Deep Surrender: Musically Controlled Responsive Video

Robyn Taylor[1] and Pierre Boulanger[2]

[1] Advanced Man-Machine Interface Laboratory,
Department of Computing Science, University of Alberta
T6G 2E8 Edmonton, Alberta. Canada
robyn@cs.ualberta.ca
[2] pierreb@cs.ualberta.ca

Abstract. In this paper we describe our responsive video performance, *Deep Surrender*, created using Cycling '74's Max/MSP and Jitter packages. Video parameters are manipulated in real-time, using chroma-keying and colour balance modification techniques to visualize the keyboard playing and vocal timbre of a live performer. We present the musical feature extraction process used to create a control system for the production, describe the mapping between audio and visual parameters, and discuss the artistic motivations behind the piece.

1 Introduction

We have used Cycling '74's Max/MSP and Jitter packages [1] to create a visualization environment responsive to vocal timbre and piano chord manipulation. This visualization system was used to create an interactive performance called *Deep Surrender*, a multimedia piece written for soprano, synthesizer, and responsive video (see Figure 1.)

Data from live musical performance is extracted and used to control parameters in video processing routines. The video manipulation is, therefore, responsive to the musician's performance, allowing the media piece to be dynamic and expressive of the nuances of live performance. This project creates an interaction platform that a musical performer can use to manipulate visual imagery in a fluid and natural way.

Responsive artistic spaces have been created by Ox [4] who visualized harmonic relationships in musical pieces through the use of colour, and Oliver *et al.* [3] who interpreted sung vocalizations through responsive audio-visual feedback. Our system uses similar strategies to map chord data and vocal information to responsive imagery in order to create a visualization environment that enhances our specially composed audio-visual performance piece.

This document outlines the process used to extract musical feature data from a live musical performance, the mapping between musical features and visual elements in the performance environment, and the artistic concept behind the *Deep Surrender* production.

Fig. 1. A performance of *Deep Surrender*

2 Musical Feature Extraction

Our musical feature extraction routines are encapsulated into a Max/MSP patch called the Musical Perception Filter Layer. The patch analyzes a live musical performance in real-time and communicates vocal pitch, amplitude and timbral information as well as information about chords played on a digital piano. We developed this patch as part of our previous work, where extracted musical data was used to control the behaviour of responsive avatars [7] [8].

Inside the Musical Perception Filter Layer, MIDI events are monitored in order to identify the chords that are played on the keyboard. Vocal pitch and amplitude information is interpreted from sung input using the functionality provided by the `fiddle~` object created by Puckette *et al.*[5]. The `fiddle~` object also outputs the raw frequency and amplitude data describing each of the partials forming the harmonic spectrum of the user's singing voice. Examining the amplitude of the energy found at each of these partial frequencies allows us to view a measure of a singer's timbre.

In our production, we use the extracted piano chord data and information about the peformer's vocal timbre to modify the colours in the *Deep Surrender* video stream.

3 Mapping Vocal Timbre to Colour

A sung sound consists of a *fundamental frequency* and a series of *harmonics* or *partials* that exist at multiples of the fundamental frequency. In the *Deep Surrender* visualization, vocal timbre is illustrated by visually representing the distribution of energy amongst partials in the sound, thereby creating a parameter that a singer controls by modulating her vocal timbre.

We have chosen to map the first three amplitudes output by the `fiddle~` object (the amplitudes of the fundamental and the first two partial frequencies) to the inputs of a Max colour-chooser object. This object takes as its input parameters red, green, and blue colour component values. The amplitude of the fundamental frequency is mapped to the red colour component, while the second and third partials are mapped to the green and blue components respectively. This mapping results in the generation of an RGB colour value whose hue is dependent upon the weighting of tone amplitude amongst the partial frequencies found in the singer's vocal output.

Assigning colour in this way yields predictable and repeatable results in response to vocalization. Different vowel sounds produce different colours, as the differences in the formant structures of each vowel sound produce different amplitude weightings at the partial frequencies.

If a soprano sings the closed vowels /i:/ as in 'free' or /u:/ as in 'fool' on a mid-range note (such as A4 which has a frequency of 440 Hz), the vowel's formant structure is such that it closely resembles with the fundamental frequency of the singer's pitch. In these cases, tone amplitude is high at the fundamental and the resulting colour output is in the red-to-yellow range. If she sings the open vowel /a:/ as in 'car', the vowel's formant structure sits higher than the phonated pitch, resulting in increased amplitude at the second and third partials in the harmonic spectrum. Our system's colour output is then in the green-blue range.

Vocalization of any vowel at a high pitch (soprano high C and above) allows the singer to produce intense red/orange colours. This occurs because extremely high pitches are phonated at frequencies higher than the vowel characterizing formant frequencies. Research in vocal production conducted by Sundberg [6] indicates that at high pitches, soprano singers modify the vowel's formant frequency so that it rises to approximately equal the value of the fundamental frequency. This results in a high weighting of tone amplitude at the fundamental, causing our system to display a red/orange colour.

Using this mapping strategy, extreme sounds (the high pitched notes or closed vowels) can be characterized with the vibrant reds and oranges, and less dramatic sounds (the open /a:/ vowel sung at moderate pitches) with the more restful blues and greens. This colour dynamic is used as a way to visualize the intensity of a singer's vocalization.

The piece is composed with this mapping between timbre and colour in mind. The early sections are sung on open vowels in the singer's mid-range, producing blue-green visual imagery, while the more dramatic later passages require the soprano to exercise the extremes of her vocal range (from high C to a high F) in order to illustrate intensity through vibrant colour.

4 Visualizing Chord Relationships Through Colour

The Circle of Fifths (see Figure 2) is a music-theoretical device that geometrically represents the way the twelve major and twelve minor key signatures relate to one another [2]. Each key signature is rooted by one of the twelve notes of the chromatic scale. If the Circle is traversed in a clockwise manner, each subsequent step to the right on the Circle represents an interval of a Perfect Fifth.

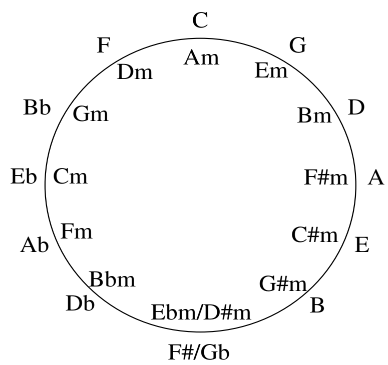


Fig. 2. The Circle of Fifths

The placement of chords on the Circle of Fifths indicates the similarity between the chords. C major and A minor appear in the same position on the Circle since A minor is the relative minor of C major. Both C major and A minor are key signatures that contain no sharps or flats. E minor is the relative minor of G major (both E minor and G major contain one sharp) and so on.

In this application the key signature relationships contained in the Circle of Fifths are visualized by mapping the Circle to the colour wheel. Chords that are adjacent on the Circle will have similar colour values.

The parallel between the circular structure of the Circle of Fifths and the colour wheel has been exploited in previous music visualization works. Jack Ox used a similar mapping in her *Color Organ* visualization [4]. Ox's work used the colour-associated Circle of Fifths to visualize the existing chord relationships in musical pieces. To represent harmonic structure in music, Ox mapped a visual colour wheel to the Circle of Fifths. Each chord present in the musical input triggered a colour representing the chord's position mapped on the colour wheel.

The mapping between the colour wheel and the Circle of Fifths was used for a subtly different purpose during the compositional process of creating *Deep Surrender*. Since the mapping between chords and colours was fixed, the colours and chords could be treated synonymously. The music was composed visually by selecting colours on the wheel that the composer felt would help illustrate the mood of each section of the piece. The sound of the song was determined by choosing between the relative major and minor chords found at the selected colour locations.

5 The Video Processing Effects

The unprocessed video footage used in this production contains images of white jellyfish on plain blue backgrounds. We use chroma-keying and colour balance modification to adjust the appearance of these video clips in response to the musician's performance.

Chroma-keying (commonly known as blue-screening or green-screening) refers to the process of isolating objects from one image source and inserting them into a different image source, making them appear to be part of a new scene. Jitter provides chroma-keying functionality through the `jit.chromakey` object.

Colour balance modification causes the entire hue of a video stream to be changed. Jitter allows the colour balance of video streams to be modified through the `jit.scalebias` object.

In the *Deep Surrender* project, we use the colours produced by the chords played on the keyboard to affect the overall colour balance of the video playback. The singer's vocalizations are used to affect the colour balance of the imagery that is chroma-keyed into the video.

Chroma-keying and colour balance modification are used in tandem to create the visual effect that certain jellyfish within the environment are changing colour while the environment remains unchanged. This is done by compositing colour-adjusted video streams in a multi-step process:

- The video stream containing the original jellyfish image is duplicated into a second Jitter matrix.
- The colour balance of the duplicate stream is altered to the colour determined by the singer's vocalization.
- The duplicated (colour-altered) jellyfish image is chroma-keyed into the original stream, overlaying the original jellyfish.

The resulting video stream then contains a colour-manipulated jellyfish while the rest of the image remains unchanged.

6 The *Deep Surrender* Performance

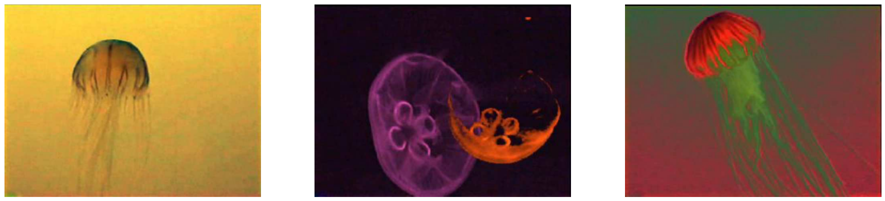
The intention of the piece is to illustrate how an artist can harness anxiety and adrenalin to produce a beautiful performance. This is achieved by utilising the visual metaphor of a jellyfish – a creature both beautiful and terrifying. The artist's musical performance manipulates the jellyfish representation in order to convey how the artist interacts with and overcomes her anxiety.

The *Deep Surrender* performance uses the video effects described in the previous section of this paper to respond to a live performer's vocalization and keyboard playing by manipulating aspects of the video footage.

In each section of the performance, different videos are played and different video processing strategies are controlled by the performer's musical input. Her interaction with the visualization is intended to illustrate her emotional state.

6.1 Part One

In Part One, the performer is merely an observer of the fearful environment. The visualization is simple, with the artist controlling only the colour of the image by playing simple chord progressions.



(a) Part 1 (b) Part 2 (c) Part 3

Fig. 3. Images from the Deep Surrender Performance

- The bold colours in this section are controlled through the `jit.scalebias` object which is triggered by keyboard input.

In Figure 3a, we see the results of the performer playing a G major chord on the keyboard. The G major chord corresponds to the orange-yellow colour tone according to our mapping between the Circle of Fifths and the colour wheel.

6.2 Part Two

In Part Two, the performer begins to experiment. She uses her voice to introduce entities to the visual environment and realizes that her actions elicit a colourful response from the ominous creatures.

- The composition modulates to an eerie set of chords which correspond to icy colour tones in the colour spectrum (purples, pinks, blues.)
- A dark and sinister video sequence begins.
- The colour balance is still modified by keyboard input.
- The performer's voice brings new and vibrantly coloured entities into the environment, using the `jit.chromakey` functionality.
- The performer's voice is used to offset the composed colour choices, making this section visually dynamic.

In Figure 3b, we see the orange jellyfish introduced to the scene as a result of the performer's vocalization. The jellyfish is orange because she is using a focused tone on a closed vowel (/i:/) which corresponds to an orange colour.

6.3 Part Three

In Part Three, the performer realizes that the most beautiful results are produced when she combines her voice with the fearful imagery. The visual effects flourish through her willingness to overcome her hesitation and thrive upon the adrenalin rush of fear.

- The song returns to the original theme, and the first video is replayed.
- Additional sound layers join the synthesized accompaniment, helping to convey new energy in this section.
- Using her voice as input, the performer can modify the video playback (through chroma-keying applied to duplicate image streams) to produce the piece's most vivid colour effects.

In Figure 3c, the performer sings extended high pitches (soprano high C and above) in order to superimpose the vivid red tones upon the scene. The underlying colour balance is controlled by the keyboard (the green colour is produced by the A chord) and the extreme vocalizations affect the highlighted portions of the jellyfish, yielding an intense visual result.

7 Discussion

A traditional musical performance conveys emotive cues to the audience through the musical structure of the composition and through the performer's nuances and body language. In addition to these traditional methods of communication, the multimedia performance environment we have created allows visual information to be presented to the audience. The content of the video streams and the visual responsivity, controlled by the performer, augment the performance and assist in communicating the concept of the musical piece.

The interaction techniques used in this application allow the performer to guide the visualization in an intuitive fashion. She need only play the piano and sing in order to manipulate the visualization, rather than function as a computer operator during the performance.

During the performance, the singer observes the visualization and adjusts her vocalizations (by modifing her vowel formation or pitch) in order to achieve the most pleasing visual results. Since the visualization of vocal timbre responds to vocal subtlety, the slight variations in vocal production that occur each time the singer performs the piece make each repetition of the visualization unique.

The *Deep Surrender* production has been performed several times in concert settings. Additionally, it is often performed during tour sessions of the Advanced Man-Machine Interface Laboratory in order to show visitors how visualization technologies can be used for artistic purposes.

8 Acknowledgments

The jellyfish video footage used in the *Deep Surrender* production was filmed by Melanie Gall.

References

1. Cycling '74. Max/MSP and Jitter.

2. Krumhansl, C.L. The Geometry of Musical Structure: A Brief Introduction and History. Computers in Entertainment, Volume 3, Issue 4, pages 1-14. ACM Press, 2005.
3. Oliver, W., Yu, J. and Metois, E. The Singing Tree: Design of an Interactive Musical Interface. In DIS'97: Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques, pages 261-264. ACM Press, 1997.
4. Ox, J. Two Performances in the 21st Century Virtual Color Organ. In Proceedings of the Fourth Conference on Creativity and Cognition, pages 20-24. ACM Press, 2002.
5. Puckette, M., Apel, T. and Zicarelli, D. Real-Time Audio Analysis Tools for Pd and MSP. In Proceedings of the International Computer Music Conference, pages 109-112. International Computer Music Association, 1998.
6. Sundberg, J. The Science of the Singing Voice. Northern Illinois University Press, 1987.
7. Taylor, R., Boulanger, P. and Torres, D. Visualizing Emotion in Musical Performance Using a Virtual Character. In Proceedings of the Fifth International Symposium on Smart Graphics, pages 13-24. Springer LNCS, 2005.
8. Taylor, R., Torres, D. and Boulanger, P. Using Music to Interact with a Virtual Character. In Proceedings of New Interfaces for Musical Expression, pages 220-223, 2005.