2018

Some approaches to representing sound with colour and shape

Lindsay Vickery
Edith Cowan University, l.vickery@ecu.edu.au

Vickery, L. (2018). Some approaches to representing sound with colour and shape. In *Proceedings of the 4th International Conference on Technologies for Music Notation and Representation* (pp. 165-173). Montreal, Canada: Concordia University. Available here

SOME APPROACHES TO REPRESENTING SOUND WITH COLOUR AND SHAPE

Lindsay Vickery
Edith Cowan University
l.vickery@ecu.edu.au

ABSTRACT

In recent times much of the practice of musical notation and representation has begun a gradual migration away from the monochrome standard that existed since the emergence of printed Non-Western music in the 16th century, towards the full colour pallet afforded by modern printers and computer screens. This move has expanded the possibilities available for the representation of information in the musical score. Such an expansion is arguably necessitated by the growth of new musical techniques favouring musical phenomena that were previously poorly captured by traditional Western musical notation. As time-critical form of visualisation there is a strong imperative for the musical score to employ symbols that signify sonic events and the method of their execution with maximal efficiency. One important goal in such efficiency is "semantic soundness": the degree to which graphical representations makes inherent sense to the reader. This paper explores the implications of recent research into cross-modal colour-to-sound and shape-to sound mappings for the application of colour and shape in musical scores. The paper also revisits Simon Emmerson's Super-Score concept as a means to accommodate multiple synchronised forms of sonic representation (the spectrogram and spectral descriptors for example) together with alternative notational approaches (gestural, action-based and graphical for example) in a single digital document.

1. INTRODUCTION

Visual representation of the multi-parametrical nature of both sound and musical notation is an enduring "wicked problem". It is also a time critical problem, complicated by the differences between the human auditory and visual systems and even mental chronometry. Since the end of the common practice period the timbral pallet employed by composers and performers has greatly expanded, and particularly since the advent of digital computing, the range and detail of the spectral description of sound has exponentially increased.

Because of the technological limitations of typesetting Common Practice Notation (CPN) developed almost entirely in monochrome and with a vocabulary of fixed symbols. Since the advent of colour printing and colour screen-based scores there is no reason for these constraints to continue. This paper explores the value and potentials of employing colour and shape to accommodate the multiparametrical description of sound and notation, in particular through the utilisation of "Cross-Modal Correspondences" between auditory and visual perception as a means for developing semantically sound strategies and methods for representing sonic phenomena and notation.

In the context of the expanding range of forms of representation and notation, Emmerson's notion of the Super-Score, a digital document bundling media relevant to a soundwork/composition together, as a means for composers, performers and researchers to synchronously document and explore sonic works.

2. CROSS-MODAL CORRESPONDENCE

The perceptual phenomenon now generally referred to as Cross-Modal Correspondence (CMC), defined by Marks as "natural correspondences between experiences in different sense modalities" [1], provides some prospects for strategies that might improve the semantic soundness of music notation.

CMC is roughly analogous to the better known rare and idiosyncratic condition Synaesthesia, which causes individuals to experience sensory input cross-modally, the most common form being the simultaneous activation of the senses colour and sound. CMC along with synaethesia has been the subject of scientific enquiry for over two hundred years [2]. In the late 1960s Luria referred to CMC as the 'remnants' of synaesthesia "that many ordinary people have, which are of a very rudimentary sort (experiencing lower and higher tones as having different colorations)" [3]. Later research has tended to separate the two phenomena [4]: the relatively rare condition of synaesthesia (occurring in 0.5% of the population [5]) characterized as absolute, unidirectional, intransitive, rigid and CMCs, some of which are universal [4] as relative, bidirectional, transitive and malleable.

- CMCs are relative: they tend to be ordered in continua – i.e. low to high, soft to loud, dense to diffuse etc.;

- Synaesthesia is unidirectional: the sensory correspondences are not invertible – a pitch may elicit a particular colour but a that colour may not elicit the same pitch;
- CMCs are transitive: "the same core correspondences should emerge whichever sensory feature is used to probe them, confirming that the en bloc alignment of the dimensions is context invariant" [6];
- CMCs are malleable, they may be used as a component of a semantic languages that are able to be learned;

Spence & Deroy [7] have proposed that there are multiple forms of crossmodal mapping: statistical, structural, and semantic. They define statistical, the most hard-wired of the three, as occurring due to the similarities of the transformation of sensory information into perceptual information; structural, more learned and environmental resulting from functional regularities that can be commonly observed in the physical environment; and semantic correspondence, the most conscious and trained occurring when two objects are linked conceptually [8].

It is notable that these categorisations align with aspects of research in other fields for example Moody's *Physics of Notations Theory* [9] (i.e. cognitive fit, semiotic clarity, visual expressiveness, semantic transparency), MacEachren's *Expanded Graphic Vocabulary* [10] (i.e. location, size, resolution, transparency, colour, texture, orientation, arrangement, shape), Wierzbicka's *Semantic Primes* [11] (quantifiers, evaluators, descriptors, actions, events, movement, contact, time, space, intensifiers) and Patel et al.'s [12] discussion of *Augmentative and Alternative Communication Symbols* (gestalt, semantic attributes, cartoon conventions, compositional distinctions, line interpretation).

Interest in CMCs may have been initially sparked by the Köhler's bouba/kiki experiment [13]. In this experiment, "because of the sharp inflection of the visual shape, subjects tend to map the name kiki onto the (pointed, star-like) figure (…), while the rounded contours of the (other) figure make it more like the rounded auditory inflection of bouba" [14]. The cross-modal mapping tendency suggested that there are "natural constraints on the ways in which sounds are mapped on to objects" [4], for human perception in general and beyond the atypical perception of synaesthetes.

The social sciences appear to have contributed the notion that mental concepts could be arranged cross-modally in oppositional continua. In 1954 Guttman [15] proposed a circular psychometric structure called the circumplex for spatially and hierarchically situating emotions, it was applied to personality by Leary [16] and Block [17] and added the further cross-modal dimension of colour to the circumplex.

> *We can compare the two dimensions of the facial-expression surface to the blue-yellow and red-green axes of the color surface. This immediately suggests that there may be a third dimension, corresponding to visual brightness. The third dimension for facial expressions might well be the intensitive one we considered earlier, level of activation.* [18]

This correlation, and other similar associations, for example between shapes and sounds, facial expressions and colours [19] and colour and a range of musical phenomena including timbre, pitch, tempo, intervals, triads and musical genres in non-synaesthetes [20, 21], continue to be explored extensively and cross-culturally [22, 23, 24] providing insight into potentially more natural means to visually represent sonic phenomena.

3. COLOUR

There are a number of perceptually based restrictions upon the use of colour to represent sound. There are no clear direct perceptual analogies between human visual and auditory processes. The ear senses sound continuously with a resolution up to 15-20kHz, while in reading visual field is sensed in grabs of detailed data through focused fixations of about $4cm^2$ for a minimum duration of approximately 5kHz or 200ms. The eye is much 'slower' than the ear. This is a crucial issue for the depiction of sound visually, as eyes are only capable of sensing visual detail of many orders lower than the ear senses sound - perhaps as much as 400-600 times lower.

The wide frequency sensitivity of the ear (in the order of ten octaves) also contrasts the single "visual octave" of the eye: colours in the visual spectrum do not repeat - ultraviolet and infrared are both invisible. In addition, although the eye can finely discriminate variations in colour, Green-Armytage suggests that, in terms of representing data with colour, a limit of 27 tones is "the largest number of different colours that can be used before colour coding breaks down" [25]. In contrast the ear can discriminate pitch differences as small as five cents [26].[1]

Although there have been numerous colour to sound mappings proposed over the centuries, the investigation of CMC between colour and sound appears to date from Schlosberg [18] and later to Plutchik who claimed that:

> *the primary emotions can be conceptualized in a fashion analogous to a color wheel-placing similar emotions close together and opposites 180 degrees apart, like complementary colors. Other emotions are mixtures of the primary emotions, just as some colors are primary and others made by mixing the primary colors.* [28]

Plutchik's mapping has been influential, underpinning a wide range of musical projects drawing on colour as a metaphor for musical expression [29, 30, 31, 32, 33, 34, 35, 36, 37]. Palmer et al's *Emotional Mediation Hypothesis* [38] expanded this research proposing, "that color and music are linked through shared emotional associations", showing systematic relationships between colour and a range of musical phenomena including timbre, pitch, tempo, intervals, triads and musical genres [20, 21].

[1] A number of the limitations of representing sound and notation are discussed in detail in [27].

Their investigation of instrument timbre interestingly showed that the average yellow-blue value was "correlated with timbre attack time whereas average red-green value is correlated spectral brightness [21]. Close inspection of their data shows that despite this correlation, the colours chosen by their participants were extremely varied: although the correlations to place the fitness of colours according to timbre attack and spectral brightness appeared to be strong, but did not point to *specific* colours as being more appropriate. This is crucial as it demonstrates that CMCs are relative, malleable and at least partially "explainable by exposure" [4] to environment and/or learning.

Application of cross-modal principals to colour is also problematic because of the difficulty of establishing a meaningful mapping of bright and dark colours. Whereas sound is mapped in a broadly linear fashion with the cochlea capturing frequencies continuously from high to low, the eye combines data from a range of different sensors – colour through three cone cells and luminosity through rod cells. The result is that vision is not mapped in a linear fashion: if it were, the light spectrum would appear as a bright to dark continuum from purple - the highest frequency colour to red the lowest frequency colour. The arrangement of rods and cones gives rise to anomalies such as the non-sequential perceptual "brightness" of colours such as yellow, cyan and magenta in the colour spectrum.

	H	S	L
Pale Goldenrod	59	30	94
Beau Blue	191	18	91
Sandstorm	53	64	90
Tea Green	108	11	90
Baby Pink	11	15	89
Thistle	291	9	83
Persian Orange	22	57	80
Ruddy Brown	21	74	78
Mulberry	328	53	74
Firebrick	2	76	71
Cool Grey	225	32	71
Dark Sea Green	155	30	70
Dollar Bill	79	59	69
Brass	54	60	62
UCLA Blue	222	53	62
Camouflage Green	115	40	61
Raw Umber	27	67	57
Dark Slate Blue	226	62	55
Cordovan	9	46	53
Viridian	176	41	52
Dark Slate Blue	274	61	48
Dim Grey	48	10	41
Dark Slate Grey	163	43	33
Caput Mortuum	341	68	32
Rifle Green	101	45	30
Smoky Black	0	0	4

Figure 1. Green-Armytage's 27 discriminable colour tones (adapted by the author in descending order of approximate perceptual brightness).

CIELAB colour space [39] attempts to mimic the non-linear response of the eye by modeling cone responses. Mapping the 27 colour tones suggested by Green-Armytage (Figure 1) to CIELAB colour space gives an approximate continuum of hues from brightest to darkest, together with a notional maximal number of discriminable hues. It is also possible to group the hues according to their proximity to spectral colours – reds, oranges, yellows etc. – in order to depict related sonic phenomena or notations: instrumental timbre variation, gesture or stick/mallet designation for example.

A further issue is the multi-parametrical nature of the representation of sound and notation. This prohibits the development of a standard method for applying colour in musical representation: it will always be necessary for the composer to make choices about the phenomena that is represented by any colour, colour continuum or colour attribute (for example mapping hue to brightness, noisiness to saturation and spectral skew to luminance [40]). What is clear is that the use of colour in representation and notation provides powerful tool for the formation of what Moody terms "Perceptual Discriminability" [9] maximising the distinctness of separate phenomena in the manner routinely employed in data visualisation [41], transport maps [25], and websites [42].

4. SHAPE

The simplest and perhaps least contested of these Cross-modal mappings is the vertical spatial depiction of frequency, in which higher frequencies are also spatially represented vertically higher on the page. This visual pitch metaphor that "while culturally diverse, may be based upon basic underlying mappings, stemming from bodily-based intermodal interactions with the physical environment" [22] and has been demonstrated pan-culturally [22] and in infants as young as 1 year old [43].

Vertically proportionality is one area in which CPN is in conflict with CMC: instruments higher on the score are not necessarily higher in pitch and a note may occupy the same vertical location if whether it is sharp, natural or flat. The same space between stave lines may represent a minor or a major third. This is a significant problem to overcome as musicians trained in this tradition can both read "music" and mentally sonify it (as opposed to visualise), but this process is only afforded by a significant range of implicit literacy skills.

The practice of music notation developed in the context of tonal/modal music many attempts have been made to "reform " this deficiency, to allow for efficient representation of the chromatic and smaller grained pitch grids/scales [44].

This issue is nowhere more apparent than in the representation of electronic sounds and field recordings. Robert Erickson's *Pacific Sirens* (1968), one of the first works to use a spectrogram-score as a means for performative engagement the timbral complexity of a field recording with instruments, is an example of the conflict. The work employed proportional notation, with an external time source (stopwatches), together with a spectrographic transcription of sound/frequency morphologies and directed the "improvising" performer's to "listen into" "the spectral complexes of the environmental noise and appropriately blend and protrude" [45]. The frequencies of the field recording are overlaid against traditional treble and bass staves and although spatially time is represented proportionally, pitch is not and therefore sliding pitches are symbolically, rather than accurately portrayed in the score (Figure 2).

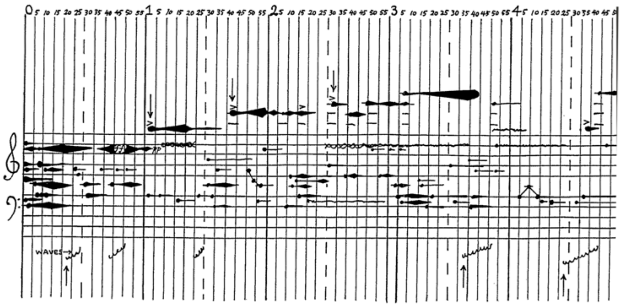

Figure 2. Excerpt of spectrographic score of Robert Erickson's *Pacific Sirens* (1968) © Smith Publications 1969.

A potential solution to this problem is a vertically proportional stave[2], however such novel systems are often unpopular with performers as indicated by the number of abandoned proposals littering the last two hundred years of music history [44].

CPN is of course not horizontally proportional either (spacing is principally determined by typographical compactness rather than the duration of note events). However horizontally "proportional notation" or "Time Notation" pioneered by Earle Brown in the 1950s [46] has been quite widely adopted by composers as diverse as Cage, Berio and Grisey. In propotional notation, the "spacing and length of the notes on the page, are put into a more or less direct relation to the timing and duration of the sounds [47].

The spatial/durational relationship in works by the New York School composers tended to be "simply observed" [48], as Cage noted, rather than executed in the context of a temporal grid (in contrast to Grisey's *Partiels* (1975) or scrolling/swiping digital notation [49] for example). This issue of precisely representing a sense of metricality, complex rhythmic structures and coordination of multiple performers in proportional notation is a significant issue that is only partially resolved by non-visual means such as a conductor, external click tracks [50] or animated notation [51].

Using the inherent semantic qualities of graphical shapes to denote sonic morphology also has valuable potential. The development of "a new graphical vocabulary based on spectromorphology" [52], the visualization of sonic phenomena, has been most fully explored in the field of spectromorphology in acousmatic music [53, 54, 55].

Blackburn refers to the cross-modal quality of acousmatic music listening: "it is frequently reported that, in concert, acousmatic music has the powerful effect of conjuring imagery, shapes, trajectories and spaces, which we as listeners proceed to describe verbally [52]. Blackburn's graphical vocabulary not only visualizes individual "sound units" but also shows how they can be "strung together to form longer phrase lengths" or "morphological strings" [54]. She emphasizes the use of perceptual metaphors, stating that words that are "more readily visualized ie. *spiral, flock, stream* and those with a clear associated physicality ie. *fly*, *drift*, *attack*, appear better suited for informing sound material creation" [55].

Spectromorphological representations, in two dimensions, share the same space as pitch/duration representations, and there are potential conflicts in the signification of other sonic parameters in this space, perhaps particularly dynamics/intensity, which are often depicted by increased size. As the size of a shape increases it also occupies the vertical continuum allotted to pitch, which is problematic when a sonic object is varying both in pitch and dynamic. Solutions to this problem might include three-dimensional representation or indicating pitch (or spectral centroid) with a line of consistent size and contrasting colour to other parameters.

The process of eye fixation (with a "gaze frame" of roughly 4cm^2 for periods of time in the order of 200–400ms) is very slow in relation to both the auditory system and the mental chronometry that allows for the execution of physical actions. The author's previous research has suggested reading becomes difficult beyond a rate of approximately 3cm/s [36, 56]. Although musicians are capable of performing nuances at extremely minute durations, the eye is not capable of capturing data quickly enough.

Music is however not always performed in a sight-reading context, and perhaps the preponderance of Western Art Music presupposes rehearsal and practice. One solution available to for screen-based representation is multiscale representation analogous to digital maps, permitting magnification of the score/representation, while maintaining a constant graphical density. In cartography, for example, Bertin [57] suggests no more than 10 semantically meaningful units should be represented per cm^2. This is a feature of Digital Audio Workstations and Spectrographic software, but could also usefully accomodate notation that is either too fast or too detailed to read, providing a less detailed representation at lower resolutions.

5. HEURISTICS

Although CMC is malleable and relative it is possible, in conjunction with the perceptual restrictions discussed previously, to develop some heuristics or "rules of thumb" that might guide their implementation in sonic representation. Statistical, structural and semantic correspondences are somewhat fluid due to environment and training, however statistical CMC appears to be the most difficult to unlearn[3]. Tsiros' compilation of crossmodal sight-to-sound research [8], shows that the strongest correspondences (in order of strength) are spatial height to pitch association (although size to pitch is also significant), amplitude to light intensity, duration to horizontal length and texture granularity to timbre. (By extension shape is therefore

[2] Discussed in [36].

[3] For example cellists can "unlearn" the correspondence between rising pitch and lower spatial position on the fingerboard. Apparently jazz pianist Joe Zawinul, however used the technique of inverting the pitch of his keyboard in order to break his familiarity with its spatial layout [58].

associated with amplitude related sonic morphology). These correspondences should therefore be the most crucial to consider in the creation of a semantically sound notation.

More overtly semantic correspondences such as symbols, pictograms and text are more flexibly applicable for specific representation requirements. They are capable of being learned and indeed many are already embedded in CPN in varying degrees.

It is important to remember that although the acquisition rate of the eye for linear information is potentially as high as 20cm/s, practical examples indicate rates beyond 4cm/s become uncomfortable for the performer to read in an accurate and synchronous manner [59]. Therefore, detailed depictions of sonic events for performative reading or as a representation of audio presented in real-time, are restricted to approximately 4cm per second of sound.

Colour parameters such as hue, saturation and luminance may be mapped to spectral descriptors such as brightness, noisiness and roughness for example, to produce visual representations[4]. As mentioned colour discrimination for the purpose of identifying distinct phenomena is restricted to approximately 27 hues. Figure 3 shows the pallet of 23 hues used to depict separate instrumental parts for the chamber orchestra and fixed media work *bascule.* The colours are segmented into groups by instrument family (yellows for flutes, blues for strings for example) as well as subset variations of those hues to depict individual instruments within a family (Firebrick red for clarinet and Caput Mortuum for bass clarinet for example). This arrangement exploits the CMC between higher pitch and brighter hue.

piccolo	alto sax	e. guitar 3	violin 1
flute 1	tenor sax	e. guitar 2	violin 2
flute 2	baritone sax	e. guitar 3	viola
clarinet 1	trumpet	accordion	cello
clarinet 2	horn	cymbal	double bass
bass clarinet	trombone	bass drum	

Figure 3. Colour-to-Instrument coding employed in Vickery's *bascule* (2016).

Figure 4 shows an excerpt from the score *wellington forest* (2017) in which a spectrographic representation of an accompanying field recording is combined with temporally proportional traditional and rhythmic notation. The spectrogram was produced by "threshing" the field recording to remove all but the highest amplitude sounds (in this case frog croaks) and serves as a guide to the filed recording for the performers. Pitches are indicated via "cut out" staves or a five-line stave. Beams are used to indicate emergent phrase structures and stems are placed on the left side of each square notehead to aide coordination of the performers with the scrolling score and embedded soundfile. The score scrolls as a rate of 6.19mm/s.

In the excerpt from *[opi'lka* (2017) for septet and fixed media (Figure 5), a number of conventions are employed simultaneously to convey a variety of performance practices. Like *wellington forest* performers are provided with cut out pitch indications. The flute reads their material entirely from yellow colour-coded spectrographic representations, inferring the pitch content from noteheads on the cut out stave and emulating the morphology of the spectrographic shapes. The bass clarinet (colour-coded purple) reads from symbolic notation indicating whitenoise-like "breath jets" as well as spectrographic representations of those sounds. The electric guitar, playing with a slide, has a more gestural form of notation in which the shape indicates the contour of the slide movements. The score scrolls as a rate of 16mm/s.

Like *wellington forest* the 4 instruments in *kuroinami* (2016) are provided with a spectrographic representation of the fixed media part (Figure 6). Cut away staves are only used when the instruments are playing non-open strings. The double bass part primarily uses a semantically symbolic language to depict combined sul ponticello bowing and left hand pizzicato. The biwa, viola and cello parts mostly employ a gestural notation similar to that used in Lachenmann's *Pression* (1968) with specific instructions written in English. The score scrolls as a rate of 11.73mm/s.

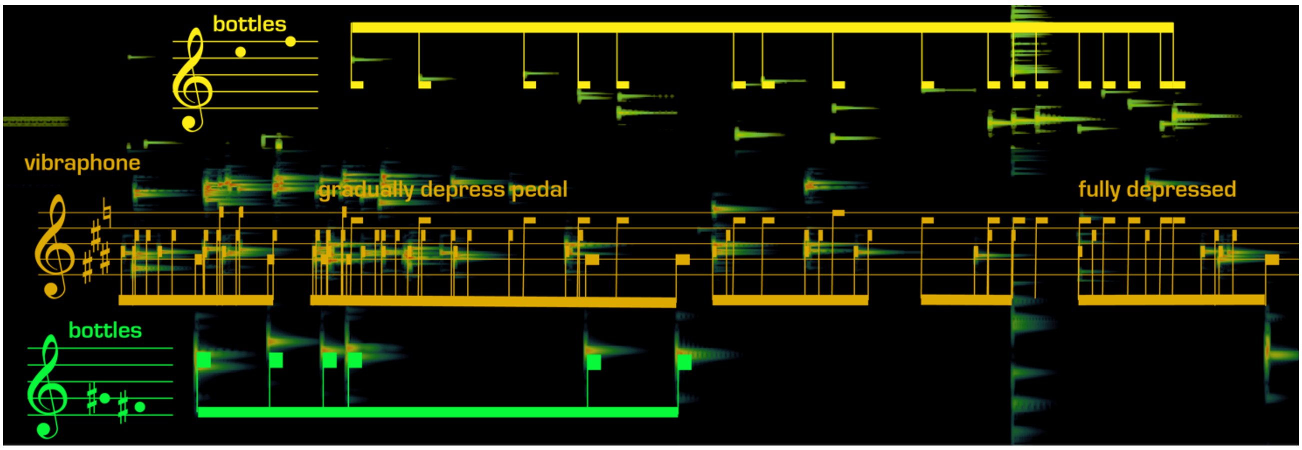

Figure 4. Excerpt from Vickery's *wellington forest* (2017) for percussion trio and field recording.

[4] A number of these issues are discussed in [40].

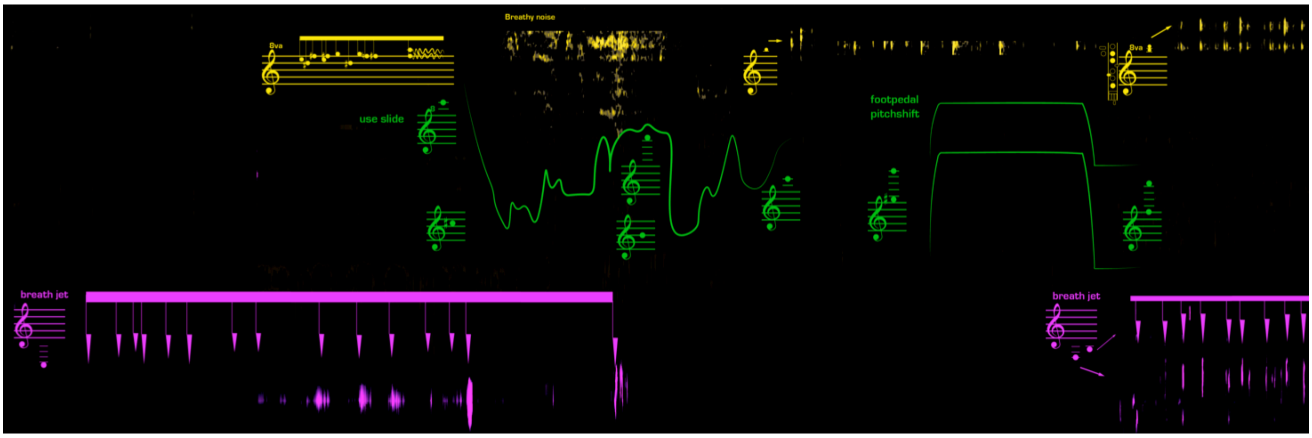

Figure 5. Detail showing excerpt of flute (yellow), bass clarinet (purple) and electric guitar (green) parts from Vickery's *[opi'lka* (2017) for flute, trumpet, soprano saxophone, alto saxophone, bass clarinet, electric guitar, prepared piano and fixed media.

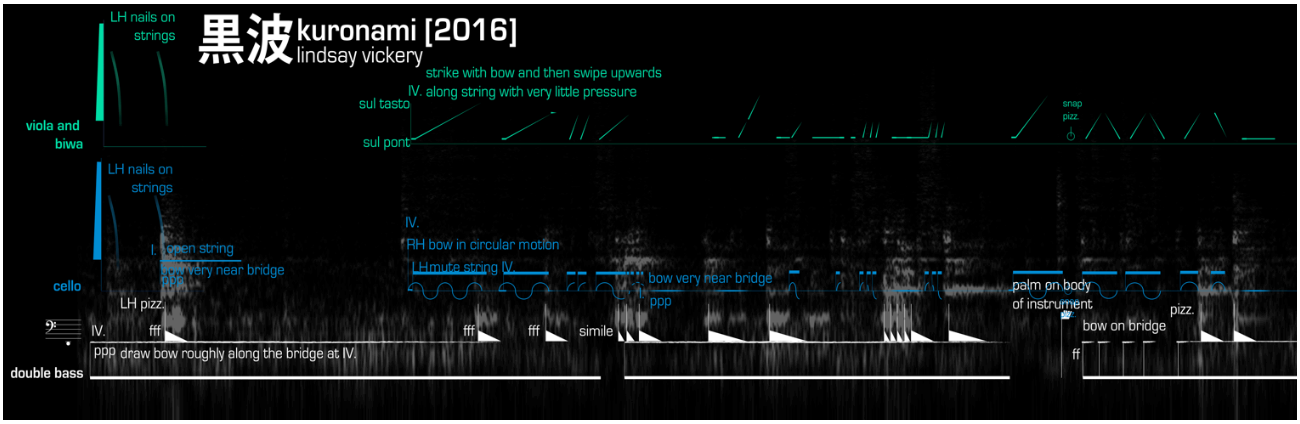

Figure 6. Excerpt from Vickery's *kuroinami* (2016) for biwa, viola, cello, double bass and fixed media.

These three scores combine spectrographic representations of sound with a different form of notation. The application of diverse notational forms highlights the fact that in each case a choice of notation most appropriate to a particular purpose was made to the exclusion of other forms. Each notational approach, CPN, proportional, tablature, gestural graphical and so on, favours different aspects of the performative requirements. Furthermore the prominence of the spectrographic representation is fixed and cannot be intensified or diminished. These developments point toward the possibility of a more multidimensional score, in which the performer may choose between and blend notational and representational approaches.

6. SIMON EMMERSON'S SUPER-SCORE REVISTITED

No single approach to musical representation can accommodate every existing notational and representational requirement. In addition to approaches that are well established in Europe such as CPN, tablature, graphic notation and the spectrogram, we might add those from Non-Western music [60, 61], Jazz [62], Popular Music [63] and European Early Music [64] and emerging approaches such a gestural [65] and "action-based" [66] notations. Related, but the other end of the resolution continuum is music analysis which often involves the schematization and compression of musical structures into meaningful components.

In 2000 Emmerson proposed that "the super-score of the future" could be a multimedia object bringing together all the necessary materials to define a sonic work [67]. His concept incorporated traditional notation, extended notation, audio, video, software and documentation. Emmerson's concept would allow for an all-encompassing digital document that would accommodate multiple synchronised forms of sonic representation, that could be viewed in multiple modes (in the way a digital map can be viewed in satellite or terrain mode) allowing the sonic phenomena and/or notations to be easy alternated.

Such a document might allow for:

- the synchronised alternation between and/or superimposition of, multiple forms of musical representation;
- linked supporting annotative media;
- multiscale representation [68] of image files (zooming);
- communication and synchronisation with digital audio and analysis tools.

- There are great advantages to "bundling" the performance or realisation materials into a single unit [67, 50].

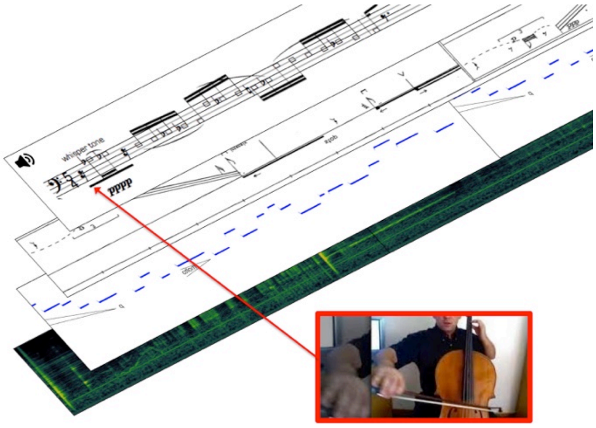

Figure 7. Depiction of the arrangement of multiple forms of notational visualization in an Integrated Score File Format.

The integrated score file format (Figure 7) would obviate the need to collect all required specifications within a single text and stream-lines inclusion of alternative forms of score and annotations by collecting them in an aligned format allowing the reader, performer, researcher to swipe between representations. The score could also be annotated with embedded text, audio and video and additional resources, such as necessary audiofiles, software, technical papers and so on.

7. CONCLUSION

The efforts to extend notation discussed here are part of an ongoing effort to better capture nuances of sound such as timbre, temperament and envelope morphology using shape and colour parameters in a manner that is concise and semantically sound. Although CMC does not provide a "magic bullet" solution, the current state of research does give helpful guidelines in regard to the appropriateness of deploying colour and shape in the service of sonic parameters.

Although the malleable quality of CMCs suggest that any system of associations can be learned it seems likely that the spatial metaphor of pitch and duration is particularly strong, and that the pre-existing (at least in English) cross-modal metaphor timbre/colour suggests the useful retention of that association. The human visual system's non-linear response to the light spectrum may by potential exploited in the service of representing multiple parameters. The contest for vertical space between pitch and dynamics is a persistent issue that will most likely elicit multiple idiosyncratic solutions.

The use of colour and shape to represent the mutliparametrical musical space embraces advances in printing and presentation technologies that will likely continue to improve.

In the context of the multiple means for representing sound and musical works, it is proposed that Emmerson's notion of the Super-Score, a digital format accommodating text, graphics, sound, video and algorithmic resources is reconsidered as a goal.

8. REFERENCES

[1] L. E. Marks, "On Perceptual Metaphors", *Metaphor and Symbolic Activity*, vol. 11, no. 1, pp. 39–66, 1996.

[2] C. van Campen, "Artistic and Psychological Experiments with Synesthesia", *Leonardo*, vol. 32, no. 1, pp. 9–14, 1999.

[3] A. R. Luria, *The Mind of a Mnemonist*. Harvard University Press, 1968.

[4] O. Deroy and C. Spence, "Why we are not all synesthetes (not even weakly so)", *Psychonomic Bulletin & Review*, vol. 20, no. 4, pp. 643-664, 2013.

[5] J. Harrison and S. Baron-Cohen, "Acquired and inherited forms of cross-modal correspondence", *Neurocase*, vol. 2, no. 3, pp. 245–249, 1996.

[6] P. Walker, "Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright", *Attention Perception and Psychophysics*, vol. 74, no. 8, pp. 1792–1809, 2012.

[7] C. Spence and O. Deroy, "How automatic are crossmodal correspondences?", *Consciousness and cognition*, vol. 22, no. 1, pp. 245-260, 2013.

[8] A. Tsiros, "The parallels between the study of cross-modal correspondence and the design of cross-sensory mappings," in *Proceedings of the conference on Electronic Visualisation and the Arts (EVA'17)*, London, UK, pp. 175-182, 2017.

[9] D. L. Moody, "The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering", *IEEE Transactions on Software Engineering*, vol. 35, no. 6, pp. 756-779, 2009.

[10] A. MacEachren, "Map complexity: comparison and measurement", *American Cartographer*, vol. 9, pp. 31–46, 1982.

[11] A. Wierzbicka, *Semantics: Primes and Universals*. Oxford University Press, 1996.

[12] R. Patel, K. Schooley, and J. Wilner, "Visual Features that Convey Meaning in Graphic Symbols: A Comparison of PCS and Artists' Depictions", *Assistive Technology Outcomes and Benefits*, vol. 41, pp. 62–80, 2007.

[13] W. Köhler, *Gestalt Psychology*. Liveright, 1929.

[14] V. S. Ramachandran, and E. M. Hubbard, “Synaesthesia - A Window into Perception, Thought and Language”, *Journal of Consciousness Studies*, vol. 8, no. 12, 2001, pp. 3–34, 2001.

[15] L. A. Guttman, “A new approach to factor analysis. The radix”, in P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*, Free Press, pp. 238–248, 1954.

[16] T. Leary, *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation*, Ronald Press, 1957.

[17] J. Block, “Studies in the phenomenology of emotions”. *Journal of Abnormal and Social Psychology*, vol. 54, no.3, pp. 358–363, 1957.

[18] H. Schlosberg, “Three dimensions of emotion”, *Psychological Review*, vol. 61, no. 2, pp. 81–88, 1954.

[19] O. da Pos and P. Green-Armytage, “Facial Expressions, Colours and Basic Emotions”, *Colour: Design & Creativity*, vol. 1, no. 2, pp. 1–20, 2007.

[20] W. S. Griscom and S. E. Palmer, “The Color of Musical Sounds in Non-Synesthetes”, in *Annual Meeting of the Vision Science Society*, Naples, FL, USA, 2012.

[21] W. S. Griscom and S. E. Palmer, “Cross-modal Sound-to-Sight Associations with Musical Timbre in Non-Synesthetes”, in *Annual Meeting of the Vision Science Society*, Naples, FL, USA, 2013.

[22] Z. Eitan and R. Timmers, “Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context”, *Cognition*, vol. 114, no. 3, pp. 405–422, 2010.

[23] Y. Kazuhiko, Y. Natsumi, K. B. Schloss, L. R. Prado-León, and S. E. Palmer, “Cross-Cultural Studies of Color Preferences: US, Japan, and Mexico”. *Journal of Vision*, vol. 10, no. 7, 2010.

[24] L. R. Prado-Leon, K. B. Schloss, and S. E. Palmer, “Bach to blues, Color, music and emotion across cultures”, in W. Anderson, C. P. Biggam, C. Hough, and C. Kay (Eds.), *Colour Studies: A Broad Spectrum*, John Benjamins, pp. 366-379, 2014.

[25] P. Green-Armytage, “A Colour Alphabet and the Limits of Colour Coding”, *Journal of the International Colour Association*, vol. 5, 2010.

[26] L. K. Miller and H. Clausen, “Pitch identification in children and adults: Naming and discrimination”, *Psychology of Music*, vol. 25, no. 1, pp. 4-17, 1997.

[27] L. Vickery, “The Limitations of Representing Sound and Notation on Screen,” *Organised Sound*, vol. 19 no. 3, pp. 215-227, 2014.

[28] R. Plutchik, “The Nature of Emotions”, *American Scientist*, vol. 89, no. 4, pp. 344-350, 2001.

[29] I. Whalley, “Broadening Telematic Electroacoustic Music by Affective Rendering and Embodied Real-time Data Sonification,” *Proceedings of the International Computer Music Conference (ICMC)*, Athens, Greece, pp. 301-307, 2014.

[30] G. S. Vercoe, “Moodtrack: practical methods for assembling emotion-driven music,” PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

[31] W. Jauk, “Interfaces–a Musical Situation”, *Musicological Annual*, vol. 43, no. 1, pp. 43-51, 2007.

[32] S. R. Livingstone, R. R. Muhlberger, A. F. Brown, and W. Thompson, “Changing musical emotion through score and performance with a computational rule system”, *Computer Music Journal*, vol. 34, no. 1, pp. 41-65, 2008.

[33] S. Baldan, A. Barate, and L. A. Ludovico, “Automatic Performance of Black and White n. 2: The Influence of Emotions Over Aleatoric Music”, in *Proceedings of international symposium on Computer Music Modeling and Retrieval (CMMR)*, London, pp. 437-448, 2012.

[34] I. R. O’Keeffe, “Soundtrack Localisation: Culturally Adaptive Music Content for Computer Games”, *The Journal of Internationalization and Localization*, vol. 2, no. 1, pp. 36-65, 2012.

[35] B. E. Carey, *Notation Sequence Generation and Sound Synthesis in Interactive Spectral Music*, Master's thesis, University of Sydney, Australia, 2013.

[36] L. Vickery, “Exploring a Visual/Sonic Representational Continuum,” in *Proceedings of the International Computer Music Conference (ICMC)*, Athens, Greece, pp. 177-184, 2014.

[37] S. Canazza, G. D. Poli and A. Rodà, “CaRo 2.0: An Interactive System for Expressive Music Rendering”, *Advances in Human-Computer Interaction*, 2015.

[38] S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, “Music–color associations are mediated by emotion”, in *Proceedings of the National Academy of Sciences of the USA*, vol. 110, no. 22, pp. 8836-8841, 2013.

[39] G. Hoffman, CIELab Color Space. [Online]. Available: http://docs-hoffmann.de/cielab03022003.pdf

[40] L. Vickery, “Visualising the Sonic Environment”, in *Proceedings of the Australasian Computer Music Conference (ACMC’17)*, Adelaide, Australia, pp. 43-50, 2017.

[41] E. Tufte, *Envisioning Information*, Graphics Press, 1990.

[42] P. Stanicek, Color Scheme Designer, 2009. [Online.] Available: http://colorschemedesigner.com

[43] S. Wagner, E. Winner, D. Cicchetti, and H. Gardner, ""Metaphorical" mapping in human infants", *Child Development*, vol. 52, no. 2, pp. 728–731, 1981.

[44] G. Read, *Source Book of Proposed Music Notation Reforms*. Greenwood Publishing Group, 1987.

[45] J. Mackay, "On Community, Institutions and Politics in the Life and Work of Robert Erickson," in J. Wall (Ed.), *Music, Metamorphosis and Capitalism: Self, Poetics and Politics*, Cambridge Scholars Publishing, pp. 18-32, 2007.

[46] E. Brown, "The Notation and Performance of New Music", *The Musical Quarterly*, vol. 72, no. 2, pp. 180-201, 1986.

[47] C. Cardew, "Notation: Interpretation, etc", *Tempo*, vol. 58, pp. 21-33, 1961.

[48] P. D. Miller, "Indeterminacy and performance practice in Cage's Variations", *American Music*, vol. 27, no. 1, pp. 60-86, 2009.

[49] L. Vickery, "The Evolution of Notational Innovations from the Mobile Score to the Screen Score", *Organised Sound*, vol. 17 no. 2, pp. 128-136 2012.

[50] L. Vickery, "Mobile Scores and Click-Tracks: Teaching Old Dogs", in *Proceedings of the Australasian Computer Music Conference (ACMC'10)*, Canberra, Australia, pp. 63-70, 2010.

[51] R. R. Smith, "An atomic approach to animated music notation," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR'15)*, Paris, France, pp. 39-47, 2015.

[52] M. Blackburn, "The Visual Sound-Shapes of Spectromorphology: An Illustrative Guide to Composition", *Organised Sound*, vol. 16, no. 1, pp. 5–13, 2011.

[53] A. Pasoulas, "Temporal Associations, Semantic Content and Source Bonding", *Organised Sound*, vol. 16, no. 1, pp. 63–68, 2011.

[54] D. Tanzi, "Extra-Musical Meanings and Spectromorphology", *Organised Sound*, vol. 16, no. 1, pp. 36–41, 2011.

[55] M. Blackburn, "Composing from Spectro-morphological Vocabulary: Proposed Application, Pedagogy and Metadata", in *Electroacoustic Music Studies Conference 09 (EMS09)*, Buenos Aires, Argentina, 2009.

[56] L., Vickery and T. Goh, "Music Screen-Reading: indicative results from two pilot studies", *Proceedings of the Australasian Computer Music Conference (ACMC)*, pp. 119-125, 2015.

[57] J. Bertin, *Semiologie graphique : Les diagrammes, les réseaux, les cartes*. Edition Gauthier-Villars, 1967.

[58] M. Jenkins, "Using and programming analog," in *Analog Synthesizers*, Focal Press pp. 142-207, 2009.

[59] L. Vickery, "Notational Semantics in Music Visualisation and Notation," in *Proceedings of the Australasian Computer Music Conference (ACMC'14)*, Melbourne, Australia, pp. 101-112, 2014.

[60] B. Nettl, *The Study of Ethnomusicology: Thirty-Three Discussions*. University of Illinois Press, 2015.

[61] H. Schippers, *Facing the Music: Shaping Music Education from a Global Perspective*. Oxford University Press, 2009.

[62] R. Rusch, K. Salley, and C. Stover, "Capturing the Ineffable: Three Transcriptions of a Jazz Solo by Sonny Rollins". *Music Theory Online*, vol. 22, no. 3, 2016.

[63] D. Temperley, "Syncopation in rock: a perceptual perspective", *Popular Music*, vol. 18, no. 1, pp. 19-40, 1999.

[64] R. Donington, *The Interpretation of Early Music*, Faber & Faber, 1963.

[65] P. Antoniadis and F. Bevilacqua, "Processing of symbolic music notation via multimodal performance data: Brian Ferneyhough 's Lemma-Icon-Epigram for solo piano, phase 1", in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR'16)*, Cambridge, UK, pp. 127-136, 2016.

[66] J. Kojs, "Notating Action-Based Music", *Leonardo Music Journal*, vol. 21, pp. 65-72, 2011.

[67] S. Emmerson, "Crossing cultural boundaries through technology?", in *Music, Electronic Media and Culture*, pp. 115-137, 2000.

[68] G. W. Furnas and B. B. Bederson, "Space-Scale Diagrams: Understanding Multiscale Interfaces," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, pp. 234-241, 1995.