

Decomposition and Classification of Timbres for Music Notation

Initial research and strategy to establish a scoring protocol for the resynthesis and performance of artificial sound sources

Keywords: Timbre Properties; Reduced Listening; Causal Listening; Invariant Classes, Decomposition of Complex Sounds; Preservation of Standard Scoring; Extensible Structure; Physical Actions; Metaphysical Qualities; Sources; Filters

Abstract

The deployment of synthesisers and artificial sounds should now be as natural as choosing a violin in orchestral pieces. However, there is a barrier to scoring, in that these instruments have no fixed timbre. This document proposes research into establishing a succinct system for describing timbres, which a sound designer can use to program the instrument. The instruction is to be placed before or after the score, defining not only the base timbre, but also how the sound will change with performance actions. Once programmed, the performer is afforded to read fairly standard and consistent music notation rather than a graphical score.

Introduction

Performance and interpretation has always been the lifeblood of music, but for a time recording was where money was made. However, the recent decline of monetary value in recorded music has rejuvenated the importance of performance as a source of funding. To be practical for anything more ambitious than a band of musical acquaintances, scoring is needed to efficiently communicate musical ideas from composer to performer. All is well and good with this time honoured method until synthesisers and artificial sounds are used. How does the performer recreate the intended sound? There are obvious solutions, such as hearing a recording of the timbre and emulating that, or a long list of parameter settings that assumes the performer has the same or similar equipment.

These methods work, but impose limitations on future instruments, plus from a music vocabulary stance, are somewhat clumsy. So much so, that the use of electronics in an orchestral setting still seems like a novelty, even though the technology has come of age and can rival any acoustic instrument for expressive versatility.

To further highlight the problem in scoring for electronics, consider how timbre variation is notated for acoustic instruments. This is usually achieved by giving a written cue for an action, such as bowing near the bridge for a brighter sound, or placing a mute into the bell of a horn for a thinner tone. No consideration of how to aurally describe that timbre variation is needed, as the instrument is a fixed physical object where both composer and player know to high degree of certainty to how an action will affect the sound. Conversely for electronics, performance direction has tended to be given via a graphical score, which works quite well and produces a level of synchresis if read while listening. However, scores like this are probably a nightmare to typeset, and require the performer to learn a non-trivial chunk of bespoke symbolism for each and every piece. What is being proposed in this document, is the development of a clear succinct system that not only describes the multifaceted character of the sound, but what happens to it through register, dynamic and controller change, clearly set out as a diagram before or after the music score. If this is well defined, then only simple direction for an action is needed during the score to achieve the desired colour, many of which are already standardised in music notation.

Of course, the exact intended timbre cannot be captured by a generalised description, but an approximation is argued to be good enough or even desirable, given that music is a performance art. Further, research into the perception of timbre suggests that the sum of the parts is what characterises an instrument. (Risset & Wessel, 1982)

One final disclaimer: Given the ambitious scope, this stage of research will concentrate on a special case where the resynthesis of artificial sounds is played on a suitable interface. It is believed that a general system, that will encompass Acousmatic music and other difficult to score forms, can be developed from this grounding.

Strategy

According to Mauricio A. Lourerio et.al

“Timbre is perceived by means of the interaction of a variety of static and dynamic properties of sound grouped into a complex set of auditory attributes.” (2004, p. 1)

Timbre is more than just the steady-state relationship of a pitch and overtones as long conjectured, (Rasch & Plomp, 1982, p. 13) and even though this time-invariant element is important, especially the formant or spectral envelope¹, (Risset & Wessel, 1982, p. 28) it doesn't express the vital temporal information that our hearing is so sensitive to. This mix of aperiodic and periodic components can also transmute with time, register and dynamic intensity. Consequently, to describe a complex multidimensional timbre will require decomposition into simpler, but self-contained timbre blocks which for conceptual distinction, shall be called sonic components from here on. A visual example is shown in Figure 1 where the horizontal axis is time and the vertical shows where sonic components can overlap.

Fortunately, sound is additive and no component of a complex timbre will occlude another. (Handel, 1995) Given this property, it is tempting to find a way to describe each of these components like the ingredients of a recipe.

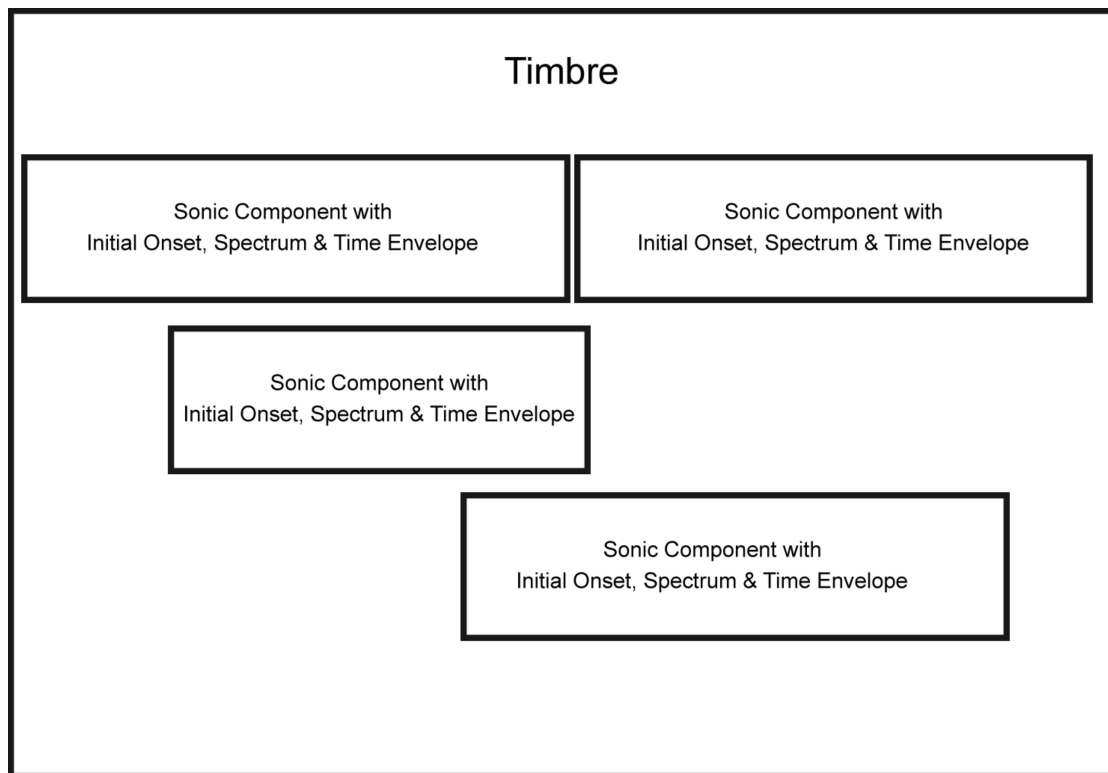


Figure 1

Classification is required to keep the description formal and obvious, and should be non-specific to any technology. One possible approach for suitable terms is to leverage the

¹ Formants are quite invariant for a register band and are a robust identifier of a source, even when there is significant harmonic distortion. Formants often have temporal metamorphosis, which further sharpens the identification.

tendency of people to assign a cause or process to sounds. (Chion, Audio-Vision, 1994) This is near blasphemy for Acousmatic proponents, but a causal description² of each sonic component may help to shortcut a long and non future-proof parameter list, but remain specific enough to allow for resynthesis. In addition, it would also be prudent to include Pierre Schaeffer's abstracted morphological classification into the system, as it will clarify the musical and perceptual character of the timbre. (Schaeffer, 2017) (Chion, Guide Des Objets Sonores, 1983)

So the first stage is to see if it is reasonable to describe an artificial sound in terms of a physical process and if limitations are found, can sensible workarounds be used? If all appears well and good, a diagrammatic system to convey the terms and numeric values that is succinct and adaptable will be created. Finally, a scheme for performance with controllers needs to be established, while preserving as much as possible, the standard and well-understood conventions of written music.

Plausibility

Is it reasonable for an artificial sonic component to be described by a similar sounding physical cause?

Sound Design

Foley artists and sound designers have always understood that many unrelated sound sources have similar sonic properties, and this is used to great effect in film and television through synchresis, but the results can be convincing enough to be also used in radio broadcasting or any aural-only medium. They also understand that to fully create an illusion often requires the stacking and time offsetting of disparate sounds that combine to produce a desired output.

In order to argue if this observation can be applied to all forms of timbre description, whether natural or artificial, it is worth assessing from first principles what is invariant for all sound generation.

Filters

Harmonic oscillation at a single frequency and amplitude is unchanging, thus carries³ no information. (Watkinson, 2002) (Shannon, 1948) At the other extreme, white noise,

² A causal listening description for a reduced listening outcome

³ The term *carry* is important, since the phenomena has intrinsic meaning, but carries no contextual or new information

having random fluctuating amplitudes for all frequencies, averaged over an entire given bandwidth, has maximum information entropy and carries no decipherable source origin. In contrast, if a given subset of all frequencies is at zero amplitude, and the remaining frequencies have amplitudes at disparate and time varying values, then the sound has a characteristic and unique quality, which carries information about its source, as opposed to the first two examples. This process is clearly recognised as filtering in engineering, but the relevant point being made, is that all resonating bodies do just that. Physical objects reinforce and attenuate particular frequencies, and the variable latitude for this filtering is unimaginably large, and perhaps, all encompassing. Given this scope, it appears plausible to find a way of equating an artificial filtering process with the way a physical object reacts to excitation, allowing one to describe the spectra of artificial sonic components in terms of familiar constructs and materials.

However, there is a caveat. Although using causal description seems like a good way to circumvent the ‘heavy lift’ of describing the complex spectral envelope of a variable filter, the choices have to be fairly nonspecific in order to limit the number of terms used⁴, and yet not restrict the descriptive precision, hence the need for classification. Plus, if a generic (and timeless) synthesiser process is simpler to describe, then that can be used in conjunction with the causal terms to further clarify the intended timbre⁵.

Sources

In signal theory all of the filtering of sinusoids in materials or a system, can be fully characterised as a response to an ideal impulse called a Delta Function. (Hartmann, 1995) This is stated to show that an ideal exciter analytically isolates the filter from the source of energy, but real exciters impart filtered energy to form part of the resulting sound. However, this can be turned to an advantage, as the most complex part of a timbre is the transient phases, such as the initial onset. It is certainly easier to describe this in terms of an action with a particular object than to give values for a barrage of multistage function generators controlling filters, gain, noise etc.

So, to have a reasonable description of a particular sonic component, the action and material qualities of the exciter have also to be considered, with the terms chosen from common experience. Therefore, forming an interaction⁶ of exciter and resonator in the

⁴ ‘Large, resonant timber-like box’ is better than ‘Tea Chest’ or ‘Furniture Cabinet’ etc.

⁵ For example, [Roughness >> 10% modulation of gain using 32Hz sine LFO] would create intermodulation distortion

⁶ Convolution

mind of the sound designer. To make terms more universal, from now on the exciters can be labelled as source and the resonators as filters. (Handel, 1995)

There is also a second class of source, which can be thought of as a disruption or an influence, which can be introduced in tandem with the primary source or at a later time. These can range from simple damping (perhaps with a fabric) through to feedback, where the whole sound recursively excites itself.

What has been argued in this section is that the causal model of a source and filter, both used in physics and physical modelling synthesis is conjectured to lend itself to describing a single sonic component of any artificial timbre. However, this does not stipulate that a physical modelling synthesiser should be used, or indeed that the resulting completed timbre has an actual physical equivalent. Also, since a practical and clear lexical system is being proposed, then well understood universal techniques and effects⁷ from electronic sound synthesis, which may be difficult or impractical to convey with the above model, may be used in conjunction with the physical descriptors.

The Search for Appropriate Terms and Numeric Values

Schaeffer's Morphological Sound Description

Pierre Schaeffer's model of a sound object extends out to a connected musical idea, explained well in the paper *Morphological sound description: computational model and usability evaluation...*

“In a sound stream, any entity perceived as having its own internal properties and rules is considered as a sound object.” (Ricard & Herrera, 2004, p. 2)

However, much of the model can lend itself to conceptually smaller units such as a timbre, and since the criteria is well defined, complete with shorthand notation, it would be foolhardy to ignore.

Schaeffer's criteria can be split into three major categories:

- Matter – The steady-state character of the sound
- Shape – The envelopes of amplitude
- Variation – The temporal change of the Matter

⁷ For example, a high resonance filter sweep is probably easier to state as it is, than to describe from an equivalent sound source from nature. Although twanging and stretching a rubber band comes close!

Each of these criteria has a great deal of detail, which will be looked at during the next stage of research, but testing by Ricard and Herrera (2004, pp. 6-7) just using Schaeffer's Morphology, produced encouraging results. It suggests that the system, with adaption, may consolidate the more specific causal language proposed in this document, leading to better results.

Causal Descriptors

The source can be thought of as a material object that transfers energy via an action. To describe them, it should be quite trivial to collate terms for the actions, and if one is left out, then that can be simply added. Common musical terms are struck, bowed, blown and plucked. These are then associated with the material object that performs the action, like a metal beater or a plastic pick. The basic actions can also be extended non-exhaustively to include scraping, brushing, bouncing, settling (like a coin spinning down), twanging and squeaking.

Filters are more difficult to pin down, one could use basic materials such as wood or metal, but a quick experiment will show the folly of this approach. For instance, striking with one's knuckle, a lightly held earthenware bowl will create a metallic-glass enharmonic ringing, like a bell. However, held firmly to damp, the quality is more like a wooden block. The experiment suggests that shape and material elasticity are more important for the spectral envelope, but that may be vague and difficult to classify without further inquiry. For future research, impulse response analysis and a statistical technique called Principle Component Analysis (PCA) may uncover some commonality in material-shape combinations, which can then be reduced to a generic description. (Güven & Özbayoglu, 2012) (Sandell & Martens, 1995) (Serra & Smith, 1990)

Another approach is to give coordinates for a two dimensional plane, where the x-axis is harmonic to enharmonic, and the y-axis is resonant to damped. This would account for many resonant systems as a computer algorithm, however it is open to misinterpretation as a sole written description.

It is proposed that the way forward is to use all the above approaches. Using a judicious choice of criteria that when used in combination, provides the required accuracy through triangulation, without invoking an arbitrarily large value of object, material and coordinate terms.

Timbre Synthesis Diagram and Temporal Notation

Proposal for the Diagram

Below is a first draft of the diagram to be placed in the preface or appendix of a score. Figure 2 describes an example of a timbre with three⁸ components (C1 to C3), at a middle pitch register and dynamic. C1 can be considered the core component timbre; therefore the other components are given gain and register values relative to C1 marked up in their header. Components can be offset in time, as with C3, and this time can be marked as accurately as desired. Each component description has two rows, where the top row is written in *Schaeffer* (S) style syntax, and the lower being the *Physical* (P) depiction. Finally, a block dedicated to what happens after the playing action has terminated (released) is provided and can be as detailed as required. To account for deviations in sonic colour over register and dynamics, a matrix of the affected parameter amendments forms the second part of the diagram. The example given in figure 3 is quite obvious and simple, but the system is completely extensible allowing any level of detail. It can also include more detailed targets such as specific key ranges, or further intermediate dynamics.

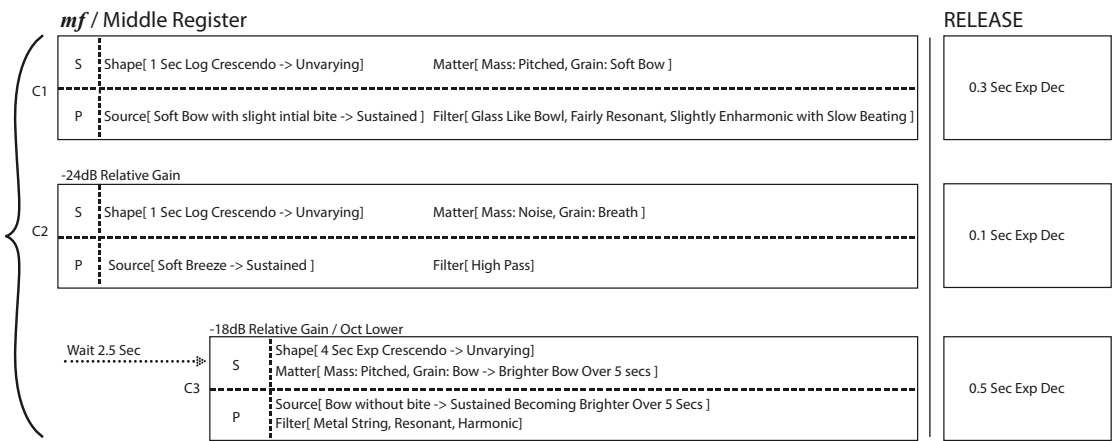


Figure 2

⁸ Of course, there can be as many as needed.

	Low Register	Middle Register	High Register
<i>pp</i>	[All Relative To MF LR] C1(Source:Less Initial Bite/ Filter:Darker/ Release:Slightly Shorter); C2(Less Relative Gain); C3(Filter:Darker/ Release:SlightlyShorter)	[All Relative To MF] C1(Source:Less Initial Bite/ Filter:Darker/ Release:Slightly Shorter); C2(Less Relative Gain); C3(Filter:Darker/ Release:SlightlyShorter)	[All Relative To MF HR] C1(Source:Less Initial Bite/ Filter:Darker/ Release:Slightly Shorter); C2(Less Relative Gain); C3(Filter:Darker/ Release:SlightlyShorter)
<i>mf</i>	C1(Source:More Initial Bite/ Release:Slightly Longer); C2(Less Relative Gain); C3(Release:Slightly Longer)		C1(Source:Less Initial Bite/ Release:Slightly Shorter); C2(More Relative Gain); C3(Release:Slightly Shorter)
<i>ff</i>	[All Relative To MF LR] C1(Source:More Initial Bite/ Filter:Brighter/ Release:Slightly Longer); C2(More Relative Gain); C3(Filter:Brighter/ Release:Slightly Longer)	[All Relative To MF] C1(Source:More Initial Bite / Filter:Brighter/ Release:Slightly Longer); C2(More Relative Gain); C3(Filter:Brighter/ Release:Slightly Longer)	[All Relative To MF HR] C1(Source:More Initial Bite/ Filter:Brighter/ Release:Slightly Longer); C2(More Relative Gain); C3(Filter:Brighter/ Release:Slightly Longer)

Figure 3

Metamorphic Control

The matrix in figure 3 can be further extended to include variations introduced by a control. However, if that control is to invoke a completely different timbre, then a whole secondary diagram can be created and labelled for that control when set at the maximum value, implying that a control at median values will have elements of both timbres. This method can be further extended for an XY control to allow the transformation of several timbres.

Scoring for Controls

Figure 4 shows that using a staff to be expedient for marking control values. The bottom line is designated to be zero, the top line is the maximum value, and each inner space and line is an increment of 12.5%. More exact values (if required) can be written next to the limiting event, which are written by using a percussion type note-head⁹. Slurs can be used to imply transitions between two events and these can be shaped to guide the performance of the control. XY control will be two staves clearly grouped, but the principle is the same.

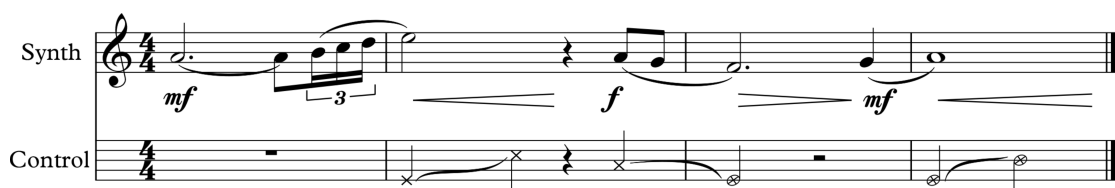


Figure 4

⁹ Easier to distinguish from the pitched staff when sight reading

Conclusion

What has been achieved?

This document started with a research strategy to devise an effective pre-score notation system for timbre resynthesis. It has given a plausibility argument that by describing simple sonic components of a timbre in terms of a physical cause is likely to shortcut long and detailed parameter lists. Finally, a draft is provided for a simple and extensible diagram system to place these terms, given in tandem with Schaeffer style morphological description. At the beginning of this research it wasn't clear if a system could be achieved, but the results at this stage, are very encouraging. However, the glaring omission is a concise and formal set of terms that will require more research to pin down.

What is the next stage of investigation?

Schaeffer's reduced listening experiments of recorded objects frustrated him, because of the tendency of making a causal association with the sound source. Brian Kane quotes Schaeffer's exasperation in his book *Sound Unseen*.

"Isn't the noise of [train] buffers first and foremost anecdotal, thus antimusical? If this is so, then there's no hope, and my research is absurd. " (2014, p. 28)

An effective description of artificial sounds in terms of cause has exactly the opposite problem, in that the composer has to listen very carefully to the desired abstract sound, and break it apart into meaningful causal components. In order to help with this, analysis to find classes of source and filter with invariant properties is key to the success of this system.

Once terms are decided, then testing is required. This will involve handing out a printed example to experienced sound designers (who are not to hear the original intended timbre) and the results will be played to a group of listeners who will score various criteria based on their perception. The results from the perception test can then be processed using mathematics to assess where improvements can be made.

References

- Chion, M. (1994). *Audio-Vision*. (C. Gorbman, Trans.) New York: Columbia University Press.
- Chion, M. (1983). *Guide Des Objets Sonores*. Paris: INA/GRM.
- Güven, E., & Özbayoglu, A. M. (2012). Note and Timbre Classification by Local Features of Spectrogram. *Procedia Computer Science*, 12, 182-187.
- Handel, S. (1995). Timbre Perception and Auditory Object Identification. In B. C. J. Moore, *Hearing - Handbook of Perception and Cognition* (2nd ed., pp. 425-461). London: Academic Press.
- Hartmann, W. M. (1995). The Physical Description of Signals. In B. C. J. Moore, *Hearing - Handbook of Perception and Cognition* (pp. 1-40). London.
- Kane, B. (2014). *Sound Unseen: Acousmatic Sound in Theory and Practice*. Oxford: Oxford University Press.
- Loureiro, M. A., de Paula, H. B., & Yehia, H. C. (2004). Timbre Classification of a Single Musical Instrument. Universitat Pompeu Fabra.
- Rasch, R. A., & Plomp, R. (1982). The Perception of Musical Tones. In D. Deutsch, *Psychology of Music* (pp. 1-24). New York: Academic Press.
- Ricard, J., & Herrera, P. (2004). Morphological sound description: computational model and usability evaluation. *Convention Paper 6012* (pp. 1 - 9). Berlin: Audio Engineering Society.
- Risset, J.-C., & Wessel, D. L. (1982). Exploration of Timbre by Analysis and Synthesis. In D. Deutsch, *Psychology of Music* (pp. 25-58). New York: Academic Press.
- Sandell, G. J., & Martens, W. L. (1995). Perceptual Evaluation of Principal-Component-Based Synthesis of Musical Timbres. *Journal of the Audio Engineering Society*, 43 (12), 1013-1028.
- Schaeffer, P. (2017). *Treatise on Musical Objects*. (C. North, & J. Dack, Trans.) Oakland: University of California Press.
- Serra, X., & Smith, J. (1990). Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition. *Computer Music Journal*, 14 (4), 12-24.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.
- Watkinson, J. (2002). *An Introduction To Digital Audio* (2nd ed.). Oxford: Focal Press.