music.dk/pages/la_quintrala_ex.html. Musical quality is of a highly subjective nature, but the system presented in this paper can be considered successful in regards with producing an accompaniment that has proved its stated purpose through many performances.

2 Background

In 2002 I was commissioned to compose an opera for five singers and interactive computer as a joint Danish, Swedish, and German commission through EU's Culture 2000 Foundation. It was to be scored for five singers and interactive, computer generated music, and it was to have a duration of two hours. The opera was later entitled 'La Quintrala' after its main character [15], and it was premiered in Copenhagen on September 2nd 2005 at Den Anden Opera. A subsequent 28 performances took place in Sweden and Germany, with the last performance in Sweden on November 11th 2005.

While composing the vocal parts, it became clear to me that a versatile and flexible tool for unifying the electronic accompaniment with the vocal lines and for supporting tonal structure was needed. Specifically, it was necessary to tonally bind the singers to their accompaniment by cueing their pitches and providing supportive harmonic structures, to sonically unify their voices with the electronics, and to provide sufficient compositional freedom for serving the dramatic and musical needs of the composition. No existing software tools seemed to serve my purpose.

3 Functionality Requirements

Opera is a musical drama where feelings of love, longing, hope, fear, and hatred are the essence of tragedy and drama. The electronic score for ‘La Quintrala’ would have to be able to project emotions into the soundscape as they would be exposed by the singers on stage. But the soundscape would also have to be able to relate the pitches of a singer to a chord succession to be synthesized, thereby providing the singers with tonal orientation. It is hard not to see these two requirements as opposites: the needs of a singer at a given point can be very different from the musical needs according to the composer. This, however, is not a characteristic of interactive music as such, yet attempting to meet these two requirements could turn the soundscape into an effective, compositional tool for shaping the formal development of the dramatic content.

Questions of timing and event coincidence needed to be decided on at an early stage as it would influence the system design. My interactive music generate exactly timed events ‘on the fly’ by letting the performance analysis affect the chosen synthesis method(s). I therefore didn’t need an accompaniment tool that would emphasize exact timing of events or action advance based on timing prediction, but I would need as low a latency as possible to get good time coincidence. This is fundamentally a compositional issue, to have the performer define the perceived rhythm periodicity. When the rhythm texture is sufficiently complex to be ambiguous in itself, yet sufficiently simple to maintain the feeling of pulse, then the performer would ‘define’ the periodicity through the periodicity of the sung melismas. So, rather than classical score following, what I needed can be termed score correlation.

4 Extracting Expressive Data

The analysis of the voice could have been done with pitch estimation only, as this would have given information of the fundamental of the note sung. This representation would be directly applicable to the score notation of the voice and the underlying chord structure. But 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 the music notated, that is, the auditory information that arrives at the listener’s ears [32].

The analysis had to provide an indication of the emotional triggers which the singers would embed in their vocal performance. Expressive information manifests itself in fluctuations of spectrum components at the onset of and during a note, in changes of dynamics, and in spectral and amplitude contours. Expressive information is readily appreciated by expert and non-expert listeners alike, as the continuous, complex sound reaches their ears. But extracting such expressive content of an audio signal is not easily done. In fact, algorithms created for this purpose are in some respects still easily outperformed by even non-expert music listeners as described in [26]. Fortunately, ‘La Quintrala’ is a musical composition and non-utilitarian, and not an analysis tool. Handling the expressive data became an issue of finding the best possible way to have the data influence and enhance the soundscape, based on my subjective, compositional judgment.

Exactly what kind of expressive data I needed to access for would vary considerably 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 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, tenuto, lontano, 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 appreciable 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 but integral part of any musical composition, and even though it is defined at the moment of composing it only comes only into existence at the moment of performance. We

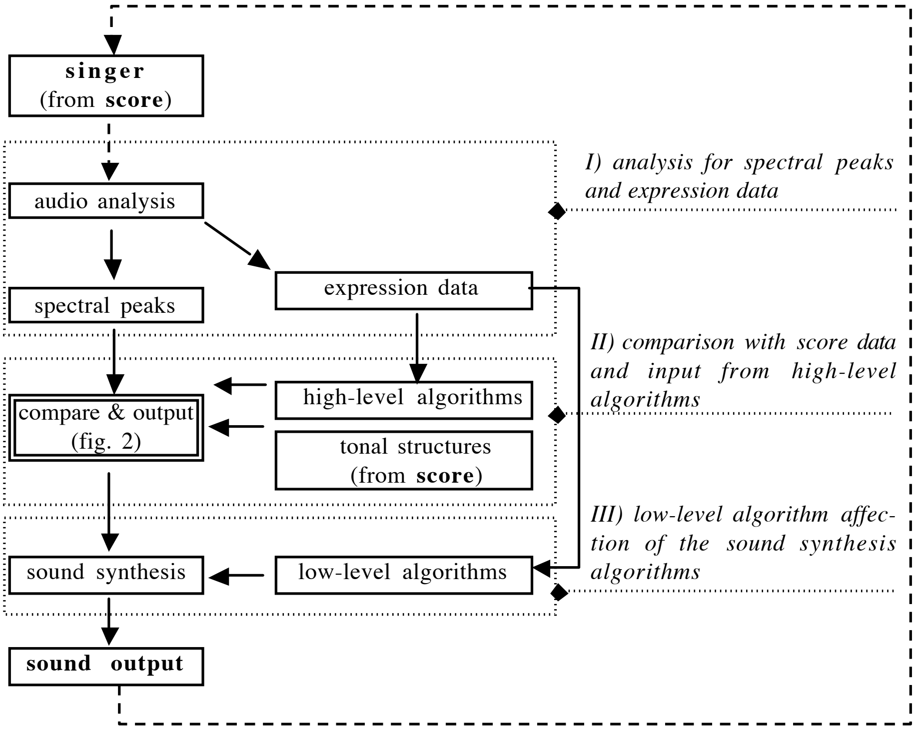

Fig. 1. Overall data fow (dashed arrows represent sound)

rely on performers to add this layer to their interpretation of a score, and it would be of major importance for me to extract and apply in some form this expressive data to the accompanying soundscape throughout the opera.

5 The System

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

The system is divided into three parts, I) analysis for spectral peaks and expression data, II) comparison with score data and input from high-level algorithms, and III) low-level algorithm affection of the sound synthesis algorithms (fig. 1).

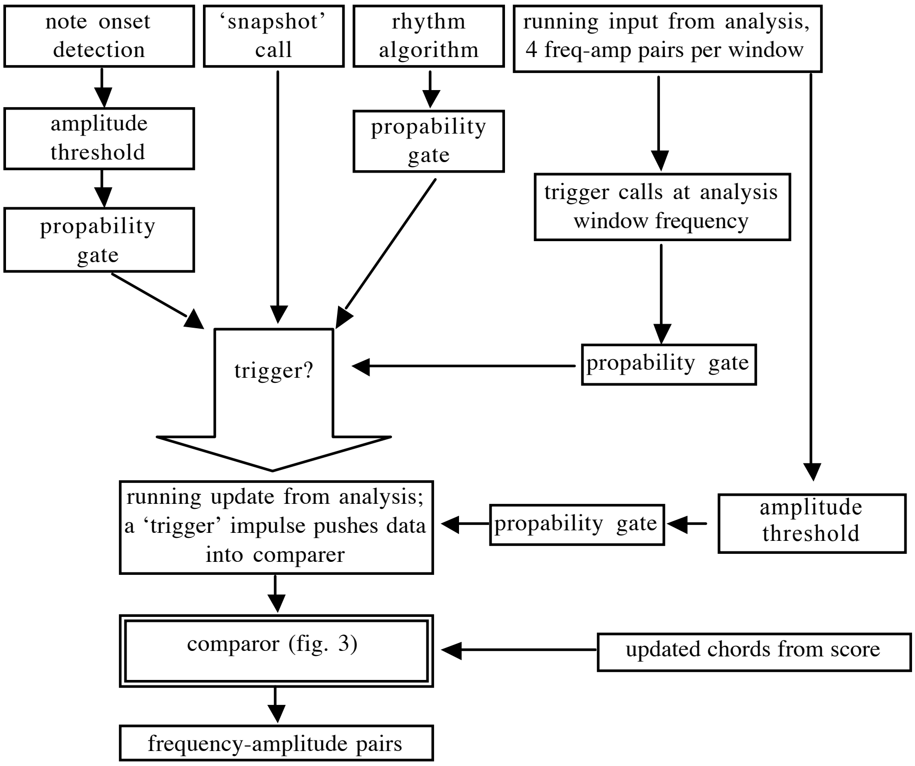

Fig. 2. Input-output flow for comparison of spectral peaks

5.1 Analysis

The analysis phase outputs the four most prominent spectral peaks as frequency-amplitude pairs, an impulse upon note-on detection, a continuous amplitude envelope and a fundamental pitch estimation. The original algorithm used is described in [30]. The spectral peaks are gated with an amplitude threshold and a probability gate, and the running update can – but may not necessarily, depending on the settings of these gates – store new values at analysis window frequency, for possible passing on to the comparer.

5.2 Data Triggering

Input to the comparer is triggered in parallel by four methods, individually controlled in terms of density, and combined into one output (fig. 2):

1. analysis window frequency,
2. a rhythmizised trigger,
3. by note onset detection, and
4. single comparison trigger (a 'snapshot'), called from within the application.

Triggering at analysis window frequency is basically a simple, additive resynthesis approach with four waveforms. The rythmizised trigger is produced by an algorithm generating onset pulses at timed intervals, with control of regularity of the pulse and probability of the trigger impulse taking place. Note onset triggering is determined by changes in amplitude or fundamental and passed through a threshold and probability gate, while single comparison triggers are called from within the application when required.

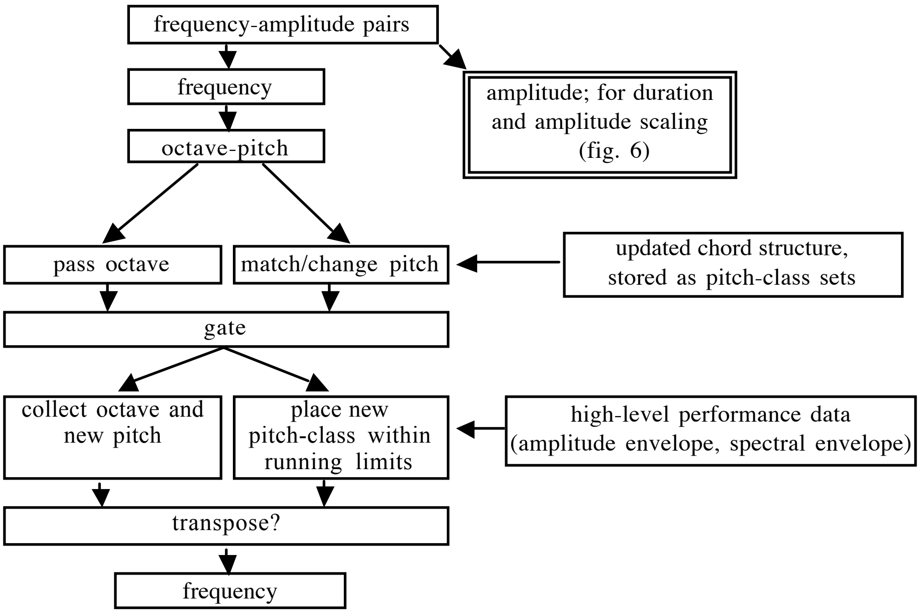

Fig. 3. The comparor

When a trigger is effectuated, it is pushes the most recent frequency-amplitude pairs received from the running analysis into the comparer for parsing, which in turn will return new frequency-amplitude pairs.
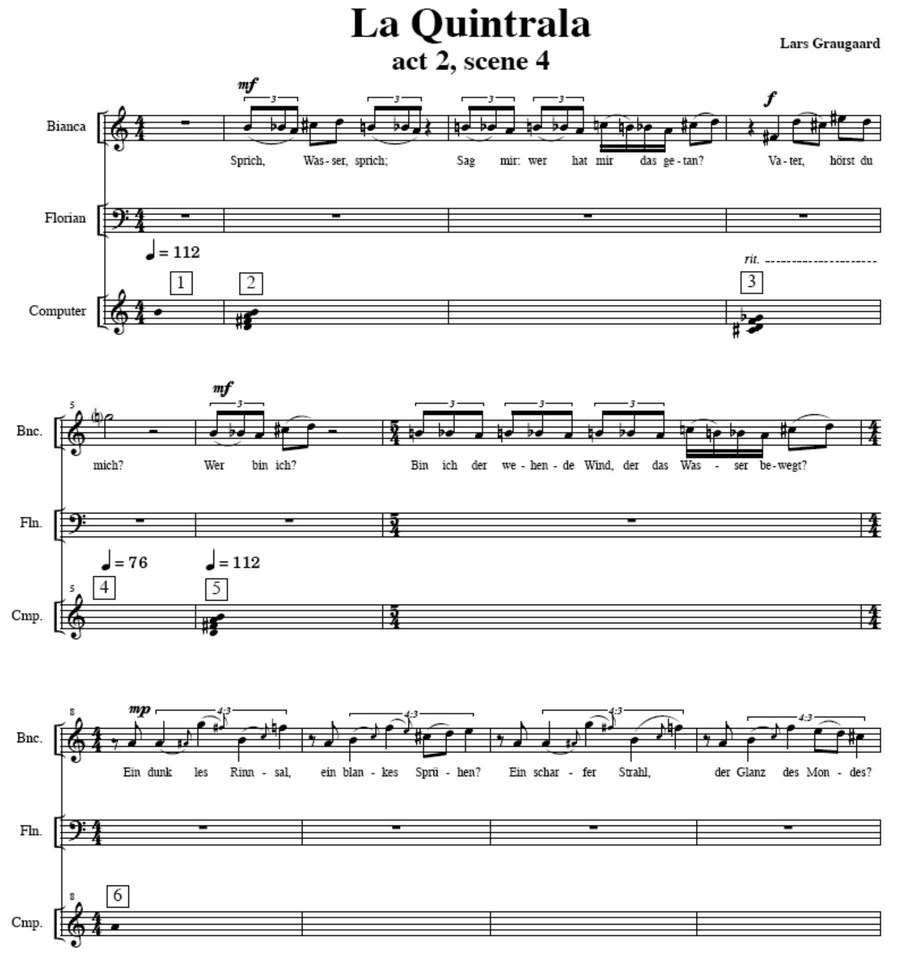
Fig. 4. Score example from La Quintrala, third staff is supporting pitch-classes

5.3 Data Comparison

The second part parses the triggered frequency-amplitude pairs based on the stored pitch-classes and a set of parameters. The frequencies from the analyzer are converted to midi pitches and octave and pitch class are separated (fig. 3). The octave is passed, and the pitch is matched to a reference pitch-class set. This pitch-class set is stored in memory, and passed to the comparer as the events in the score are advanced by hand in performance.

The stored pitch-classes define which pitch-classes are the basis of the melodic material at those particular bars of the accompanying score (fig. 4). The pitch-class sets have been defined beforehand as a chordal structure that supports a given subsection of the melody. I derived these chordal structures from the melodic material they supported, which in turn was developed from the requirements of the libretto in terms of dramatic content and musical needs. The structures were made with an informal technique of floating pitch connection, with no functional harmony. The chords can be seen as chroma, i.e. pitch-classes independent of octave placement and inversion, and each chord usually delimits a time segment corresponding to a melodic segment that can be sustained by a set of up to five pitches.

The comparison produces a pitch-class output describing the best musical match to accompany the singing. The best match is modified by an index so as to make it possible to vary the consonance/dissonance relation between singer and soundscape/accompaniment, since the best musical match for compositional reasons wouldn’t necessarily be the most consonant match. The resulting pitch-class is then either re-collected with the octave information, or the octave information is discarded and the pitch-class is placed within a running low-high pitch limit. The harmonic structures become ‘associative’ as pitch centers attract the melody and thereby is perceived to shape and guide the melodic contours. Equation (1) handles this, as well as micro-tonal deviations according to a dissonance factor.

$$\begin{gathered} y = \mathrm{int}(x) + d^{*}x \bmod 1 \text{ for } 0. \leq x \bmod 1 < 0.5 \\ y = \mathrm{int}(x) + 1 - d^{*}\mathrm{abs}(x \bmod 1 - 1) \text{ for } 0.5 \leq x \bmod 1 < 1\ . \end{gathered} \tag{1}$$

where *d* is dissonance factor.

The comparisor handles pitch-classes as a container of microtonal pitch information 0.5 semitones above and below its center. Applying gradual controls to these containers makes it possible to ’pull in’ the generated frequencies towards the stored chord, or to ‘release’ them. This works well in conjunction with running pitch output, even though it defeats the tonal precision of the accompaniment.

The running low-high pitch limit is correlated to the running amplitude envelope with parameters for center, follow, and range (equation 2). The center value follows the amplitude envelope either directly or inversely according to the follow value, which specifies the center offset at maximum amplitude.

$$\begin{gathered} c + (a^{*}f) + 0.5^{*}r = \text{high limit} \\ c + (a^{*}f) - 0.5^{*}r = \text{low limit}\ . \end{gathered} \tag{2}$$

where c is pitch center, r is pitch range, a is amplitude envelope, and f is the follow factor expressed as the ratio pitch-steps/max-amplitude.

The singer can hereby affect the pitch range of the output of the comparer in performance, and the code for equations (1) and (2) can be seen in (fig. 5).

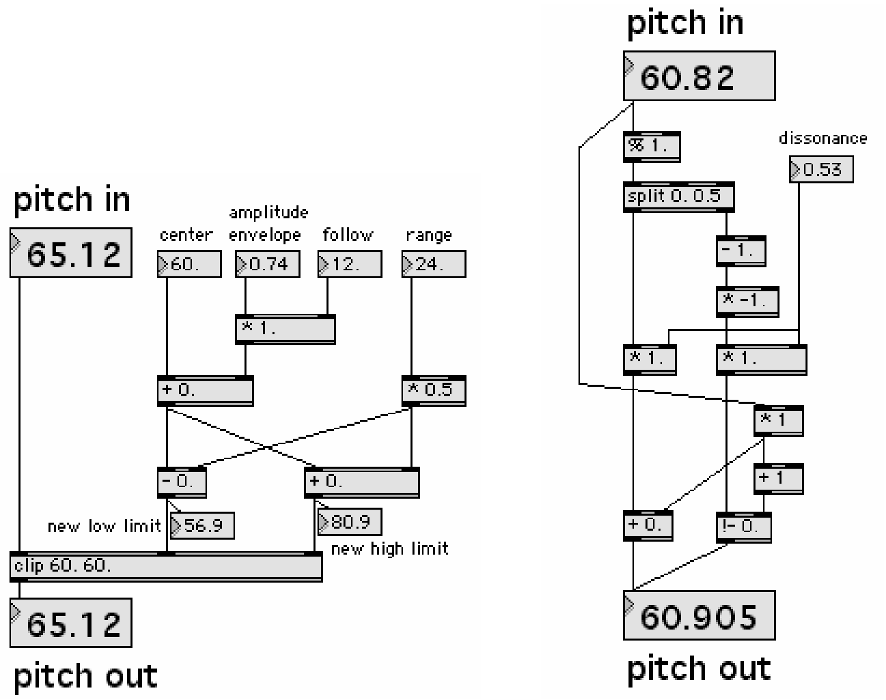

Fig. 5. Code examples of equations (1) and (2)

5.4 Amplitude as Expressive Parameter

The amplitude obtained during analysis is also applied to note duration and base amplitude of the corresponding frequency-amplitude pair from the comparer (fig. 6). It made sense to have the ability to apply amplitude envelopes to the comparison results, because this would make it possible to affect sound synthesis taking place elsewhere in the application, provided it recognized such data. It would serve the need for subjective, dynamic mapping of expressive data according to the evolving musical needs as I perceived them. Such expressive data doesn't interfere with the pitch estimation algorithm but affects the duration and amplitude of each comparison result on an ad-hoc basis (as well as the pitch range, as described by equation 2). High-level parameters were devised for generalized control, so that duration information was submitted to a scaling value where positive scaling would yield longer durations for higher amplitudes, zero would be unity duration, while negative scaling value would yield shorter durations.

5.5 Direct Audio Output

Finally, a direct output was added whereby each comparison result would be played back as an audio waveform, using the above mentioned amplitude envelope.

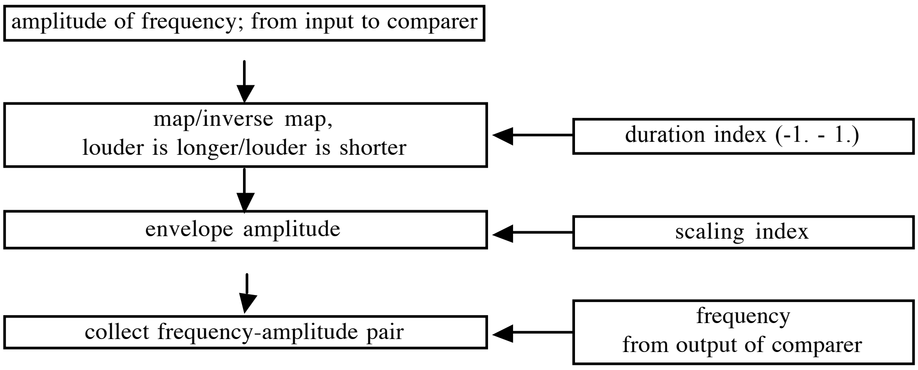

Fig. 6. Duration scaling by amplitude

Parameters are exposed for switching waveform (not necessarily being a sinusoid), setting the amplitude envelope, and for spatial placement in a four channel XY-field.

5.6 The System in Performance

The system shapes the accompaniment by some combination of stored information pertaining the requirements of the composition together with the needs of the singers for vocal cueing, and the dramatic content of the moment as it develops on stage in interaction between the characters. As a result of this combined approach, the accompanying electronic soundscape in ‘La Quintrala’ could reconcile formal, compositional needs for structure and tonal support with sonic manifestation of immediate emotion expressed during performance by the singers. Furthermore, the possibility to make gradual transitions between these two requirements provided flexibility in handling transitions between different musical contexts. The combination of sound algorithm and analysis data affecting the spectral content of the synthesis proved an effective way to enhance the presence of the performer in the electronic accompaniment and project the singer’s musicality into the accompaniment. In fact, the accompaniment in ‘La Quintrala’ was very often described to me by members of the audience as an immediate, aural manifestation of the psychological disposition of the singers, and of the emotional charge of a scene.

The singers’ expressivity and musicality projected into the electronics of ‘La Quintrala’ is one of interplay rather than actual control, since the algorithms driving the sound synthesis as well as my decisions on handling the correlation exceptions of analysis output versus stored chordal structure are partially hidden to them. This should not cause concern, because interaction implies some degree of lack of control on part of the musical performer [25]. The purpose of giving presence to the singers did not coincide with giving (full) control over the electronic score, since the drama being developed isn’t determined by such local action-reaction mechanisms, but by the larger-scale dealings and consequences. The singers’ psychological dispositions are 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.

6 Research Background

The system relates primarily to score following, signal decomposition, performer-computer interaction, interactive music accompaniment systems, and presence enhancement and mood creation.

6.1 Signal Decomposition

The score correlation method presented doesn't include original work on signal decomposition, but it was important to choose the proper method for audio analysis. The algorithm used is described in [30], and subsequently AltiVec optimized and further developed by Jehan et al. [20], even though the authors don't specify how the algorithm was enhanced. The analysis algorithm outputs doesn't distinguish between sinusoidal and noise component, but it performs reasonably well for my purpose, since the signal analyzed in 'La Quintrala' is known to be a – mostly – periodic signal.

Signals with a large noise component or in circumstances of low signal-to-noise ratio are not well handled by pure spectral peak estimation. A procedure is proposed in [16] for preanalyzing a signal for spectral peaks corresponding to true sinusoidal components, whereby the influence of noise on the analysis result is considerably reduced. The technique seems presently too slow for real time application. An approach for classifying the spectral peaks into noise and sinusoidal peaks is presented in [31], where the authors suggest to limit the number of candidate peaks in order to reduce the computational cost. This approach is relevant to my purpose, as I only need a limited number of peaks.

Conventional content-based audio representations use statistical characteristics, but this has limitations in terms of content representation as described in [26]. A content-based retrieval system has been described in [1] and the authors 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, and accommodates for fluctuations resulting from the expressive musicality of the performance. Such analysis methods would enhance the decomposition for extracting high-level expressivity data.

The need for real time analysis puts restrictions on the algorithms. Not all approaches mentioned are therefore feasible at this moment, even if they suggest results that could be very useful for my purpose. But given the development in computational power and in the algorithms themselves, it must be expected that real time analysis methods will develop further features such as those mentioned.

6.2 Score Following

Score correlating in 'La Quintrala' has the purpose of cueing and tonally supporting a singer, by relating the spectral content to a stored chord, and to sonically enhance the presence of the singer in the accompaniment. Score correlating is related to score following, and it was indeed my intention to include score following functionality in 'La Quintrala'. Score following gives score position and future tempo estimate, for advancing the computer's reading of the electronic score in synchronization with the performance of the music. For this purpose score following attempts to transform the performance into a real-time sequence of note onsets and their corresponding discrete pitch.

Score following was first presented by Barry Vercoe and Roger Dannenberg in 1984 [34, 9, 10]. Later systems pioneered at IRCAM [28] tried to minimize system latency rather than predicting tempo, since steadfast tempo often would be absent from performance, or possibly only vaguely implied. One of the first examples of these efforts was Philippe Manoury's composition 'Jupiter' for flute and interactive computer. It consists of a great many events to be automatically triggered through the proper advancing by the score following algorithm. This works fairly well, but not sufficiently reliable to avoid human supervision during performance. But while the flute is a relatively simple signal to track, it was of concern to me that the singing voice is a particularly difficult subject for such detection [28]. The voice is an extremely flexible 'instrument' characterized by timbral and dynamic richness – the voice has a dynamic range of up to 55dB – and capable of a great many expressive performance parameters such as vibrato and glissandi, let alone the particular influence of the sung text. All these performance 'artifacts' make fast and accurate onset and pitch detection for score following purposes unreliable, if at all possible. For this reason any score following functionality was eventually left out of 'La Quintrala', since the purpose was to rely solely on the follower to advance the events.

A method for tracking the score position of a singer based on stochastic procedures was later presented in [14]. In order to increase reliability, the authors suggested to extend the score following model to include features other than fundamental pitch. Tracking through stochastic modeling a variety of performance data using hidden markov models was presented in [7] and [27]. In the latter system, the authors treated the problem as a subclass of sequence alignment, and expanded on techniques first developed in speech recognition and in molecular genetics. A technique for use with nonlinear, open-form score notation has been described in [23]. The follower was mainly used for audio routing and mixing purposes, based on the 'crossroad' concept known from some open-form compositions. The linearity of standard score notation was avoided by using a noise-reduced confidence-index to locate the tracker in the sections. The index would then be used for making global changes at signal routing and mixing level, but also of signal processing algorithms.

Score following in its extreme is an automated accompaniment system, and attempts have been made at refining and commercializing such systems. [18], [21], and [19] describe systems that accompany amateur vocalists performing pop music. The first two rely on pitch detection for tracking the performer, while the last applies speech processing techniques for vowel recognition. The systems attempt to identify both the score position and the tempo of the performer, and to adjust the computer accompaniment in response. CODA Music Group introduced with SmartMusic™ a commercial accompaniment system for amateur musicians.

The latest score following algorithms have reached a fair degree of accuracy, but the somewhat more simple score following algorithms I considered for 'La Quintrala' are prone to errors. On the one hand the performer may make errors in performance, while real time audio analysis algorithms on the other hand aren't fail-safe concerning pitch estimation. And if the musician falters then it is of little comfort if the following algorithm works correctly. Consequently, it becomes mandatory to compose for easier performance and score following, such as avoiding writing which cannot be accurately followed because it doesn't translate easily into simple notation, or avoid passages which are of great difficulty in interpretation or execution. The compositional consequences of these stability issues made me refrain from including score following in 'La Quintrala': it would limit the score notation unacceptably (and hence the musical possibilities), and it seemed moot when the performance in any case would have to be supervised, due to the risk of failings of the tracking.

6.3 Mood Creation and Presence Enhancement

Most score following systems are concerned with extracting pitch and duration of a musical performance. Yet expressivity is one of the salient features in music appreciation. It has been my intention to map expressiveness onto the electronic soundscape through mapping salient performance aspects onto high-level output parameters. Observations are made in [8] concerning how parametrization is capable of trivializing or enhancing interactivity in a human-machine relationship, and mood creation through adding expressiveness to an automatic musical performance is described in [4]. In [3] a performance was analyzed and the data applied to a computer-generated performance, resulting in robust detection of the emotional intention by expert listeners. It could therefore be considered to map such data onto synthesis algorithms, and signal routing and mixing.

Real time emotion tracking is described in [13], where a limited set of cues can accurately predict a set of emotional expressions, without using any score information. The system compares the extracted cues to a previously stored reference input, and the strong intercorrelation of the cues in a given emotion makes for quite accurate prediction. The authors use the output for a graphical representation of the intended emotions, but this data could just as well be used for high-level control of synthesis algorithms, and signal routing and mixing. Another system, designed to work with MIDI instruments, is described in [6]. The basis is a 'Perceptual Parametric Space' described in [5] which relate sets of coefficients to acoustic quantities. The system would require accurate fundamental pitch estimation, note duration, etc. in order to work with an audio signal.

Emotion as acquired sensibility towards art and music is presented with reference to its Japanese word 'kansei' in [17]. The author describes this as the third target of information processing referring to feelings, intuition, and sympathy, while the second target is semantic symbol processing, and the first target being the physical signal. Contributions in the area of 'kansei' encompassing not only music are described in ao. [2]. Emotion can be broken down to a fairly limited set of parameters, and following these parameters have proven to give fairly precise results for predicting the emotion intended. I have tentatively interpreted selected performance parameters and projected them onto synthesis algorithms, signal routing and mixing, and other aspect of an expressive space. This seems as a fertile area which requires a fine balance between the trivializing and enhancing, the obvious and obscure, and the traceable and multidimensional [24].

7 Conclusions, Further Development, and Other Applications

The described performer-accompaniment unification approach proved very effective in musical drama. Projecting emotions into an accompanying soundscape as they are exposed by the singers on stage enhances the audience's appreciation of their presence and mood which in turn enhances the dramatic content and development, strongly engaging the audience's attention.

A content-based system for analyzing the audio as suggested will be considered.

Interactive instrumental music without dramatic action may contain a high emotional, yet abstract impact, 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. This presumes that we accept that music really is a language of emotion, primarily expressing the composer's knowledge of human feeling, as expressed in [22]. The performer-accompaniment unification approach seems therefore readily adaptable to interactive instrumental music as well, where attaching performance expressivity to the electronic soundscape in parallel with more 'autonomous' evolution of the soundscape could be used to advantage. I therefore intend to apply the technique of expressive projection in interactive, instrumental music, and expect to find further development possibilities in this area, and most likely quite different from those explored in 'La Quintrala'.

References

1. Cai, R., Lu, L., Zhang, H-J., and Cai, L-H.: Improve Audio Representation by Using Feature Structure Patterns. ICASSP04 (2004).
2. Camurri, A., Trocca, R., Volpe, G.: Interactive systems design: A KANSEI-based approach. In: Proceedings of NIME2002. Dublin, Ireland (2002).
3. Camurri, A., Dillon, R., Saron, A.: An Experiment on Analysis and Synthesis of Musical Expressivity. In: Proceedings of the XIII Colloquium on Music Informatics. L'Aquila, Italy (2000).
4. Canazza, S., Rodá, A.: Adding Expressiveness in Musical Performance in Real Time. In: Proceedings of the AISB 1999, Symposium on Musical Creativity, pp. 134-139. Edingburgh, Scotland (1999).
5. Canazza, S., De Poli, G., Rodá, A., Vidolin, A., Zanon, P.: Kinematics-energy space for expressive interaction in music performance. In: Proceedings of MOSART, Workshop on current research directions in Computer Music, pp. 35-40. Barcelona, Spain (2001).
6. Canazza, S., De Poli, G., Rodá, A., Soleni, G., Zanon, P.: Real Time Analysis of Expressive Contents in Piano Performances. In: Proceedings of the 2002 International Computer Music Conference, pp. 414-418. Gothenburg, Sweden (2002).
7. Cano, P., Loscos, A., Bonada, J.: Score-performance matching using HMMs. In: Proceedings of the 1999 International Computer Music Conference. Beijing, China (1999).
8. Choi, I.: Interactivity vs. control: human-machine performance basis of emotion. In: Proceedings of 1997 'KANSEI - The Technology of Emotion Workshop', pp. 24–35 (1997).

9. Dannenberg, R. and Mukaino, H.: New Techniques for Enhanced Quality of Computer Accompaniment. In: Proceedings of the 1988 International Computer Music Conference, pp. 243-249. Cologne, Germany (1988).
10. Dannenberg, R.: An On-line Algorithm for Real-Time Accompaniment. In: Proceedings of the 1984 International Computer Music Conference pp. 193-198. Paris, France (1984).
11. Forte, A.: The Structure of Atonal Music. Yale University Press. New Haven and London (1973).
12. Freescale Semiconductor: Altivec technology; www.simdtech.org/altivec
13. Friberg, A., Schoonderwaldt, E., Juslin, P. N., Bresin, R.: Automatic Real-Time Extraction of Musical Expression. In: Proceedings of the 2002 International Computer Music Conference, pp. 365-367. Gothenburg, Sweden (2002).
14. Funkhauser, T., Jot, J-M, Tsingos, N.: Sounds Good to Me. SIGGRAPH 2002 Course Notes (2002).
15. Grubb L., Dannenberg, R.: A stochastic method of tracking a vocal performer. In: Proceedings of the 1997 International Computer Music Conference, pp. 301-308. Thessaloniki, Greece (1997).
16. Graugaard, L.: La Quintrala. Opera for five singers and interactive, computer generated accompaniment. Premiered September 2nd 2004 at Den Anden Opera, Copenhagen, Denmark (2003-04).
17. Hanna, P., Desainte-Catherine, M.: Detection of sinusoidal components in sounds using statistical analysis of intensity fluctuations. In: Proceedings of the 2002 International Computer Music Conference, pp. 100-103. Gothenburg, Sweden (2002).
18. Hashimoto, S.: KANSEI as the Third Target of Information Processing and Related Topics in Japan. In: A.Camurri (Ed.) Proceedings of the International Workshop on KANSEI: The Technology of Emotion, pp.101-104 (1997).
19. Inoue, W., Hashimoto, S., Ohteru, S.: A computer music system for human singing. In: Proceedings of the 1993 International Computer Music Conference, pp. 150-153; Tokyo, Japan (1993).
20. Inoue, W., Hashimoto, S., Ohteru, S.: Adaptive karaoke system—human singing accompaniment based on speech recognition. In: Proceedings of the 1994 International Computer Music Conference, pp. 70-77. Århus, Denmark (1994).
21. Jehan, T., Schoner, B.: An Audio-Driven Perceptually Meaningful Timbre Synthesizer. In: Proceedings of the 2001 International Computer Music Conference. Havana, Cuba, (2001)
22. Katayose, H., Kanamori, T., Kamei, K., Nagashima, Y., Sato, K., Inokuchi, S., Simura, S.: Virtual performer. In: Proceedings of the 1993 International Computer Music Conference, pp. 138-145. Tokyo, Japan (1993).
23. Langer, S. K.: Philosophy in a New Key. Harvard University Press. New American Library, New York, New York
24. Lawter, J., Moon, B.: Score following in open form compositions. In: Proceedings of the 1998 International Computer Music Conference,. San Francisco, USA (1998).
25. Lippe, C.: A Look at Performer/Machine Interaction Using Real-Time Systems. In: Proceedings of the 1996 International Computer Music Conference, pp. 116-117. Hong Kong (1996).
26. Lippe, C.: Real-Time Interaction Among Composers, Performers, and Computer Systems. In: Information Processing Society of Japan SIG Notes, Volume 2002, Number 123, pp. 1-6. (2002)
27. Martin, K. D., Scheirer, E. D., and Vercoe, B. L.: Music Content Analysis through Models of Audition. ACMMM98 (1998).

28. Orio, N., Lemouton, S., Schwarz, D.: Score following: state of the art and new developments. In: Proceedings of the 2003 Conference on New Interfaces for Musical Expression. Montreal, Canada (2003).
29. Puckette, M., Lippe, C.: Score following in practise. In: Proceedings of the 1992 International Computer Music Conference, pp. 182-185. (1992)
30. Puckette, M.,: Score following using the sung voice. In: Proceedings of the 1995 International Computer Music Conference. Banff, Canada (1995).
31. Puckette, M., Apel, T., Zicarelli, D.: Real-time audio analysis tools for Pd and MSP. In: Proceedings of the 1998 International Computer Music Conference, pp. 109-112. San Francisco, USA (1998).
32. Roebel, A., Zivanoviz, M., Rodet, X.: Signal decomposition by means of classification of spectral objects. In: Proceedings of the 2004 International Computer Music Conference, pp. 446-449. Florida, USA (2004).
33. Rowe, R.: Machine Musicianship, pp. 191-193. MIT Press, ISBN 0-262-68149-8 (2001).
34. Vercoe, B.: The Synthetic Performer in the Context of Live Performance. In: Proceedings of the 1988 International Computer Music Conference, pp. 199-200. Paris, France (1998).