An Approach to the Generation of Real-Time Notation via Audio Analysis: *The Semantics of Redaction*

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

**ABSTRACT**

This paper discusses an approach to mechanically and conceptually to generation notation in real-time in MaxMSP in the work *The Semantics of Redaction*. The work uses analysis data to generate scrolling, but conventional looking notation. The issues discussed include scaling and mapping of audio data to visual representation, the formal structure of the work and its relation to the generative processes, semantic principles guiding the notational concept and the development of a performance practice for the work. The intention of the work is to provide a flexible vehicle for a performer to explore audio from the media-world in a near-real-time manner using a recent, topical speech recording as the generative artifact for each performance.

1. **INTRODUCTION**

*The Semantics of Redaction* [2014] (SoR) is a work for solo performer, pre-recorded audio and generative score. To perform the work, a recent, topical speech recording is chosen and loaded. The intent here is that the performer chooses a new recording for each performance emphasizing the “near real-time”, “daily news”-like aspect of the process. The audio is played through the work’s analysis patch and the analysis parameters (frequency, amplitude, brightness, noisiness and attack) are scaled to provide desirable outputs, in terms of range, density and diversity of the generated notation. Instrumentation is chosen by the performer as a commentary upon the subject matter of the recording. It is intended that the generative score is simple enough to perform after a brief period of familiarisation.

The work was written for percussionist Vanessa Tomlinson and officially premiered in her program 8 Hits [1], an ABC Classic FM live broadcast on November 1 2014. Tomlinson chose a speech made by climate-change activist Naomi Klein as part of the promotional tour for her book *This Changes Everything* [2014] [2].

Along with its sister work *Lyrebird* [3], SoR began development during Tomlinson’s residency at the Orpheus Institute for Advanced Studies & Research in Music in December 2013.

Tomlinson is both a strong improviser and reader and the work specifically aimed to open a space in which she could interact with a chosen recording with a combination of precision and freedom. Like previous “series” by this composer, *Delicious Ironies* [2001] [4], *Splice* [2002] [5] and *Lyrebird* [2014] SoR is intended to be a structured improvisation environment that takes on different qualities in each performance.

2. **GENERATIVE APPROACH**

The notation for SoR (see Figure 1) is principally created by using accents detected in a speech recording in real-time to generate graphical symbols of varying vertical position, size and colour, determined by the frequency, amplitude and timbre of the recording at the accent point.

![Figure 1](image-url). The opening section of *The Semantics of Redaction* [2014].

Analysis of the recording is derived from Tristan Jehan’s [6] analyzer~ object, using attack, frequency, amplitude, brightness and noisiness data. The flow of analysis data is captured at the point of each detected attack. By default, the values defining an attack are an increase in the incoming signal of +10db in a period less than 100ms, however the performer may alter these values to achieve a better score/visualisation result.

The data is mapped in the following manner:
- **Frequency**: Vertical height and notehead colour hue
- **Amplitude**: Notehead size
- **Brightness**: Notehead colour saturation
- **Noisiness**: Notehead colour luminance

Copyright: © 2015 Lindsay Vickery. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
The data is drawn to MaxMSP’s jit.lcd object using the standard paintrect command. A stem is also drawn on the left side of the notehead above a central “beam”. A left side stem is used regardless of whether the notehead is above or below the beam as an aid to visualising the precise onset of the accent and therefore performer’s note execution.

Because of the variation of the acoustic characteristics of potential recordings and performer preferences, an analysis window allows the performer to scale incoming data (See Figure 2). Scaling changes may be necessary because of variations in the average frequency of the speaking voice or ambient noise in the recording. From the performer’s perspective the attack rate and sensitivity values create changes in the density of note events in the score.

The analysis window provides the performer with “peak”, “trough” and “mean” values for each of the analysis parameters from the recording, as well as a graphical display showing the most recent 100 values for each parameter. A swatch graphically displays the variations in notehead colours. A preset object allows the performer to save the chosen scalings for each new recording.

Although it would be possible to automate the scaling process, allowing the performer to set the values provides a means to individualise the score that is generated by each recording: for example to achieve variations in the number of noteheads that fall within different “registers”.

Unlike Lyrebird, which has a range of colour interpolations and generates a spectrogram-like score, only five hues (Yellow, Orange, Red, Green and Blue) are used in SoR (although there is continuous variation in the saturation and luminance of each colour). This approach was taken to achieve a clear differentiation between five registers in the score, which are conceived as instrument groups by the percussionist: for example ranging from yellow (high-pitched instruments) to blue (low-pitched instruments). The notehead size is intended to correlate to dynamic range.

Unlike Lyrebird, which has a range of colour interpolations and generates a spectrogram-like score, only five hues (Yellow, Orange, Red, Green and Blue) are used in SoR (although there is continuous variation in the saturation and luminance of each colour). This approach was taken to achieve a clear differentiation between five registers in the score, which are conceived as instrument groups by the percussionist: for example ranging from yellow (high-pitched instruments) to blue (low-pitched instruments). The notehead size is intended to correlate to dynamic range.

In the Commentary sections the noteheads are extended indicating the note event should be held proportionally for that duration. The wave-like black lines that widen or narrow, indicate timbral or dynamic variation of the note that remains constant, (See Figure 4a.)

In the Interlude section some noteheads appear without stems indicating non-accented/slurred gracenotes, rather like liquecent neumes in medieval notation, while stems indicate accents. Graphical symbols in the form of black vertical lines are added and are interpreted as note or noise repetitions (See Figure 4b.). During this section the performer plays solo: the recorded speech is not heard by the audience (although it still drives the generation of notation). Audio from the speech is recorded into buffers that are heard during the “redactio” sections.

Unlike Lyrebird, which has a range of colour interpolations and generates a spectrogram-like score, only five hues (Yellow, Orange, Red, Green and Blue) are used in SoR (although there is continuous variation in the saturation and luminance of each colour). This approach was taken to achieve a clear differentiation between five registers in the score, which are conceived as instrument groups by the percussionist: for example ranging from yellow (high-pitched instruments) to blue (low-pitched instruments). The notehead size is intended to correlate to dynamic range.
3. SEMANTIC PRINCIPLES

In any generative, real-time score, “semantic soundness” is an imperative for efficiently communicating notational aims to the performer. The conceit of the traditional beam, stem and notehead is employed as it is presumed to be already familiar to the performer. However, the generative score also differs from a traditional score in a number of ways.

The score scrolls from right to left on a jit.lcd screen with the dimensions 1300x700px. Graphics are drawn on the right side of the screen at the horizontal 1200 px position. The screen is scrolled leftwards at a rate of 100 px/s. Previous research on the semantic issues of screenscores had suggested that this is a middle-range scroll rate \[3, 8, 9\] roughly equivalent to reading a traditional score at a moderate speed. At play here is the trade-off between excessive note density (at a slow scroll-rate) and the limits of the eye’s fixation rate (at a fast scroll-rate) \[10, 11, 12, 13\].

The performer executes events in the notation at the point they reach a line (play-head) on the left-hand side (0 px) of the screen. The graphics take 12s, from the point of drawing, to reach to left-edge of the screen and the audio that generated them is also delayed for 12s. Graphics reach the play-head simultaneously with the audio events that caused their generation. In this manner, the performer has a visualisation of what will occur in the recording 12s in advance of it being heard. The process gives the performer a significant amount of time to view and access the up-coming passages of music, significantly greater than the eye-hand span (the distance between the point of fixation and the point of performance) for music readers suggested by sight-reading studies (between 1.5 and 3 cm) \[14\].

Eye-tracking tests \[15\] were conducted upon a number of participants reading SoR. The results indicate that the majority of performer fixations do indeed fall within the first 3 cm of the score, however there were a significant number of examples of “look-ahead” fixations spanning the entire score (See Figure 5).

![Figure 5. Eye-tracking heat-map of Semantics of Redaction. The colours indicate red as areas of most frequent fixation and green as least frequent fixation.](image)

Although it is possible to draw other notehead shapes, (including more traditional looking rounded noteheads), the rectangle is ideal to indicating a precise onset of the performer’s note execution. As previously mentioned, to aid clarity in the correspondence between the spatial distribution of the note-events and their temporal distribution, note stems always occur on the left hand side of the notehead.

Frequency data is converted to just five HSL Hue values: Yellow (165), Orange (90), Red (0), Green (330), Blue (670). The ordering is suggested by research conducted at the University of California Berkeley, which shows that the brighter-darker hue continuum is strongly correlated with higher-lower frequency continuum \[16, 17\], due to what has been termed Weak Synaesthesia \[18\] or Crossmodal Correspondence \[19\]. Saturation and Luminance (and vertical position) are all continuously scaled with a potential resolution of 700 degrees (the number of vertical pixels).

The additional graphic symbols discussed in the previous section are intended to imply natural, heuristic correspondences both gesturally and sonically, due to “natural constraints on the ways in which sounds are mapped on to objects” \[20\]. In this way the graphic adjacent to Commentary I in Figure 4 should imply to the performer a wave-like gesture/sound increasing in magnitude.

Interestingly, because the brightness and noisiness of accents are represented visually, there are also some observable correspondences between phonemes and the saturation and luminence of noteheads: for example, sibilant sounds tend to be brighter and less saturated in the current system. This may prove to be an interesting phenomena to explore in future work.

The first versions also inverted the function of speech: text was printed onto to the score and intended to be either read by the performer or by the computer using Masayuki Akamatsu’s aka.speech object (See Figure 6).

![Figure 6. An early version of The Semantics of Reduction.](image)
The performance practice of muting the speech and allowing an instrumental interlude is similar to the approach taken by Alvin Lucier in his work Carbon Copies (1989), in which a field recording is used as an “audio-score” for performers that is heard alone, then in addition to the performer(s) who are emulating it, is muted so the audience only hear the performer and then muted even for the performer leaving them to improvise an emulated version of the recording.

The performance is perhaps best situated at the nexus between interpretation and improvisation, in which the performer ranges between precise interpretation of symbols and leaving them to improvise an emulated version of the recording.

4. CONCLUSIONS

The Semantics of Redaction is a flexible vehicle for a performer to explore audio from the media-world in a near-real-time manner. The author has also performed the work on a number of occasions on bass clarinet, reading the colour-coded notation as five instrumental register rather than instrumental families. Multiple performers have also successfully performed the work. It is possible to perform with multiple networked scoreplayers, allowing for interaction with visualisations that focus on varied frequency, amplitude and timbral parameters of the same recording, although this has not yet been attempted.

Acknowledgments

Research on this work and into notational semantics, eyetracking and screenscores was funded by an Early Career Researcher Grant from Edith Cowan University. The Semantics of Redaction was commissioned by Vanessa Tomlinson with funding from the Australia Council for the Arts. Early experiments were conducted at the Orpheus Instituut for Advanced Studies & Research In Music. Very special thanks to percussionist Vanessa Tomlinson for the dialogue that lead to the creation of this work.

5. REFERENCES

[15] The experiments were conducted using the Tobii X2-30 eye-tracker.