AN INTERACTIVE, MULTIMODAL INSTALLATION USING REAL-TIME SPECTRAL ANALYSIS

Benedict Eris Carey
Hochschule für Musik und Theater, Hamburg

John R. Taylor
Sydney Conservatorium of Music

Damian Barbeler
Sydney Conservatorium of Music

ABSTRACT

This paper describes an interactive sound and light system based on real-time analysis of an augmented musical instrument called the 'motor bow', that when played, makes new, unconventional, often 'noisy' tones. The interactive system comprises a timbre-matching engine that performs real-time analysis of the motor bow performance and algorithmically governed multi-channel diffusion of harmonically similar, but timbrally contrasting material coordinated with visual representations. This interactive installation was designed as part of the Tasmanian International Arts Festival 2015. Of primary importance was that the installation be an extension of the motor bow itself and that it work smoothly in extended performance scenarios. Particular emphasis was placed on extending emergent properties of this interference based relationship between bowed string performance and a weighted, spinning motor placed at the tip of the bow. The entire system can be viewed as an extension of an augmented instrument, one that is multimodal in nature, involving visual (a custom lighting-matrix), audio (acoustic and computer music performance) and tactile (motor bow) elements.

1. INTRODUCTION

1.1. The Acoustic Life of Sheds

The installation was commissioned by the Australian arts and social change company *Big hART* to create an installation in a disused shed in country Tasmania. "Bruce's shed" was due to be demolished, so the authors wanted to make a new performance space to pay tribute to the shed's history. This installation was part of a larger exhibition called "The Acoustic Life of Sheds".

Inspiration for all elements of the installation came rom the character for the site, a disused shed in the far North West of Tasmania. The shed had housed an amateur collection of historical curios in the 60s, 70s, and 80s. The owner was a collector of diverse objects ranging from WWI paraphernalia, to milking equipment, and seashells to indigenous artifacts. The shed was made of wooden paneling that over time had cracked and warped. Light in the shed fell through cracks and pitted windows to a cobblestone floor. All elements of the installation in terms of sound, performance, visuals and the behind the scenes software were influenced by the aesthetics of this building.

The shed, while decaying, had an atmosphere which, far from being dank and sorry, had a kind of quiet, child-like joy: an echo of the enthusiasm of the original man who created the place. Live strings were imagined performing constant dusty energetic patterning sounds of various kinds. Such sounds can be difficult to achieve on a string instrument due to the need to lift the bow regularly to create more bouncing energy. The Motor Bow was invented to solve this problem.

1.2. The Motor Bow

In this installation, the input sources were a conventional viola and cello, played in an improvised style using a modified motor bow (Figure 1) in combination with a wireless microphone. This bow augmentation system was designed and developed by composer Damian Barbeler. The motor bow is a standard string bow with a small electric motor attached at the tip. Electric wires wind down the bow stick to a battery secured under the hand of the player. A moderately weighted front door key is attached to the motor shaft, which when rotating, the uneven weight causes vibration (e.g. overhung rotor unbalance). Since a string bow can be considered a type 1 lever with a dynamic fulcrum (at the string), the radial and axial vibrations coupled with the changes in mechanical advantage as the player draws the bow up and down, produces four distinct bow behaviours, with each behaviour producing a distinct sound. These vibrational behaviours and resulting sounds correspond to four phases of the bow tip draw, relative to the string.

The phases, corresponding vibrational behaviour, and sound are: (1) nearest the hand: jittering, noisy; (2) middle of the bow: tremolo, pitched; (3) top quarter of the bow: small regular ricochets; and staccato

(4) top of the bow: large irregular ricochets, pointillistic, as catalogued by Damian Barbeler. This categorisation method was used in the recording process when building the sample database, so as to provide a diverse range of compositional material for our sample playback system. What was noteworthy about these four phases is how they informed the development of the analysis system. It had to be responsive to both 'noisy' and 'pitched' sound without creating an abundance of random output to ensure accurate, predictable triggering for lighting and sound generate.

Figure 1. The 'motor bow'.

In order to facilitate efficient communication within our team it had to be discussed using terminology familiar to classically trained musicians with varying degrees of familiarity with interaction design terminology.

1.3. System Concept Design

During system design, we used the term "Contrapuntal Interaction" in our own specific way, as short hand for giving the illusion of intelligent reactions from the system. It was felt that the software should "listen" to the player, then respond with sounds which were neither a predictable echo, nor too abstract or random. Rather, a "goldilocks zone" response, reminiscent of the live sound, with some new additional timbral elements. With the motor bow providing a novel set of timbral characteristics, each linked to the physical areas of the bow, groupings of periodicity/aperiodicity, flux, and noise emerged, as the predominant spectromorphological components of the sound (Smalley, 1986; 1997). As the bow enters each of the four physical phases, significant independent variations in amplitude and noise, provided an opportunity to elicit diverse system behaviour over the course of each stroke.

Our solution was a system that identified the spectral flatness of the player's input, and matched the spectral flatness to an opposing spectral flatness value: when the player played a noisy sound, the system would play something with more perceived pitch. In addition, tonality and amplitude information was also extracted from the player's input to refine the system responses, so that they appeared to be the product of a creative mind developing the sound further. It was decided that from an artistic and interaction design perspective, a multimodal system that reacted in a contrapuntal manner would produce a clear relationship between the performer and the audio and lighting system, thus inferring dynamic behaviour. These design choices presented unique challenges for real-time spectral analysis and matching of timbres, while simultaneously maintaining relevant dynamic system behaviour and artistic vision.

2. IMPLEMENTATION

2.1. Timbre-Matching Engine

Designing a real-time 'timbre-matching' engine can be seen as a relatively impressionistic exercise, where the relationship between the input and output streams can be modified depending on the preference of the designer. Our timbre-matching engine contrapuntally triggers spectrally similar material, and operates as a flexible audio-based controller for a custom DMX lighting system. Therefore, from the outset of the prototyping phase, our sample set informed the aesthetic consequences of the system's behavior. Additionally, early design requirements indicated that such a system should be capable of being applied to a variety of audio inputs, in as yet unknown scenarios, so as to be useful for future installation and performance projects without entirely relinquishing control over the selection of the playback material to aleatoric factors.

In collaborating with musicians from a diverse range of specialist areas and educational backgrounds, the common language provided by knowledge of modern musical notation and avant-garde performance techniques by the group became the standard communication convention, and ultimately the framework on which the analysis engine was built. Monophonic pitch tracking was seen as a less useful option in our case where we were interested in triggering material from within our sample database that was differentiated from the live input audio in terms of timbre (noise/tone relationship), yet remained harmonically similar. The presence of multiphonic and chordal viola and cello material was also expected at the audio input and these two considerations impacted on the analysis methods chosen.

It remains a highly processor-intensive task to track timbral changes of input audio in real-time, prohibitively so for use in a system also dealing with multiple modes of output due to the complex calculations involved. In our case, LED mapping was to prove highly demanding in terms of system resources, as was spectral analysis within the Max/MSP environment on the available development hardware (MacBook Pro, Mid 2012, 2.3 GHz Intel Core i7, 8 GB 1600 MHz DDR3, Intel HD Graphics 4000 1024 MB, OS X 10.1 Yosemite). The aim of this analysis engine was therefore to select a series of samples for playback from within a

pre-analysed collection, following on from similar efforts by Jehan and Schoner (2001) and Carpentier and Bresson (2010) using a simple analysis model. This was approximated through commonalities of pitch material (extrapolated from FFT data) and a tonality coefficient, which can be measured by calculating the spectral flatness of an input signal (Dubnov, 2006). An overview of the system design is shown in Figure 2.

The analysis component of this system is implemented through manipulation of data derived from the real-time tracking of spectral flatness data, amplitude, and pitch information. It relies on three objects from the Max/MSP environment in particular; *zsa.flatness* and *zsa.freqpeak* (Malt and Jourdan, 2008) and part of the *Max4Live* package *live.slider* (for amplitude tracking) in combination with dynamically updated histograms and databases of 'pre-performance analysis'[1].

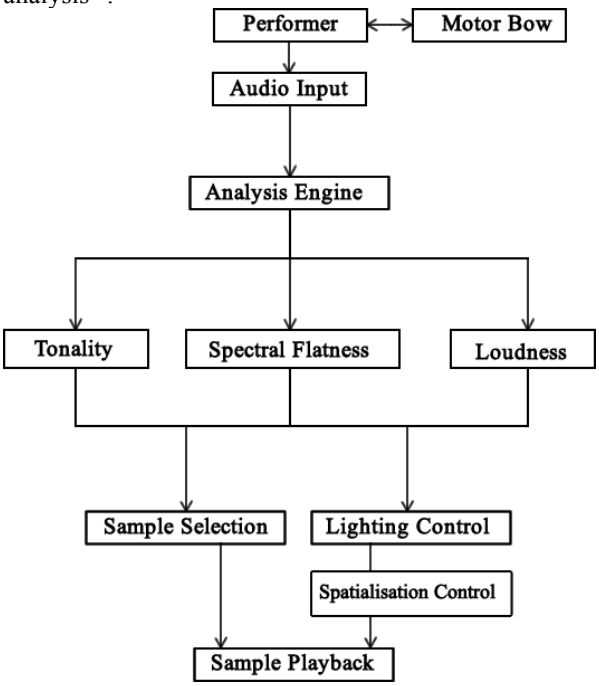

Figure 2. An overview of the analysis system.

Our pre-collected analysis data is based on two databases of motor bow performed viola audio samples recorded at 44.1kHz. The first sample database, comprising of three hundred samples was previously classified into eight groups of increasing spectral flatness, and is sorted and filtered based on the 'Tonality', 'Spectral Flatness' and 'Loudness' values deduced over an analysis period unique to each sample's length. This database was accessed and utilised through the *mxj* Java-framework provided with Max/MSP, again to reduce the overall demand on system resources.

When using the analysis system in a live context, initially, a spectral flatness factor is tracked in real-time over a variable analysis period (referred to as the analysis period throughout the following pages), typically between approximately 1 and 6 seconds, determined by the length of the previously triggered sample. This process determines the most commonly detected 'noise-like to tone-like' ratio of an audio input chunk. High spectral flatness is assumed to correlate to a generally 'less-noisy' signal, and low spectral flatness to a 'more-pitched' (more periodic) signal (Dubnov 2006). This number is stored per window, represented as an integer between 1 and 100, to a dynamically updated collection, which is continuously ranked based on the number of occurrences of each unique spectral flatness value from the most often to least often detected value. This is tracked using the *histo* object in Max/MSP (Figure 3). The amplitude threshold can be modified through the GUI during performance, to filter out lower amplitude frequencies, if desired.

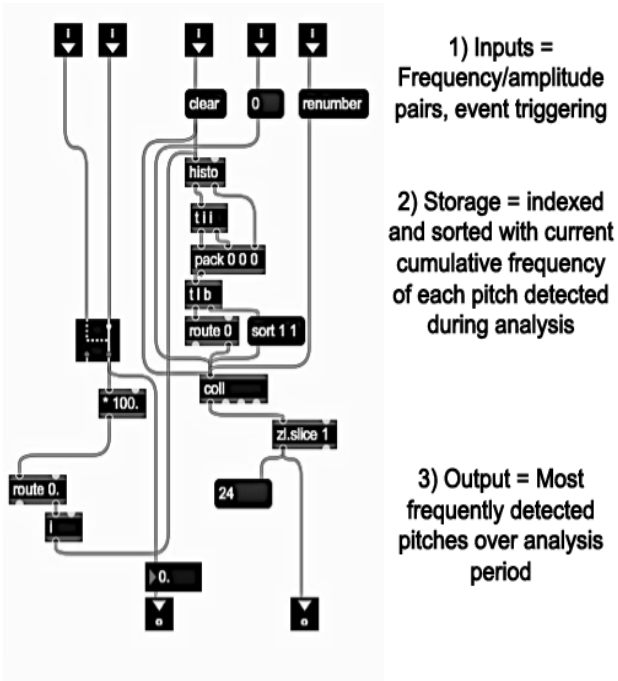

Figure 3. Histograms are used to track the most frequently detected spectral features i.e. pitch/amplitude pairs and current spectral flatness from the same input audio for comparison with the database items in a similar fashion to methods used by Tzanetakis et al. (2003).

Secondly, the most frequently detected spectral flatness value over the analysis period is temporarily stored alongside the relevant string of most significant pitch information, as calculated by the Tonality component of the analysis system. This component outputs a string of 24 floating-point values representing the loudest and most often detected frequencies over the same analysis period. This was implemented through similar use of a histogram function and array as with the Spectral Flatness component. A maximum of 24 unique values was chosen for this string due to the fact that each partial was rounded to the nearest quarter-tone based on a system similar to the one used by Tristan Murail in composing "Gondwana" (Tzanetakis and Cook, 2002). This is intended to allow the performer to interact with the system on a more familiar, music notation related

[1] See Tzanetakis and Cook (2002) and Tzanetakis et al. (2003) for a more detailed explanation of this technique.

level, it is also one way to generalise pitch material for timbre matching.[1] The pitch resolution of the system is therefore 24 discrete divisions per octave or 24-TET (see Figure 4).

The Tonality component differs in that it is based on pairs of floating point values representing pitch and amplitude as tracked by the *zsa.freqpeak* object. This data is derived using a continuously calculated FFT with a window size of 512 samples at a sample rate of 44.1kHz, which also returns 24 pitch and amplitude pairs, per window. Finally, the average amplitude recorded per frequency band is retained in order to further rank the final 24 values, output by the Tonality section of our system from 'loudest' to 'softest'.

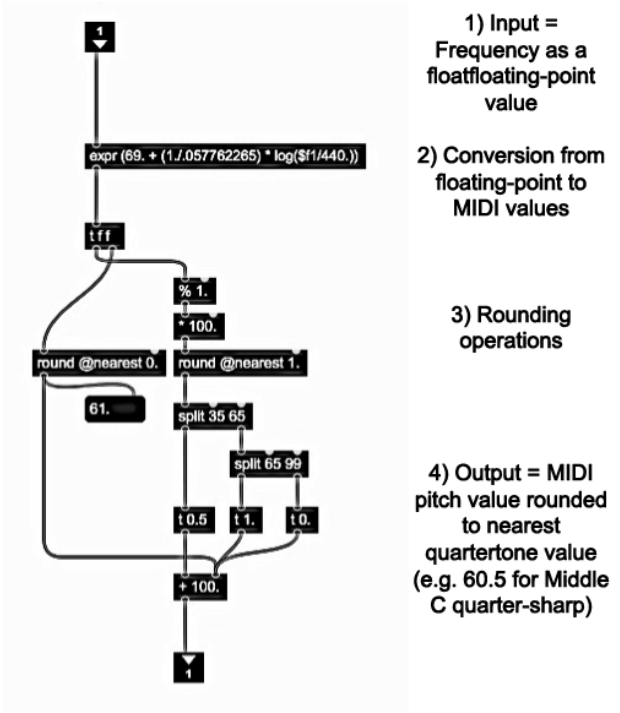

Figure 4. A pitch-quantisation system based on techniques used by Tristan Murail in his preparation of Gondwana. Frequencies are rounded to the nearest quarter-tone (Rose, 1996).

Once the aggregated tonality data and the spectral-flatness coefficient have been determined for the previous analysis period, they are concatenated into a string and stored to an array for retrieval by the Sample Selection component. This component compares the input data with the pre-collected analysis data, and outputs the index of the most similar candidate from the pre-analysed sample set. Internally, the tonality data of each member of the relevant spectral flatness collection is compared to the tonality data derived from the input audio, but only for the most relevant spectral flatness bandwidth group, to determine which sample contains the highest number of coincidence pitches. This data is then passed on to the sample playback engine, which builds musical phrases based on the collected data and synchronously spatialises the audio and lighting across the installation.

2.2. Audio Spatialisation

Spatialisation is achieved using eight individual speakers connected to an external soundcard with eight outputs, placed in the corners and equidistantly of the lighting matrix. The final output of the analysis system is fed into a sample playback probability table and the sample playback engine. In order to maintain spatial coherence between the light and audio, each channel's signal is crossfaded from interpolated coordinates of lighting movement. Since the visualization render was three dimensional, it was possible to change the spatialisation across three axes: X and Y where the audio moves with the spatialisation; and Z, where the spatialisation adds depth of field, and therefore a perceived sense of height, to the installation.

Like the primary sample database, the secondary sample database, comprising of the three hundred motor bow samples, was categorized by spectral flatness values. However, instead of a further categorization by pitch, these samples were categorised by amplitude, and are triggered by the amplitude of the performance input. Spatially, the playback of the secondary samples follows the primary sample set, although owing to the contrapuntal nature of the system, the secondary sample's main function is to add intensity and density in contrast to softer performance input.

In both instances of the playback engine, sample selection is achieved using phrases that comprise of three samples from the collection played with a predefined but adjustable gap of silence between them. The system can be activated using a wireless controller, with three modes of operation: automatic activation and control from the performer, manual activation and control, and interrupt, where certain parameters can be adjusted "on the fly" with a wireless touchscreen controller.

2.3. The Lighting System

The lighting system (Figure 5) is comprised of 1000 addressable LEDs in 10 rows of 100 LEDs, designed to be hung from the ceiling in nine-meter strips, obscured by white balloons of different sizes, in order to diffuse the light. The dimensions of the pixel matrix were determined by the dimensions of the shed. The interactive lighting was created using Max/MSP/Jitter, and then rendered in Jitter Open GL. Syphon Server (Butterworth and Marini 2015) was then used to render the Open GL window inside MadMapper, which was then subsequently transmitted to a PixLite pixel

[1] Murail's method rounds to the nearest ¼ or ½ tone, as opposed to rounding up or down consistently as is the case with the *round* object available in the Max/MSP object suite. Our implementation also has a bias towards 12-tone tempered pitches by using unequal zones where the ¼ tone zone is 30 cents in size while the tempered zone is 70 cents. Due to this generalisation, the system is more reactive. Max 5 and Max 6.1 were used in development and the performance executed in Max 6.1, so issues of compatibility were paramount.

controller connected to the LEDs via the ArtNet protocol.

Figure 5. Lighting System Prototype being tested by Composer Damien Barbeler and Violist, Nicole Forsyth.

2.4. Lighting Interaction Design

The lighting interaction was created in Jitter using the *Boids3d* Max object (Singer et al. 2015) to control Jitter objects that simulate three-dimensional animal flocking in a distributed behavioural model (Reynolds, 1987). Aesthetically, this system sought to impart a sense of depth in the space by having flocks move forward and backward along the length of 100 LEDs. In order to create a consistent visual flocking motion, the attraction point of the Boids was moved back and forth at periodic intervals, although the attraction point of the Boids are constrained to the end of the matrix in the event that the incoming performance audio is both high in amplitude, and high in spectral flatness. In addition, when instructed to play *fff,* the performer triggered colour inversion of the lighting matrix, where the background colour matrices became white, and the Boid black. The Boid behaviour returned to normal once both the amplitude and spectral flatness reduced below their respective thresholds.

Additional visual behaviour includes dynamic changes in the contrast between the "background" LEDs and the Boids, where inactive LEDs have an increase in brightness. This behaviour reduces the brightness of the Boids, reducing their visual impact. In addition, the size and speed of the Boids are also affected by the spectral flatness of the performance input, with higher spectral flatness increasing the size and speed of the Boids.

2.5. Player interaction with the System

The player reacted to this "contrapuntal" system by changing their approach to their performance. This manifested itself in the form of their approach to the inclusion of more wide ranging sounds in terms of the "pure-to-dirty" or pitched vs. noisy spectrum. Although the player was aware that the system reacted to the spectral flatness of their performance, the correlation and translation of such a spectral feature to an identifiable performance metric, led to an exploratory approach to interaction, as the performer sought to illicit more extreme responses by themselves exploring extremes of their own performance norms.

This resulted in more dynamic, contrasting and evocative improvisations by the player. Additionally, the motor bow limited the player's ability to improvise melody. When using the motor bow, the player tended to stray away from melody, instead opting to explore the motor bow equivalent of *klangfarbenmelodie*. The influence of the software in producing more extreme spectral flatness, combined with the player's colouristic, gestural approach, created highly evocative improvised responses. The interaction between the Motor Bow and the system, particularly the system behaviours can conceptually, be considered a proxy score, which in the end produced a repeatable and idiosyncratic work with very specific recognisable features, without the need for detailed notated music.

The sonic result of the final work closely resembles "call and response", where the system generates a contrapuntal response that causes tension as short bursts of noise from the player are contradicted by system generated longer fluttery sounds with more perceived pitch. When the player created more sustained, pitched sounds, the system played shorter noisier sounds.

2.6. Interaction with a Standard Bow

In order to determine the limits of the timbre-matching engine, the performance input was changed to a standard bow. Producing more stable tones with less noise, and with greater relative partials gave rise to interesting and unexpected behaviours. With the motor bow, while there was an element of "call and response" in the system, the contrapuntal nature of the spectral features produced a greater sense of tension in the way the system generated audio material. With the standard bow, the contrapuntal relationship was less evident. This was owing to the selection of samples that were not previously accessible by the system under the timbral constraint of the motor bow. Consequently, the system displayed imitative

behavioural responses to the performer. However, owing to the contrapuntal implementation of the visuals (e.g. the size and speed of the Boids) the visual element lacked the kind of variety seen with the motor bow.

3. CONCLUSIONS

The interactive audiovisual installation described in this paper displays very complex, but easily recognizable associative behaviours. The timbre-matching method was chosen owing to the relatively noisy four-phase characteristic of the motor bow. Since the motor bow is a novel extension of the traditional bow, no previous real-time spectral analysis has been undertaken on a bow of this nature, in the context of an interactive installation. The four-phases of the motor bow required constant analysis, with which larger FFT window sizes compromised system resources, since the timbre-matching engine, the DMX light system, and the sound spatialisation were running from within the same software. This window size proved adequate for the desired aesthetic consequences of the system. Additional optimisations included reducing and sorting the returned spectral and pitch values from the analysis, targeting the relevant samples through the comparison engine.

Early prototypes found latency between the Boids and the spatialisation, owing predominantly to the reporting rate of the Boids coordinates, and the pixel rendering. This was because the Jitter window, being at different dimensions (320 x 240) to the pixel matrix (100 x 10), was causing the Boids to stretch and thus producing a sense of latency.

The exhibition "Ten Days on the Island" for the Tasmania International Arts Festival has led the authors to create a complex interactive system that performs real-time analysis on a new type of augmented musical instrument. Owing to the uniqueness of the timbres created by this instrument, a real-time timbre-matching engine was created to make use of the new timbres that proved to be less processor intensive than previous efforts made by the authors with similar software tools.

Further work could include extending the timbre-matching matching engine to other instruments, such as vibraphones, idiophones or membranophones and, as previously described, with a standard viola bow. By focusing on the analysis engine, it may be possible to use a greater percentage of the computational overhead, thus improving the resolution of the FFT windowing size and the overall resolution. In addition, there may be wider applications of real-time analysis, by searching for other spectral features, for example, spectral centroid or bark scale, and could be used in conjunction with other instruments, in addition to more complex feature extraction techniques such as that proposed by Khadevich and Omologo (2013) to potentially cover larger sample databases. A transportable version of this installation is currently being developed, with the lighting matrix housed in an inflatable structure, for demonstration in temporary exhibits.

3.1. Acknowledgments

The authors would like to thank Big hART, The Bundanon Trust, Tasmania International Arts Festival, Synergy, Georg Hajdu for his comments on the pitch quantisation system and Bruce's family for the loan of his shed.

4. REFERENCES

Butterworth, T., and Marini, A. 2015. "Syphon Framework." Available online at syphon.v002.info. Accessed 19 July 2015.

Carpentier, G., and J. Bresson. 2010 "Interacting with symbol, sound, and feature spaces in orchidée, a computer-aided orchestration environment." *Computer Music Journal* 34(1):10-27.

Dubnov, S. 2006. "Spectral Anticipations." *Computer Music Journal* 30(2):63-83.

Jehan, T., and B. Schoner. 2001. "An Audio-Driven Perceptually Meaningful Timbre Synthesizer." In *Proceedings of the International Computer Music* Conference*,* pp. 381-388.

Khadkevich, M., and M. Omologo. 2013. "Large-Scale Cover Song Identification Using Chord Profiles." In *Proceedings for the International Society for Music Information Retrieva*l, pp. 233-238.

Malt, M., and E. Jourdan. 2008. "Zsa. Descriptors: a library for real-time descriptors analysis." In *Proceedings of Sound and Music Computing (SMC)*, pp.134-137.

Reynolds, C. 1987. "Flocks, herds and schools: A distributed behavioral model." In *Proceedings of ACM SIGGRAPH Computer Graphics,* vol. 21, pp. 25-34.

Rose, F. 1996. "Introduction to the pitch organization of French spectral music." *Perspectives of New Music* 34(2):6-39.

Singer, E., et al. 2015. "Boids for Max." Available online at s373.net/code/. Accessed September 2015.

Smalley, D. 1986. "Spectro-morphology and structuring processes," In S. Emmerson, ed. *The language of electroacoustic music.* Basingstoke: McMillan, pp. 61-93.

Smalley, D. 1997. "Spectromorphology: explaining sound-shapes." *Organised Sound* 2(2):107-126.

Tzanetakis, G., and P. Cook. 2002. "Musical genre classification of audio signals," In *Proceedings of IEEE transactions* on *Speech and Audio Processing,* vol. 10, pp. 293-302.

Tzanetakis, G., Ermolinskyi, A., and P. Cook. 2003. "Pitch histograms in audio and symbolic music information retrieval." *Journal of New Music Research* 32(2) pp. 143-152.

Wechsung, I. 2014. "What Are Multimodal Systems? Why Do They Need Evaluation? - Theoretical Background." In I. Wechsung, ed. *An Evaluation Framework for Multimodal Interaction*, Switzerland: Springer, pp. 7-22.