http://dx.doi.org/10.14236/ewic/EVA2016.11

A Perceptually Motivated Visualisation Paradigm for Musical Timbre

Sean Soraghan
ROLI / Center for Digital Entertainment
2 Glebe Road, London E8 4BD
United Kingdom
sean@roli.com

Alain Renaud
Center for Digital Entertainment
Bournemouth University, BH12 5BB
United Kingdom
arenaud@bournemouth.com

Ben Supper
ROLI
2 Glebe Road, London E8 4BD
United Kingdom
ben@roli.com

Background is given on existing research into the perception and description of timbre. This is used to contextualise the development of visual mappings for acoustic timbre features. Two example systems are explained and discussed, which demonstrate the use of perceptually motivated mappings from acoustic features to visual properties, in specific contexts. Possible applications and extensions of such systems are proposed.

Timbre. Audio analysis. Timbre visualisation. Generative computer graphics.

1. INTRODUCTION

Timbre is a complex, multidimensional attribute of sound. It is typically defined as a collection of sound attributes by which two sounds can be distinguished while having identical pitch, volume and duration (American Standards Association 1960). This definition has become outdated for a number of reasons. Firstly, it assumes a separation between timbre, pitch and volume. There is existing research that suggests that the three can interfere in audio perception (Wapnick & Freeman 1980, Platt & Racine 1985, Beal 1985, Wolpert 1990, Melara & Marks 1990, Krumhansl & Iverson 1992, Warrier & Zatorre 2002, Marozeau & de Cheveigné 2007, Caruso & Balaban 2014). The definition also leaves no possibility for the quantification or measurement of timbre, since it is defined in terms of what it is not (Dannenberg 1993). Finally, the definition originated in the context of acoustic sound sources, where the main consideration of timbre was with respect to the *identification* of sound sources (Grey 1975, Wessel 1978, Wedin & Goude 1972, Berger 1964).

The introduction of computer music and digital synthesis techniques have made possible the creation and manipulation of widely varying timbres that are not inherently linked to any acoustic sound source, since they are digitally produced. Although a more sufficient definition of timbre has not been proposed, there is a growing body of research that suggests that it is better understood in relation to a large set of audio descriptors that capture different aspects of the temporal, spectral and spectro-temporal qualities of the sound.

For example, research into the perception of timbre using the technique of multidimensional scaling (MDS), in combination with regression analysis using different audio features, has identified various audio features that are important in the perception of timbre (Grey & Gordon 1978, Krimphoff *et al.* 1994, McAdams *et al.* 1995, Kendall, Carterette & Hajda 1999). Similarly, research into the semantic description of timbre has identified various visual and material descriptors that are commonly used in the description of timbre (von Bismark 1974, Pratt & Doak 1976, Seago, Holland & Mulholland 2004, Seago 2013).

The objective of this paper and the systems described within is to unify these existing avenues of research in order to produce novel audio-visual environments for timbre visualisation that are based on the perception of timbre. The development of effective mappings from acoustic to visual features could aid in the design of software interfaces and environments for timbre creation and manipulation.

2. BACKGROUND

An extensive list of audio descriptors that can potentially be used in the characterisation of timbre is provided in the "Timbre Toolbox" (Peeters *et al.* 2011). Peeters *et al.* conduct information redundancy analysis on the extensive list of features, resulting in a collection of clusters of descriptors. The audio features used to drive each of the visualisation systems described in this paper are taken from this list. They will be detailed in Section 3.1.

Analysis of the semantic description of timbre using the techniques of semantic differential and verbal attribute magnitude estimation (VAME) have shed light on the way in which people understand timbre without relation to direct sound sources (Seago 2013). Multiple studies have identified that listeners often use visual metaphors when describing timbre, with semantic descriptors being material, textural and physical in nature (Pratt & Doak 1976, Disley, Howard & Hunt 2006). Three salient clusters relating to: *volume/wealth*; *brightness and density*; and *texture and temperature* have been identified by Zacharakis, Pastiadis & Papadelis (2011). This has since been referred to as the "luminance texture mass" model in a follow-up study (Zacharakis, Pastiadis, Reiss & Papadelis 2012) and confirmatory studies have been conducted which show that the model is consistent across two different languages – English and Greek (Zacharakis, Pastiadis & Reiss 2014, Zacharakis & Pastiadis 2015).

An in-depth discussion on the complexities of audio-visual associations is given by Tsiros (2013). Tsiros talks of the links between modal experience (from embodied experience of the world) and amodal concepts (from abstract thought). It is argued that knowledge generated from modal experience could be used creatively in order to "inform the design of multimodal systems, for organisation, comprehension and interaction with modal and amodal information." The aim of the systems described in this paper is to explore this possibility. In the context of these systems, the 'amodal concepts' are the visual metaphors used in the description of timbre. The 'knowledge from modal experience' consists of the acoustic properties of timbre and their visual metaphorical mappings.

The systems were designed using existing research into the links between acoustic features and semantic descriptors of timbre. A very similar process is described by Tsiros and Liplatre (2015) in their introduction of the AniMorph system. The difference is that the AniMorph system extracts visual features in order to drive audio synthesis. This paper effectively describes the reverse – the extraction of audio (timbre) features to drive generative visualisation.

3. EXAMPLE SYSTEMS

This section discusses two example systems that have been developed using perceptually motivated mappings between acoustic attributes and visual attributes. Each system makes use of specific graphical techniques and therefore has its own context-dependent visual attributes / parameters. The choices of mappings between acoustic attributes and visual attributes is justified in each case with reference to the existing literature. Firstly, the acoustic parameters that are used in these systems will be discussed.

3.1 Acoustic Timbre Features

The acoustic features used to drive the different timbre visualisation systems are all taken from Peeters *et al.* (2011). They are divided into spectral parameters and harmonic parameters. The spectral features describe the structure and shape of the spectral energy distribution of the audio for very short windows of time. The harmonic features are calculated over longer windows of time and describe the harmonic content of the audio. Audio analysis is performed continuously on the audio, providing real-time tracking of the spectral and harmonic evolution of the audio. Short windows of audio are firstly transformed to the frequency domain, using a fast Fourier transform (FFT). The generated audio is windowed using Bartlett windowing, with 256 samples per window for spectral analysis and 2,048 samples per window for harmonic analysis, at a sample rate of 48,000 Hz. Therefore, the total frequency bandwidth covered is 24,000 Hz and this is split into 128 real-valued frequency bins for spectral analysis and 1024 real-valued frequency bins for harmonic analysis. Table 1 shows all of the acoustic features used, along with their mathematical definitions. Table 2 shows a mathematical notation key.

Table 1: *Acoustic features*

Acoustic Feature	Mathematical Definition
Spectral Centroid	$\mu_1(t_m) = \sum_{k=1}^{K} f_k \cdot p_k(t_m)$
Spectral Spread	$\mu_2(t_m) = \left(\sum_{k=1}^{K} (f_k - \mu_1(t_m))^2 \cdot p_k(t_m)\right)^{1/2}$
Spectral Flatness	$SFM(tm) = \frac{\left(\prod_{k=1}^{K} a_k(t_m)\right)^{1/K}}{\frac{1}{K}\sum_{k=1}^{K} a_k(t_m)}$
Harmonic Energy Ratio	$HER(tm) = \frac{1}{E(t_m)} \sum_{h=1}^{H} a_h^2(t_m)$
Inharmonicity	$inharm(t_m) = \frac{2}{f_0(t_m)} \frac{\sum_{h=1}^{H} (f_h(t_m) - hf_0(t_m)) a_h^2(t_m)}{\sum_{h=1}^{H} a_h^2(t_m)}$

***Table 2:** Mathematical notation key*

Notation	Definition
t_m	The audio signal within time window m
K	The number of real-valued frequency bins
f_k	The centre frequency of bin k
a_k	The magnitude of energy in frequency bin k
p_k	The normalised magnitude of energy in bin k
a_h	The magnitude of energy in the frequency bin that contains frequency h
f_0	An estimation of the fundamental frequency of the signal
$f_h(t_m)$	The magnitude of the partial with frequency closest to harmonic frequency h
E_{tm}	The total energy in the spectrum at time window m

The audio feature extraction is implemented in a standalone application, developed in C++ using the JUCE framework. The features are sent via OSC to the different visualisation applications.

3.1.2. Spectral Parameters

The **spectral centroid** measures the central tendency of the frequency spectrum. The **spectral spread** measures the spread of the frequency spectrum, around the mean. The **spectral flatness** is a measure of the flatness of the spectrum. Noisier signals tend to have flatter spectrums (spectral flatness value close to 1). The **spectral slope** is also used. The spectral slope is the result of linear regression on the spectral amplitude values. The spectral slope is linearly dependent on the spectral centroid value, so they measure very similar qualities of the spectrum.

3.1.3. Harmonic Parameters

The **harmonic energy ratio** is an estimate of the ratio of harmonic energy in the signal, as opposed to noise energy. The **inharmonicity** is an estimate of how much each partial deviates from pure harmonicity.

3.2 Generative Geometry

System 1 makes use of a geometric spherical mesh to visualise the timbre features. Deformations of the mesh and bump mapping in the rendering are used to produce different textural, shape and brightness qualities that visualise the various timbre features. The decision to use a 3D object for the visualisation in this system was influenced by the fact that one's notion of timbre is often concerned with the physical properties of the sound source (Tsiros 2013). The symmetric nature of a 3D spherical model makes it quite simple to deform in complex ways, providing various surface textures and forms, as will be discussed in Section 3.2.2. The system was developed using the Unity game engine. Two example videos of this system can be viewed under https://youtu.be/te_0plnNOE0 and https://youtu.be/BYC2jgnMpZM.

3.2.1. Luminance / Brightness

The sphere is rendered using Blinn-Phong shading (Blinn 1977). The brightness of both specular and diffuse components can be controlled individually. A third parameter is the specular radius (the radius of the specular highlight). Figure 1 shows the effect of varying the specular radius value. A small specular radius leads to local specular highlights, whereas a large specular radius leads to widespread illumination.

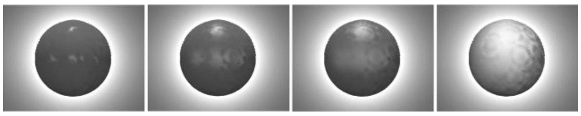

***Figure 1:** Varying specular radius value from low (left) to high.*

3.2.2. Texture

The spherical structure consists of a spherical model. A vertex extrusion shader is used to extend the vertices along their normal vectors according to two sinusoidal functions: *sin (az * faz) * aaz* and *sin (inc * finc) * ainc*, where *az* = the position of the vertex along the azimuth arc of the sphere and *inc* = the position of the vertex along the inclination arc of the sphere. faz and finc are multipliers that control the frequency of deformation along the azimuth and inclination arcs, respectively. Similarly, aaz and ainc control the amount of deformation. Figures 2 and 3 show the effect of varying aaz and ainc individually. Figure 4 shows the effect of varying both simultaneously.

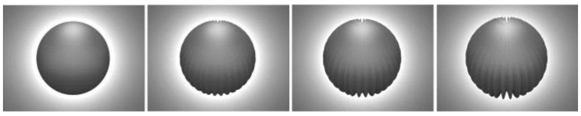

***Figure 2:** Varying azimuth deformation amplitude value from low (left) to high.*

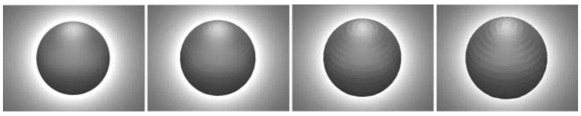

***Figure 3:** Varying inclination deformation amplitude value from low (left) to high.*

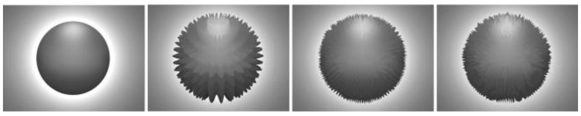

***Figure 4:** Varying inclination and azimuth deformation amplitude values from low (left) to high.*

In addition to the vertex extrusion deformation, bump mapping is implemented by perturbing the normal vectors slightly, using a Perlin noise function. This noise function has two parameters: the amount by which the normals are perturbed and the granularity of the Perlin noise. This bump mapping technique produces the impression of coarse surface deformations on the sphere. Figures 5 and 6 show the effect of changing the amount and granularity, respectively.

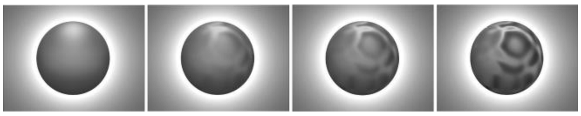

Figure 5: *Varying bump mapping amount value from low (left) to high.*

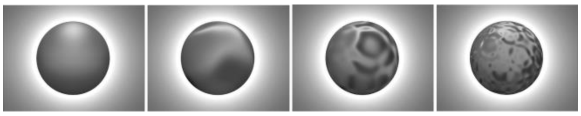

Figure 6: *Varying bump mapping granularity value from low (left) to high.*

3.2.3 Mappings

3.2.3.1 Luminance / Brightness

Multiple studies have found that the spectral centroid provides a decent indication of the brightness of a sound (Beauchamp 1982, Schubert, Wolfe & Tarnopolsky 2004, Schubert & Wolfe 2006). Therefore, the spectral centroid is used to drive the brightness in the visualisation. Additionally, the spectral spread is mapped to the specular radius in the visualisation. This way, more broadband audio will produce more evenly dispersed lighting, whereas narrow-band audio will produce particular highlights (Figure 1).

3.2.3.2 Texture

The spectral flatness measure is mapped to the frequency of inclination deformation in the vertex extrusion. The amount of inclination deformation is controlled by the spectral centroid. This means that noisier signals will produce rougher deformations on the sphere, and this deformation will be heightened when signals have a high centre of mass of spectral energy. Links between texture roughness and noisiness have been shown in Giannakis (2001) and Berthaut, Desainte-Cathrine, and Hachet (2010).

The frequency of the azimuth deformation is controlled by spectral centroid, and the azimuth amount is inversely controlled by the spectral spread. This means that the frequency of the azimuth deformation will be dependent on the centre of mass of the spectrum, and the amplitude of this deformation will be accentuated when the spectral energy is centred around the mean.

The harmonic energy ratio is inversely mapped to the bump mapping amount, meaning that audio with less harmonic content will produce more pronounced deformations. Harmonic energy distribution was linked to texture in Zacharakis, Pastiadis & Reiss (2014). The granularity of the bump mapping is controlled by inharmonicity, meaning that 'out of tune' audio will produce more granular bump deformations (Figure 6). Inharmonicity has been linked to texture-repetitiveness in Giannakis (2001).

3.2.3.3 Mass / Volume

The overall deformation amount in the vertex extrusion is controlled by the root-mean-square amplitude (RMS) of the audio, multiplied by the spectral centroid. Thus, louder audio will produce more pronounced deformations, with high centre of mass in the spectrum emphasising these deformations. A link between mass and spectral centroid was suggested in Zacharakis, Pastiadis & Reiss (2014).

3.3 Fluid Simulation

System 2 uses a fluid simulation algorithm for visualisation, which is driven by the timbre features. In the visualisation, fluid is emitted in bursts from the centre. The properties of the fluid are then controlled parametrically by the audio features in real time. Conversely to the generative geometry system, the main motivation behind the use of fluid simulation for visualisation was the fact that sounds are experienced as events in time that can mask one another and interact with one another (Tsiros 2013). A graphical fluid simulation has similar properties, in that events in time affect the system as a whole, and events can be seen to interact with one another, affecting the overall system in the process. An example video of this system can be viewed at https://youtu.be/LnoJ6eLfAK4.

The Cinder C++ framework is used for the fluid simulation. The 'cinderfx' cinder block (framework extension) is used, which provides a Fluid2D class. This class provides a number of fluid parameters which can be altered to provide various effects. These will be detailed in this section.

3.3.1. Luminance / Brightness

The colour and luminance of the fluid is implemented using a hue-saturation-value (HSV) algorithm. The hue, saturation and value can be altered in real-time. Figure 7 shows the colour of the fluid varying from low hue to high.

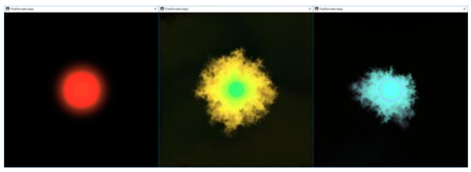

Figure 7: *Varying colour hue value from low (left) to high.*

3.3.2. Texture

The parameters of the fluid simulation that are altered in order to vary the texture are: RGB viscosity, velocity viscosity, RGB dissipation, velocity dissipation, and vorticity scale. Figures 8, 9, and 10 show the effect of varying these parameters.

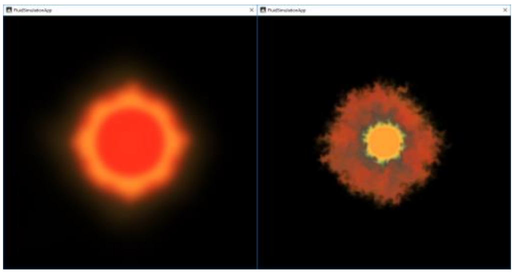

Figure 8: *Varying viscosity values from low (left) to high.*

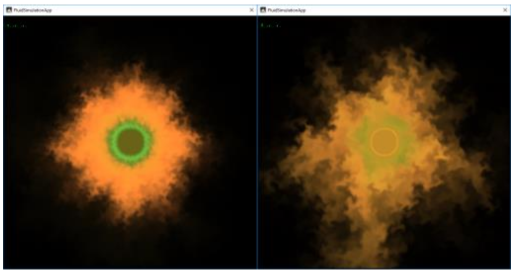

Figure 9: *Varying velocity dissipation value from low (left) to high.*

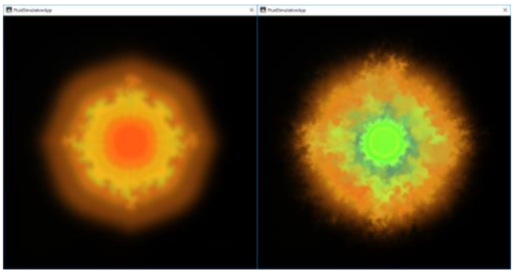

Figure 10: *Varying vorticity scale value from low (left) to high.*

The viscosity affects the momentum of the fluid. A lower viscosity value leads to 'faster', less resistive fluid movement. The dissipation value affects the rate at which the fluid dissipates throughout the environment. Both the dissipation of the velocity and the colour can be controlled individually. The vorticity scale affects the granularity of the fluid. Lower vorticity scale produces a finer-grain texture in the fluid.

3.3.3. Mass / Volume

The fluid is emitted from a central emitter. The bursts are emitted with a certain velocity (force) which affects how far the fluid initially travels. This can be controlled parametrically.

3.3.4 Mappings

3.3.4.1 Luminance / Brightness

In the construction of the HSV colour for the continuous bursts of fluid, the following mappings are implemented:

- Hue – spectral slope
- Saturation – spectral spread (inverse)
- Value – (spectral centroid + RMS) * RMS

The spectral slope is used to sample a colour from the hue wheel. Audio with more low-frequency content will sample from the lower colours in the hue wheel, and vice-versa. The spectral spread is inversely mapped to the colour saturation. Therefore, narrowband signals will produce very saturated colour, which has been sampled from the hue wheel according to the spectral spread.

Conversely, broadband signals will produce more diluted colour. The spectral centroid controls the value (brightness). As mentioned previously, there are various existing studies that suggest spectral centroid as a good indication of the perceived 'brightness' of a sound.

3.3.4.2 Texture

The various parameters of the fluid motion are parametrically controlled using the following mappings:

- Vorticity scale – spectral flatness
- Velocity dissipation – spectral centroid
- RGB dissipation – RMS
- Velocity viscosity – spectral spread (inverse)
- RGB viscosity – spectral spread (inverse)

The spectral flatness, which indicates how 'noisy' the signal is, controls the vorticity scale. This means noisier signals will produce more fine-grain textures in the fluid. As mentioned previously, this correlation has been reported in the existing literature. The spectral centroid controls the rate of velocity dissipation. Signals with higher spectral centroid values will produce more 'active' fluid.

3.3.4.3 Mass / Volume

The initial velocity of the fluid or 'burst strength' is controlled by the root-mean-squared (RMS) amplitude of the signal. This way, the size of the initial motion is controlled by the volume. The RMS also controls the base brightness of the fluid burst (which is then augmented using the spectral centroid). The spectral spread inversely controls the viscosity of the fluid. This means that signals with more low-end energy will produce more 'dense' fluid. An association between the 'density' of a sound and the low frequency energy content has been reported in Alluri & Toiviainen (2010).

4. DISCUSSION

The key objective of this paper and of the systems described within is to introduce a perceptually motivated methodology by which musical timbre can be visualised. Existing research into the perception and description of timbre suggests that it is often conceptualised with reference to physical, material and visual qualities. These visualisation systems attempt to draw upon perceptually and semantically meaningful mappings from low-level audio features to low-level visual features such that certain visual metaphors for timbre can be brought to life in real time.

4.1 Generative Geometry

The generative geometry visualisation system visualises the audio as a graphical 3D object with different surface texture, physical form, and luminance qualities. As mentioned in Section 3.2, one of the main motivations for the choice of visualisation was its effectiveness in representing the notion of a sound *source*. This is particularly important in the context of digitally produced sound, where the physicality of the sound source is less important, and visual metaphors may be drawn upon to a larger extent.

Since the audio is represented as a singular physical entity, this kind of visualisation could be well suited for monophonic sounds, visualised in isolation. The system could be put to use in contexts involving the construction and manipulation of timbres (e.g. sound design). In the context of interfaces for sound design, there is often a gap between what Seago, Holland and Mulholland (2004) call task-language ('bright,' 'punchy') and core-language ('filter cut-off,' 'envelope attack time'). Users may have a clear idea of the kind of sound they are aiming for, but may be restricted by their lack of technical knowledge of sound design. This kind of visualisation could be used to provide users with a perceptual – rather than engineering-focused – approach to sound design.

The audio object is visualised as being situated in a virtual space. This spatial quality opens up the possibility that the object could be interacted with in the virtual space. For example, 3D movement within the space could alter the physical qualities of the sound object, and in turn alter the timbre of the audio. The main author is currently exploring such an extension to the system involving motion tracking.

4.2 Fluid Simulation

The fluid simulation visualisation system visualises the audio as a continuous system of fluid. The audio features drive the parameters of different parts of the system in real time. The fluid system is very complex, and the behaviour of the fluid can be quite unpredictable due to the global interaction of the fluid. For this reason, this kind of visualisation is more suited to solely the representation (rather than manipulation) of timbre. For example, generative visualisations such as this are often used in audio-visual performances and installations.

The stochastic nature of fluid simulations is often drawn upon in order to simulate real-world phenomena that also have seemingly stochastic properties (e.g. fire, water). One possible avenue of research could be to use the audio features of an audio sample of one such real-world phenomenon (or even a Foley artist's impression) in order to influence the graphic simulation.

5. CONCLUSION

This paper has provided some background on the perception and description of timbre. It has discussed apparent links between aural perception of timbre and visual, material and textural qualities that have been reported in existing literature. These correlations have been used to introduce a perceptually motivated methodology for the visualisation of timbre. Mappings from audio features to visual features have been implemented in specific contexts, with reference to audio-visual perceptual correlations. Two example systems have been used in order to demonstrate this methodology: one using generative geometry and one using fluid simulation. Each system has been discussed in terms of potential applications and extensions, and further avenues of research have been proposed in each context.

The demonstration systems outlined in this paper have been developed using existing research into visual metaphors for timbre. Future work could include the use of such systems in lab studies and user studies in order to collect user data for mapping preferences, which can be compared and contrasted with existing results in the literature.

6. REFERENCES

Alluri, V. and Toiviainen, P. (2010) Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception,* 27(3), pp.223-241.

American Standards Association and Acoustical Society. (1960) *American standard acoustical terminology (including mechanical shock and vibration)*. American Standards Association.

Beal, A. L. (1985) The skill of recognizing musical structures. *Memory & Cognition,* 13(5), pp.405–412.

Beauchamp, J. W. (1982) Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society*, 30(6), pp.396–406.

Berger, K. W. (1964) Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, 36(10), pp.1888–1891.

Berthaut, F., Desainte-Catherine, M., and Hachet, M. (2010) Combining audiovisual mappings for 3d musical interaction. *International Computer Music Conference*, pages 100.

Blinn, J. F. (1977) Models of light reflection for computer synthesized pictures. *SIGGRAPH*, 11, pp.192–198.

Caruso, V. C. and Balaban, E. (2014) Pitch and timbre interfere when both are parametrically varied. *PLOS ONE*, 9(1).

Dannenberg, R. B. (1993) Music Representation Issues, Techniques, and Systems. *Computer Music Journal*, 17(3), pp. 20-30.

Disley, A. C., Howard, D. M., and Hunt, A. D. (2006) Timbral description of musical instruments. *International Conference on Music Perception and Cognition*, pp.61–68.

Grey, J. M. (1975) *An exploration of musical timbre*. Dept. of Music, Stanford University.

Grey, J. M. and Gordon, J. W. (1978) Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, 63(5), pp.1493–1500.

Giannakis, K. (2001) Sound mosaics: a graphical user interface for sound synthesis based on audio-visual associations. *PhD thesis*, Middlesex University, UK.

Kendall, R. A., Carterette, E. C., and Hajda, J. M. (1999) Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, pp. 327–363.

Krimphoff, J., McAdams, S., Winsberg, S., Petit, H., Bakchine, S., Dubois, B., Laurent, B., Montagne, B., Touchon, J., Robert, P. (1994) Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification. *J. de Physique*, 4, pp.625–628.

Krumhansl, C. L. and Iverson, P. (1992) Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), pp.739.

Marozeau, J. and de Cheveign´e, A. (2007) The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America*, 121(1), pp.383–387.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995) Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological research,* 58(3), pp.177–192.

Melara, R. D. and Marks, L. E. (1990) Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48(2), pp.169–178.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (2011) The timbre toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, 130(5), pp.2902–2916.

Platt, J. R. and Racine, R. J. (1985) Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception & Psychophysics*, 38(6), pp.543–553.

Pratt, R. and Doak, P. (1976) A subjective rating scale for timbre. *Journal of Sound and Vibration,* 45(3), pp.317–328.

Schubert E., and Wolfe, J. (2006) Does timbral brightness scale with frequency and spectral centroid? *Acta acustica united with acustica*, 92(5), pp.820–825.

Schubert, E., Wolfe, J., and Tarnopolsky, A. (2004) Spectral centroid and timbre in complex, multiple instrumental textures. *Proc. 8th Int. Conf. on Music Perception & Cognition (ICMPC)*.

Seago, A. (2013) A new interaction strategy for musical timbre design. In *Music and human-computer interaction*, pp.153–169. London: Springer.

Seago, A., Holland, S., and Mulholland, P. (2004) A critical analysis of synthesizer user interfaces for timbre. *Proceedings of the XVIIIrd British HCI Group Annual Conference*. Retrieved from Citeseer.

Tsiros, A. (2013) The Dimensions and Complexities of Audio-Visual Association. *Proceedings of Electronic Visualisation and the Arts,* London.

Tsiros, A., Leplatre, G. (2015) AniMorph: Animation Driven Audio Mosaicing. *Proceedings of Electronic Visualisation and the Arts*. London.

von Bismarck, G. (1974) Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica united with Acustica,* 30(3), pp.146–159.

Wapnick, J., Freeman, P. (1980) Effect of dark-bright timbral variation on the perception of flatness and sharpness. *Journal of Research in Music Education,* 28 (3), pp.176–184.

Warrier, C. M. and Zatorre, R. J. (2002) Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, 64(2), pp.198–207.

Wedin, L. and Goude, G. (1972) Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology,* 13(1), pp.228–240.

Wessel, D. L. (1978) Low dimensional control of musical timbre. In *Audio Engineering Society 59th Convention,* p.1337.

Wolpert, R. S. (1990) Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception,* pp.95–105.

Zacharakis, A., and Pastiadis, K. (2015) A confirmatory approach of the luminance-texture-mass model for musical timbre semantics. *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, page 39. ACM.

Zacharakis, A., Pastiadis, K., Papadelis, G., and Reiss, J. D. (2011) An investigation of musical timbre: Uncovering salient semantic descriptors and perceptual dimensions. *ISMIR,* pp.807–812.

Zacharakis, A., Pastiadis, K., and Reiss, J. D. (2014) An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception: An Interdisciplinary Journal*, 31(4), 339–358.

Zacharakis, A., Pastiadis, K., Reiss, J. D., and Papadelis, G. (2012) Analysis of musical timbre semantics through metric and non-metric data reduction techniques. *Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC12) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM 08)*, pp.1177–1182.