Towards a perceptual framework for interface design in digital environments for timbre manipulation

Sean Soraghan*
ROLI Ltd.
2 Glebe Road
London, E8 4BD, UK
sean@roli.com

Alain Renaud
Center for Digital Entertainment
Bournemouth University
BH12 5BB
arenaud@bournemouth.ac.uk

Ben Supper†
ROLI Ltd.
2 Glebe Road
London, E8 4BD, UK
ben@roli.com

ABSTRACT

Many commercial software applications for timbre creation and manipulation feature an engineering-focused, parametric layout. This paper argues the case for a perceptually motivated approach to interface design in such tools. 'Perceptually motivated' in this context refers to the use of common semantic timbre descriptors to influence the digital representation of timbre. A review is given of existing research into semantic descriptors of timbre, as well as corresponding acoustic features of timbre. Discussion is also given on existing interface design techniques. The perceptually motivated approach to interface design is demonstrated using an example system, which makes use of perceptually relevant mappings from acoustic timbre features to semantic timbre descriptors and visualises sounds as physical objects.

Author Keywords

Timbre, mappings, visualisation, interface design, timbre space

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing—Signal analysis, synthesis, and processing, H.5.2 [User Interfaces]—Graphical user interfaces (GUI), J.5 [Arts and Humanities]—Performing arts, I.3.5 [Computational Geometry and Object Modeling]—Curve, surface, solid, and object representations

1. INTRODUCTION

Historically, the notion of timbre has been linked to instrumentation or, more generally, sound source [27]. Similarly, alterations of timbre have been linked to articulation and gesture (e.g. pizzicato, arco). Digital audio synthesis by machines has facilitated the creation and manipulation of widely varying timbres. With digital synthesis, the manipulation of such timbres can be controlled via any number of conceivable interfaces that link with the digital synthesis engine [16].

The use of multidimensional scaling (MDS) techniques in the study of timbre perception has highlighted the multidimensional nature of timbre [13]. Using such techniques, timbres can be situated as points within a multidimensional 'timbre space' where distances between points are indicative of perceived difference in timbre [19]. The global timbre space can then be conceived as a continuous control interface where movement between different points can be used to alter the timbre [30].

One can consider the timbre spaces afforded by traditional musical instruments. The mode of exploration through such timbre spaces exists in the different articulation techniques used on the instruments. With digital synthesis engines, there are often high numbers of parameters that can be altered in order to manipulate the timbre. Together, these parameters make up a parameter space, where each parameter alters the position of the timbre along one or more perceptual dimensions within the timbre space. In commercial software, the presentation of such parameter spaces usually consists of arrays of rotary knobs and sliders. The layout and positioning of the individual controls follows the conventions of sound design, beginning with sound generation and following the signal flow through processes such as filtering and effects. Such interface design is suited to the focused creation of individual timbres, but can be limiting to the process of exploration or performative manipulation.

This paper makes the case for a *perceptually* motivated approach to interface design in digital tools for timbre creation and manipulation. It is argued that the exploration of parameter spaces in digital environments could be visually guided using a perceptually motivated visualisation framework that makes use of mappings between acoustic features of timbre and semantic descriptors of timbre. It is also suggested that a perceptually motivated interface design framework could be effective in building more performative, exploration-oriented interfaces.

The rest of the paper is structured as follows. Section 2 discusses some existing research into the semantic description of timbre, as well as research into mappings between acoustic features and semantic descriptors. It also discusses existing modes of interaction for the creation and manipulation of timbre, categorising them into the different use-cases to which they are suited. Discussion is given on how a perceptually motivated visualisation framework could be useful to each mode of interaction. Section 3 describes an example system developed using existing research into timbre perception, semantic description of timbre, and acoustic features of timbre. This example system is used to argue the case for a novel interface design paradigm. Section 4

*EngD Researcher with the Centre for Digital Entertainment (CDE) and R&D software engineer at ROLI.

†Head of research at ROLI.

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

provides some discussion and section 5 provides conclusions and summation.

2. RELATED RESEARCH

2.1 Semantic Descriptors

Anlaysis 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 [25]. Multiple studies have identified that listeners often use visual metaphors when describing timbre, with semantic descriptors being material, textural and physical in nature [29, 21, 6].

Three salient clusters relating to: *volume/wealth*; *brightness and density*; and *texture and temperature* have been identified in [32]. This has since been referred to as the 'luminance texture mass' model in a follow-up study [34] and confirmatory studies have been conducted which show that the model is consistent across two different languages (English and Greek) [33, 31].

2.1.1 Mapping acoustic features to semantic descriptors

A brief summary of research into mappings between acoustic features and semantic descriptors will now be given. The acoustic features that are mentioned here will be explained in detail in section 3.2.

There has been a lot of research into the correlation between acoustic features of timbre and perceptual dimensions in timbre spaces. Studies have shown associations between spectral centroid [14, 18, 17, 19], attack time [18, 19], and spectral variation [18, 19] for particular dimensions. Three key techniques have been used in the attempt to identify mappings between acoustic features and semantic descriptors: user studies, correlation analysis and machine learning.

User Studies

Giannakis and Smith have carried out a number of studies looking at auditory-visual perceptual correlates [11, 9, 10]. An investigation into the similarities between visual texture and auditory timbre is presented in [9]. The following associations were highlighted: *texture contrast – spectral centroid*; *texture granularity – noisiness*; *texture repetitiveness – distance between pairs of partials*.

In [2], a user study was conducted in which participants were presented with various audio-visual stimuli and their mapping preferences were measured. The audio stimuli involved pitch and volume. The timbre-related features used were noisiness and spectral centroid. Results suggested that participants preferred brightness ('colour lightness') as a visual mapping for spectral centroid, and texture rugosity as a visual mapping for noisiness.

Correlation Analysis

By conducting correlation analysis between acoustic features and semantic descriptors, Zacharakis et al. [33] found the following correlations: *harmonic energy distribution – texture*; *spectral centroid variation and inharmonicity – thickness (related to mass and luminance)*; *fundamental frequency – mass or luminance (depending on native language)*. There is a large consensus for the association between spectral centroid and the perceived 'brightness' of a sound [1, 5, 23, 22].

The results of various user studies and regression analysis studies are summarised in the 'suggested mappings' column of table 1.

One issue that has been reported with such correlation analysis is that results often show that specific acoustic features or perceptual dimensions (in the case of timbre space analysis) are correlated with *multiple* semantic descriptors [7, 8]. This makes the construction of mapping strategies complicated, since it suggests the requirement of *n-m* mappings (as opposed to 1-1) from acoustic feature to semantic descriptor.

Machine Learning

Machine learning techniques can be used for automatic classification of timbres (see [15] for a review). By defining a set of synthesis parameters and using them as inputs to a classifier, a measurement can be obtained of how much a given parameter configuration fits a given timbre label. An iterative process can then be employed in order to quantify how much each parameter needs to change in order to approach a given timbre label, or semantic descriptor. Such a technique was used by Gounaropoulos and Johnson [12].

An advantage of using machine learning to extract mappings in this way is that arbitrarily complicated *n-m* mappings can be obtained, which facilitates the representation of groups of synthesis parameters using global perceptual labels. However, this may also be a drawback, as such mappings will be black-box in nature, and hard to inspect.

2.2 Interface Design

This section gives a brief review of three different strategies for user interface design, namely: *sound design*, *performance*, and *search and retrieval*.

2.2.1 Sound Design

The term *sound design* is being used in this context to refer to the construction of specific timbres, from the ground up. Most commercial applications for sound design are geared towards this use-case. As mentioned previously, the presentation of parameter spaces in such interfaces usually consists of arrays of rotary knobs and sliders and follows the conventions of sound design from an engineering perspective. This type of interface design is rooted in the mimicking of modular synthesis hardware from the sixties to the eighties, and has remained the industry standard despite the development of novel synthesis techniques [26]. It is based around what Seago et al. [26] call a 'user specified architecture'. The user patches together sound sources, filters and effects, and then sets the parameters of these various modules in order to construct a required timbre. In the context of timbre spaces, this interaction paradigm effectively involves two steps: the construction of the timbre space (linking modules), and the subsequent extraction of a point, or area, within that timbre space (setting parameters). Often what is lacking with interfaces such as these is a strong link between 'task language' (e.g. 'bright,' 'punchy') and 'core language' (e.g. 'filter cutoff', 'envelope attack time') [26]. One way in which this link could be achieved is through the use of a central visualisation of the timbre, based on a perceptually motivated visualisation framework.

2.2.2 Performance

Performance in this context refers to the real-time manipulation of timbre in a performative setting. Such interfaces are usually built on top of what Seago et al. [26] refer to as a 'fixed architecture'. In other words, what the majority of these interfaces offer is some mode of exploration through an existing underlying timbre space. Indeed, commercial software packages based around user specified architectures often have simplified counterparts, where the architecture is fixed, and a collection of specific parameters or macro controls (that alter multiple parameters) is presented to the user for simple – sometimes performative – exploration. Again, interfaces such as this would benefit from a perceptually motivated visualisation that would guide such exploration.

In the case of performance-oriented interfaces, the issue of mappings is of particular importance. Since the connection between interaction and sound generation is disconnected in digital performance interfaces, effective mappings need to be designed that afford intuitive and expressive performance capabilities [16]. The use of perceptually motivated visualisation of timbre features could aid in the design of intuitive modes of interaction. Such mappings could allow direct interaction with a virtual object whose visual features represent different aspects of the timbre.

Some performative interfaces make use of novel synthesis techniques. One such technique is corpus-based concatenative synthesis (CBCS), which involves the re-synthesis of a target set of sound characteristics [24]. These target characteristics are either explicitly provided, or extracted through audio analysis of a target sound. A corpus of small sound snippets or 'grains' is then referenced to find a set that together can provide a sound with the desired characteristics. CBCS therefore provides access to a high-dimensional timbre space, where the dimensions are defined as the various sound characteristics that are used in the labeling and referencing of sound grains.

2.2.3 Search and Retrieval

The third type of interaction is based around the search and retrieval of timbres using some form of user-guided search algorithm [25, 12]. A common method is to use interactive genetic algorithms (IGAs), however these can take a long time to converge on a solution since GAs require multiple generations, and user-evaluation of individual solutions is very time consuming [28]. Alternative techniques are proposed in [25] and [12]. Effectively, user exploration through a high-dimensional timbre space is guided by continuous user evaluation of reference sounds in comparison to a required target sound. This kind of interface is suited to users without knowledge of sound design but who require specific types of sounds.

By constructing robust complex mappings between sound parameters and semantic labels, the process of searching through a timbre space can be guided by semantically meaningfully labeled controls [12]. Again, the drawback to this is that the mappings are hidden within the black-box machine learning algorithm.

3. EXAMPLE SYSTEM

This section will give details of an example system that makes use of mappings between acoustic features and semantic descriptors of timbre. This interface affords performative exploration of a synthesis parameter space which is guided by semantically motivated visualisation of the timbre. The system constitutes a 'simplified counterpart' (as discussed in section 2.2.2) to an underlying synthesis engine with user specified architecture.

3.1 System Overview

The system makes use of the Microsoft Kinect 2.0 sensor in order to track the user's hands. The front-end of this system consists of a virtual 3D space within which the user's hands are represented, along with a 3D visualisation of the audio being generated by the sound engine. Real-time audio analysis functionality is built into the audio engine, which continuously extracts timbre descriptors from the audio that is generated. These timbre descriptors are used to drive the visualisation in the user-facing front end. In turn, the X, Y and Z distances between the user's hands are mapped to different parameters within the sound engine. The user can thus drive the synthesis engine through gestural motion of their hands within the space. This leads to both audible changes in the sound and corresponding visual changes to the visualised object. Figure 1 shows a flow diagram of the system.

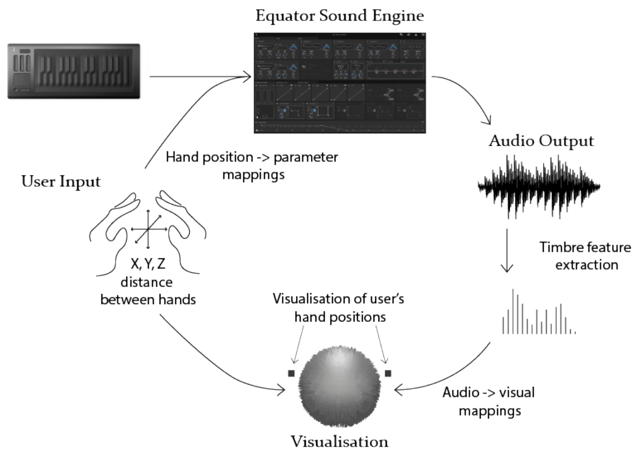

Figure 1: System overview.

3.2 Timbre Feature Extraction

There are two types of features used in this system: spectral and harmonic. Spectral features are calculated for windows of 256 samples, and harmonic features for windows of 2048 samples. Spectral features consist of *spectral centroid*, *spectral spread* and *spectral flatness*. Harmonic features consist of *harmonic energy ratio* and *inharmonicity*.

These timbre features were chosen based on their correlation with semantic descriptors from previous research. An extensive list of acoustic features is provided in [20]. The features used in this example system are all taken from that list. Table 1 shows the mathematical definition of these features, and table 2 provides a mathematical notation key.

F0 estimation

An estimation of the fundamental frequency - f_0 - is required in order to calculate the harmonic features. It can be estimated by performing peak detection on the spectrum. A histogram of frequencies can then be constructed that contains the differences between each peak and every other peak. These frequencies can then be weighted using some sort of scoring function. The scoring function used here is the harmonic energy ratio. The peak frequencies in the spectrum resulting from peak detection are used as the main partials.

3.3 Visualisation

The system visualises the audio being generated by the sound engine as a spherical structure with surface deformations, where the brightness, shape, and texture of the sphere are used to indicate different semantic descriptors of the timbre.

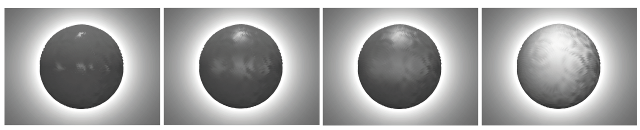

Figure 2: Varying specular radius value.

Table 1: Acoustic Timbre Features and Semantic Mappings

Acoustic Feature	Mathematical Definition	Suggested Semantic Mapping(s)
Spectral Centroid	$\mu_1(t_m) = \sum_{k=1}^{K} f_k \cdot p_k(t_m)$	Brightness [1, 5, 2, 33]
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 spread is a measure of the spectrum bandwidth
Spectral Flatness	$SFM(tm) = \frac{\left(\prod_{k=1}^{K} a_k(t_m)\right)^{1/K}}{\frac{1}{K}\sum_{k=1}^{K} a_k(t_m)}$	Texture [9, 2]
Harmonic Energy Ratio	$HER(tm) = \frac{1}{E(t_m)} \sum_{h=1}^{H} a_h^2(t_m)$	Texture [33] Texture granularity [9]
Inharmonicity	$inharm(t_m) = \frac{2}{f_0(t_m)} \frac{\sum_{h=1}^{H} (f_h(t_m) - h f_0(t_m)) a_h^2(t_m)}{\sum_{h=1}^{H} a_h^2(t_m)}$	Texture repetitiveness [9]

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 $h^{\cdot}$
E_{tm}	The total energy in the spectrum at time window m

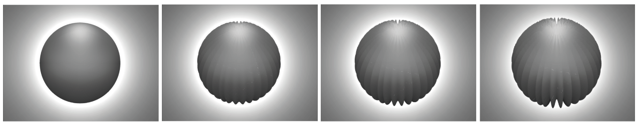

Figure 3: Spherical vertex extrusion dependent on azimuth position.

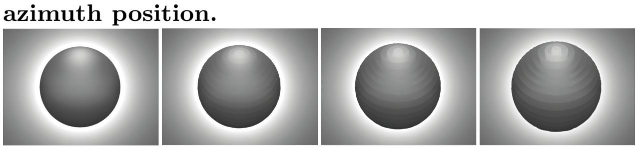

Figure 4: Spherical vertex extrusion dependent on inclination position.

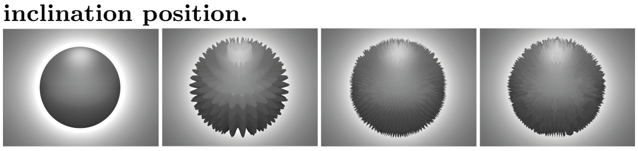

Figure 5: Spherical vertex extrusion dependent on azimuth and inclination position.

Luminance / Brightness

The sphere is rendered using Blinn-Phong shading [3]. 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 2 shows the affect of varying the specular radius value. A small specular radius leads to local specular highlights, whereas a large specular radius leads to widespread illumination.

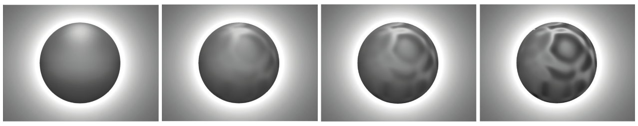

Figure 6: Perlin noise bump mapping with varying amount.

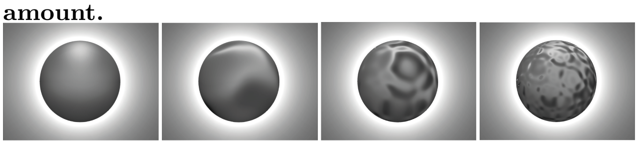

Figure 7: Perlin noise bump mapping with varying granularity.

Texture

The spherical structure consists of a spherical model and makes use of a vertex extrusion shader, which extends the vertices along their normals. The amount of deformation per vertex is dependent on its spherical coordinates (azimuth and inclination position). Figures 3 and 4 show the affect of varying the azimuth and inclination extrusion amounts. Figure 5 shows the affect of varying both simultaneously. The frequency of deformation can also be altered parametrically.

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 6 and 7 show the affect of changing the amount and granularity, respectively.

3.4 Mappings

Luminance / Brightness

As mentioned previously, multiple studies have found that the spectral centroid provides a decent indication of the brightness of a sound [1, 5, 23, 22]. Therefore, the spectral centroid was 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 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 [9] and [2].

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 [33]. The granularity of the bump mapping is controlled by inharmonicity, meaning that 'out of tune' audio will produce more granular bump deformations (figure 7). Inharmonicity has been linked to texture-repetitiveness in [9].

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 [33].

3.5 Gestural Control of Synthesis Parameters

The user‘s hands are tracked and visualised in virtual 3D space. The X, Y and Z distances between the hands are tracked and used to control different parameters within the Equator sound engine. The parameters that these distances control can be varied, and depends on the configuration of the sound engine for a given preset. The mapping of these distances to different synthesis parameters is what constitutes the creation of an exploration interface for a given fixed architecture configuration of the sound engine. In other words, the creation of a 'simplified counterpart' interface as discussed in section 2.2.2.

This paper focuses on the audio-visual mappings used in the visualisation system which guides the gestural control. The specific gesture-audio mappings constitute an area of research in and of themselves and are an area of interest for future research and experimentation. For a review of gestural control of music see Cadoz et al. [4].

4. DISCUSSION

The motivation behind the development of this system was the prevalence of textural, material and physical vocabulary involved in the semantic description of timbre. The intention was to make use of existing research into both acoustic descriptors of timbre and semantic descriptors of timbre in order to develop a novel interface for timbre manipulation. The interface was conceived as an extension to an existing interface, which has an engineering-focused parametric layout (the Equator synth from ROLI). This extension provides a perceptually motivated front end to the sound engine, where, for example, 'rough' 'bright' sounds are presented as rough bright objects.

In the context of existing interface design, as discussed in section 2.2, this system has been designed in order to provide a performance-oriented interface. It was developed in order to demonstrate a novel form of exploration of a fixed-architecture parameter space, and the underlying timbre space. Navigation of the parameter space is semantically-driven, and involves parallel parameter control. It is guided by semantically motivated visualisation of the audio. This contrasts with the serial parameter-driven navigation afforded by the Equator interface.

Although this specific system makes use of a perceptually motivated timbre-visualisation framework for a performance use-case, such a framework would be of benefit in both sound design and search and retrieval use cases as well. In sound design contexts, it would aid in bridging the gap between task language and core language. In search and retrieval contexts, it could be used to provide visual reference points to assist in the orientation of a search.

5. CONCLUSIONS

This paper has introduced the problem of barriers to entry in many commercial interfaces for timbre creation and manipulation due to engineering-focused interface design and parametric layout. It has provided a concise review of some existing research into common semantic descriptors of timbre, and their corresponding acoustic timbre features. It has also provided a discussion on existing interface design techniques, in the context of three different user goals: sound design, performance, and search and retrieval.

An example interface has been described which provides a novel way to explore a synthesis parameter space in a performance context, by making use of a perceptually motivated timbre visualisation framework. The implementation of the audio visualisation functionality has been described in detail, with reference to existing research that motivated the choice of mappings from timbre feature to visual feature. Discussion has been given on how the perceptually motivated timbre visualisation approach could be put to use in the contexts of sound design and search and retrieval.

Future work could include in-depth evaluation of the interface, and comparisons with other existing interfaces, particularly in the area of gestural control. The interface could also be put to use in user-studies to investigate user preference for different audio visual mappings and compare results with existing research.

6. REFERENCES

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

[2] F. Berthaut, M. Desainte-Catherine, and M. Hachet. Combining audiovisual mappings for 3d musical interaction. In *International Computer Music Conference*, pages p–100, 2010.

[3] J. F. Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer*

Graphics, volume 11, pages 192–198. ACM, 1977.

[4] C. Cadoz, M. M. Wanderley, et al. Gesture-music. *Trends in gestural control of music*, 12:101, 2000.

[5] G. De Poli and P. Prandoni. Sonological models for timbre characterization*. *Journal of New Music Research*, 26(2):170–197, 1997.

[6] A. C. Disley, D. M. Howard, and A. D. Hunt. Timbral description of musical instruments. In *International Conference on Music Perception and Cognition*, pages 61–68, 2006.

[7] R. Ethington and B. Punch. Seawave: A system for musical timbre description. *Computer Music Journal*, pages 30–39, 1994.

[8] A. Faure, S. McAdams, and V. Nosulenko. Verbal correlates of perceptual dimensions of timbre. In *4th International Conference on Music Perception and Cognition*, pages 79–84, 1996.

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

[10] K. Giannakis. A comparative evaluation of auditory-visual mappings for sound visualisation. *Organised Sound*, 11(03):297–307, 2006.

[11] K. Giannakis and M. Smith. Auditory-visual associations for music compositional processes: A survey. In *Proceedings of International Computer Music Conference ICMC2000, Berlin, Germany.* Citeseer, 2000.

[12] A. Gounaropoulos and C. Johnson. Synthesising timbres and timbre-changes from adjectives/adverbs. In *Applications of Evolutionary Computing*, pages 664–675. Springer, 2006.

[13] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.

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

[15] P. Herrera, X. Amatriain, E. Batlle, and X. Serra. Towards instrument segmentation for music content description: a critical review of instrument classification techniques. In *International symposium on music information retrieval ISMIR*, volume 290, 2000.

[16] A. Hunt, M. M. Wanderley, and M. Paradis. The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, 32.

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

[18] J. Krimphoff, S. McAdams, S. Winsberg, H. Petit, S. Bakchine, B. Dubois, B. Laurent, B. Montagne, J. Touchon, P. Robert, et al. Characterization of the timbre of complex sounds. 2. acoustic analysis and psychophysical quantification. *J. de Physique*, 4:625–628, 1994.

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

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

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

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

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

[24] D. Schwarz. Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92–104, 2007.

[25] A. Seago. A new interaction strategy for musical timbre design. In *Music and human-computer interaction*, pages 153–169. Springer, 2013.

[26] A. Seago, S. Holland, and P. Mulholland. A critical analysis of synthesizer user interfaces for timbre. In *In Proceedings of the XVIIIrd British HCI Group Annual Conference.* Citeseer, 2004.

[27] D. Smalley. Defining timbre - refining timbre. *Contemporary Music Review*, 10(2):35–48, 1994.

[28] H. Takagi. Interactive evolutionary computation: Fusion of the capabilities of ec optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.

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

[30] D. L. Wessel. Timbre space as a musical control structure. *Computer music journal*, pages 45–52, 1979.

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

[32] A. Zacharakis, K. Pastiadis, G. Papadelis, and J. D. Reiss. An investigation of musical timbre: Uncovering salient semantic descriptors and perceptual dimensions. In *ISMIR*, pages 807–812, 2011.

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

[34] A. Zacharakis, K. Pastiadis, J. D. Reiss, and G. Papadelis. Analysis of musical timbre semantics through metric and non-metric data reduction techniques. In *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)*, pages 1177–1182, 2012.