

Swarm Tech-tiles

Tim Blackwell¹ and Janis Jefferies²

{¹Department of Computing, ²Department of Visual Arts}
Goldsmiths College,
New Cross,
London SE14 6NW, UK
{¹t.blackwell, ²j.jefferies}@gold.ac.uk,
<http://www.timblackwell.com>

Abstract. This paper describes an exploration of visual and sonic texture. These textures are linked by a swarm of “tech-tiles”, where each tech-tile is a rectangular element of an image or a sequence of audio samples. An entire image can be converted to a single tech-tile, which can be performed as a composition, or a swarm of small tiles can fly over the image, generating a sonic improvisation. In each case, spatial (visual) structure is mapped into temporal (sonic) structure. The construction of a tech-tile from an image file or a sound clip and the swarm/attractor dynamics is explained in some detail. A number of experiments report on the sonic textures derived from various images.

1 Introduction

The word texture (15th Century, from the Latin *texere*, to weave) most commonly refers to the tactile appearance of a surface, especially that of woven fabric. The texture of a textile arises from colored material, woven from yarn that is laid out lengthways (the warp) and width ways (the weft). The resulting fabric has large-scale structure – the design itself – as well as local structure at the scale of the yarn. Upon closer inspection, there is also a micro-texture evident in the twisted fibers that constitute a strand of yarn.

A concept of texture exists in numerous specialized domains. In each case, the concept reflects the common meaning of the term but adds (albeit loosely) domain-specific connotations. For example, in painting, *visual texture* would refer to the representation of the nature of a surface. Visually, large scale structure can emerge from dense, intricate patterning of smaller elements; the works of Jackson Pollock present a good example.

The warp/weft idea exists informally in the idea of *musical texture*. A musical work can be described at a music-theoretic level as the interleaving of “horizontal” individual parts or voices, and “vertical” strands of harmony. An introductory book on composition even illustrates this idea with pictures of textured fabric (silk, polyester yarns and cheesecloth) [7].

In contrast, *sonic texture* refers to descriptions at the sound level [11]. The breakdown of tonality in 20th Century Western art music has led to a heightened awareness of

texture within musical composition. This awareness reaches its zenith in the genres of Sonic Art and Improvised Music [14, 2], where music theoretic descriptions become almost meaningless. Ascension by John Coltrane, a remarkable abstraction within jazz, is a good illustration of sonic and musical texture [6]. In this work, the improvising ensemble (comprising five saxophonists, two trumpeters, two bassists, a pianist and a drummer) create a highly detailed and emotional sound-world. This seems to be a direct analogy of the abstraction pursued by Pollock.

Sound texture in the domain of signal-processing refers to sounds with no perceptible long-term structure [1]. A short sample of a sound texture gives no indication of when it actually occurred within the whole sound, in contrast to musical instruments which produced time-dependent envelopes. Examples of sound textures are running water and traffic noise.

Another meaning of texture is to be found within image processing. The segmentation of a digital image into regions of different *image texture* (for example the segmentation of an aerial image into fields and forests) is achieved by a quantifiable texture measure, although there are many to choose from [8].

These domain specific applications of surface texture share a more general idea: texture is concerned with the structure and relationship of the component parts of something.

In order to understand this meaning, it is suggested in this paper that large scale sonic structure might emerge from the interactions of sonically dense micro-textures. To be precise, each sonic micro-texture is a sound “grain”, where, arguing by analogy, the grain is constructed from a small element or micro-texture of the image. Small rectangular tiles from an image are mapped to sound grains of short, but perceivable, duration. The grains are emitted asynchronously from a synthesizer in dense clouds in a technique known as granulation [13]. The constituency of the micro-textures and their interactions is determined by a virtual swarm that flies over the textured image. Dynamic sonic macro-texture emerges from the self-organization of the swarm around attractors of high micro-texture.

Three uses of Swarm Tech-tiles are demonstrated here. Firstly, a “tech-tile” of a whole image can be rendered into sound. This produces a short piece of sonic art. Secondly, a user can explore interesting regions of micro-texture by selecting tile locations with a mouse click. More interesting, though, is to allow a swarm/attractor system search for interesting micro-texture, producing an improvisation of indeterminate length. Swarm techniques are pertinent in this context for a number of reasons, not least because different starting configurations lead to different outputs, so that many pieces can be generated from a single image (i.e. the output is improvisatory). Also, the self-organizing properties of swarms will endow the improvisation with temporal order, so that the piece exhibits musical structure.

The following section describes the tech-tile. This is the image-sound map which is used to construct grains (for an improvisation) or a single complete texture (for a composition). Section 3 explains the novel particle swarm used in Swarm Tech-tile. A discussion of how the swarm interprets the local micro-texture of an image and of how particle positions are interpreted as dynamic parameters of a sound stream follows in Sect. 4. The following section continues with a description of some experiments and the final section summarizes the findings of this paper.

2 Texture tiles

There is no necessary connection between images and sounds since each stimulates a different sense (modality). Moreover there are infinitely many possible maps from the two spatial dimensions of an image into the single time dimension of a sound. Some guiding principles, however, are at hand. At the least, a spectator should be able to derive correlations between visual and sonic texture. This principle, in a different context, is known as transparency [5].

A computer generates sound by rendering an audio stream at the sampling rate. An audio stream is, in turn, a succession of samples or digitized elements of a pressure wave. Alternatively, a digital image is an arrangement of adjacent pixels, where each pixel is a digitization of color, for example the three RGB pixel-values, one for each primary color. The simplest map is therefore an association between a single pixel value (R, G or B) and an audio sample. (The map has to stipulate a scaling between positive pixel values and audio samples which can be positive or negative, and with a much larger range of values.)

Locality would imply that closeness in image space relates to closeness in time so that image micro-textures are mapped to sonic micro-textures. Smoothly varying patches of similar color might correspond in a local map to continuous, harmonic waves, perceivable as tones. Edges between colors, on the other hand, which occur at thread boundaries on a textile, might map to buzzy, non-harmonic waves (square, saw-tooth and other discontinuous wave-forms). A very speckled texture might map to a noise stream.

The audio rate determines how quickly samples are rendered, and the simplest choice is to move between pixels at a time interval equal to the reciprocal of the audio rate. But there is a problem of exactly how to move (by horizontal, vertical, or diagonal steps in any combination), and what action to take at a tile edge. In order to solve this, we were guided by the manufacture of an actual textile. Textiles are made by weaving threads length-ways (warp) and width-ways (weft). An obvious way to preserve this feature is to scan the image vertically (along the warp) and horizontally (along the weft). However, in order to preserve locality at the tile edges, each warp/weft scan is continued in the opposite direction, Fig. 1.

The scanning time can be calculated by dividing the number of pixels in the image by the sampling rate. A 64 x 64 tile contains 4096 pixels/samples, and would therefore take 93 ms to scan at a sampling rate of 44100Hz, certainly long enough to hear as a micro-texture and not just as a click. A digital image of a large textile is shown in Figure 2, [9]. The original image, of size 2646 x 1760 = 4.7 megapixels, would produce 106s of sound if rendered as a single tile.

In summary, a texture tile – or tech-tile – is, visually, a rectangular portion (a tile) of an image, or aurally, it is a local parameterization of a sonic stream. The image to sound map is made by simultaneously scanning at the audio rate along the warp and the weft. R, G and B pixel values are picked up and scaled into audio samples. In order to preserve locality, and to prevent anomalous discontinuities (these would be heard as clicks in an otherwise smooth sound), the scan doubles back at the tile edges. Six audio streams are produced; for convenience the warp and weft streams are sent to separate stereo channels.

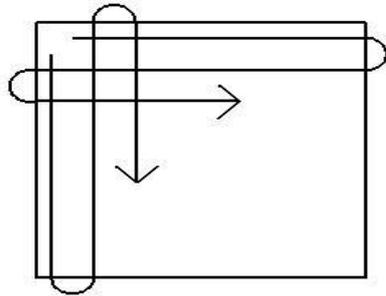


Fig. 1. Warp and weft scan lines

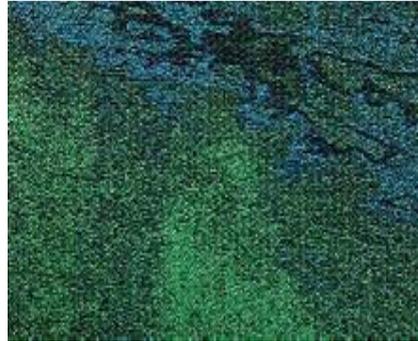


Fig. 2. Uniform and Laundry by Janis Jefferies. *Photo: David Ramkalawon*

3 Swarms and Stigmergy

Swarms of interacting particles, moving in a d -dimensional real space, are a more abstract realization of the A-Life swarm, herd and flock animations initially studied by Reynolds [12]. Reynolds' demonstrated that the collective behavior of these animal groups can be explained through local (rather than globally scripted) interactions. These swarms typically exhibit self-organization (SO). SO emerges from direct particle interactions and also through indirect environment-mediated interactions known as stigmergy [3]. Particle Swarm Optimization (PSO) and Ant Colony Optimization are practical applications of these ideas [10, 3].

The creative use of particle swarms has also been investigated [4, 5]. In the swarm/attractor systems of these papers, particles are drawn towards special points in space known as attractors. These systems are interactive: attractors are positioned as a result of input with an external system (human, or another swarm) and the output from the swarm is an interpretation of particle positions as grains of sound (Swarm Granulator) or musical events (Swarm Music). Coherence and structuring of the respective outputs, and correlations to the inputs are manifestations of SO, induced by the stigmergetic interaction of particles with attractors.

However, in PSO, attractors derive not from interaction but from the evaluation of a fitness function at each particle location [10]. The swarm described below incorporates function evaluation with biologically plausibility particle dynamics.

3.1 Particle dynamics

The particle dynamics for this swarm implementation have been modified from the system described in [5]. Here, particles' perceive attractors and each other within a

local hyper-spherical neighborhood, and not over the entire space¹. This more plausible feature favors the development of subswarms – breakaway groups of particles searching for new attractors. Spring-like interactions have also been replaced by fixed magnitude accelerations which, together with velocity clamping, lead to constant magnitude velocity vectors. This moderates the (somewhat biologically unrealistic) tendency of particles to oscillate about an attractor.

The particle positions $x_i \in [0, x_{max}]^2$ and velocity v_i are updated by determining the local neighborhood of other particles and attractors, where the neighborhoods are determined with respect to the perception radius $r = 0.25x_{max}$. The acceleration towards a particle at x_j and an attractor at p_k is given by

$$a_i(x_j) = Q^2 \frac{(x_i - x_j)}{|x_j - x_i|^3} + \frac{C}{N} \frac{(x_j - x_i)}{|x_j - x_i|} + \frac{C}{M} \frac{(p_k - x_i)}{|p_k - x_i|} \quad (1)$$

where the first term is a collision avoiding repulsion and the second and third terms are attractions towards x_j and p_k respectively. Constants Q and C determine the strength of these accelerations and are set to $x_{max}/32$ and $x_{max}/128$ respectively. N and M are the number of particles/attractors in the neighborhood. The accelerations are added to the current velocity, and the velocity is clamped to $v_{max} = x_{max}/32$. The position is finally updated by adding on the new velocity, and reflecting the particle from the sides of $[0, x_{max}]^2$ if necessary. x_{max} is an arbitrary scale, fixed at 128.

3.2 Attractor stigmergetics

There are important differences between this swarm and PSO in the treatment of attractors. PSO uses a cognitive-social model for attractor placement. In PSO, each particle has a memory of the best position (as measured by an objective function) it has attained (cognitive model) and has knowledge of the best position obtained by other particles (social model) in a topological neighborhood. Although good for optimization, the social model of PSO is not plausible from a swarm perspective (but does make more sense for the social networks found in human culture).

Stigmergy is a biological term for environment-mediated interaction [3]. In Swarm Tech-tiles, particles make decisions based on the attractors that they can *actually see* i.e. in their immediate spatial neighborhood – this implementation of stigmergy is more faithful to biological swarms. Decisions to create or move attractors depend on the value of a micro-texture measure T at the particle position. The particle only has access to values of T at visible attractor positions, but retains a memory of the best texture that it has encountered. However, the longevity of attractors and particle memory is limited in a way that will now be described.

Attractor death. The initial perception radius $r_{init} = 0.25 x_{max}$ of an attractor shrinks by a decay constant $\lambda \in [0, 0.1]$ at each particle visit, eventually leading to attractor annihilation when $r < r_{crit} = 0.5r_{init}$. The biological parallel is a food source that is pro-

¹ This modification was suggested to one of the authors by a participant of the EvoMUSART 2004 workshop.

gressively consumed on each visit. Dynamically this means that a swarm (or subswarm) cannot stagnate around an old attractor.

Attractor movement. Suppose that a particle can see one or more attractors $\{p\}$, and it is currently at a better position x than all these attractors. In this case the best attractor, p_{best} , will be moved to x . The particle always stores $T(p_{best})$, irrespective of movement.

Attractor creation: Suppose that a particle at x cannot currently see any attractors, and $T(x)$ improves upon its memory, $T(p_{best})$. The particle then deposits a new attractor at x and remembers $T(x)$. This rule is necessary to compensate for attractor death. This rule produces fascinating cooperative behavior. A breakaway particle or small subswarm can generate new attractors along a fluctuation, leading to particle “streamers”. This is an example of positive feedback which is an important ingredients of SO [3].

Particle forgetfulness. A particle will forget $T(p_{best})$ if no attractor is visible for a fixed number of interpretation iterations, even though its current $T(x)$ is no improvement. Any location will then become tempting and new attractors can be created. This rule counteracts this scenario: suppose the swarm finds a very good location, and each particle sets an attractor in the neighborhood of this location. After the attractors have been consumed, particles may then wander aimlessly, never depositing a new attractor since $T(p_{best})$ is never exceeded.

The observational effect of these attractor stigmergetics is an alternation between phases of exploration (breakaway subswarms, attractor creation and particle forgetfulness) and exploitation (attractor movement and consumption).

4 Interpretation: Swarm/environment interface

A general architecture for a swarm/attractor system has been given in [5]. Each software module corresponds to a mathematical function. The input interpretation function P explains how the system represents the environment as a pattern of dynamic attractors' p and with an objective function T . The output interpretative function Q explains how the system tries to modify the environment by generating external events $q = Q(x)$ from particle positions x . The internal workings of the system are given by the particle dynamics, $x(t) = f(p, x, v)$ where the t is real time and all the arguments of f are evaluated in the interval $[t-\delta t, t]$ (refer to [5] for a full explanation).

The patterning module has been specified in the previous section. This section continues with a description input/output modules P and Q .

Q maps a particle position x onto a tech-tile $q(\xi, w, h)$ of width w and height h . q is centered at image coordinates ξ where ξ is obtained by a simple re-scaling of $[0, x_{max}]^2$ into the image space $[0, w_{max}] \times [0, h_{max}]$. If q extends beyond the image in either dimension, it is translated so that all scan lines fall inside the image. The warp/weft scans are then performed according to the description of Sect. 2. Q runs in a separate thread to f , pausing by Δt after each interpretation. The particle update time interval, δt , is chosen to give smooth animations and is typically set to 10 ms. Each 64 x 64 tile takes 93 ms to render, giving up to 10 overlapping grains at $\delta t = 10$ ms. Tiles up to dimension 20 x 20 can be rendered without overlap, but tiles this small have no discernible timbre. Overlapping is desirable in granulation since this gives a continuous,

sonically dense stream. Too many overlaps can strain the processor and lead to unwanted audio fragmentation. In practice, a grain rate Δt of 10-50 ms is used for a 1.7 GHz processor with 250 MB memory.

P concerns attractor placement. In Swarm Tech-tiles, an objective function determines the desirability of any sampled location of the textured environment by calculating a measure of micro-texture T for the tile at ξ . T should be chosen to lead the swarm towards “interesting” texture. This is clearly an arbitrary and aesthetically-driven choice. However, certain measures of image texture are commonly used for image segmentation [8] and a couple of these have been tested in experiments.

The first measure is a generalization of the grayscale entropy of the co-occurrence matrix P_{ij} of the image I ,

$$S = - \sum_{i \in I} \sum_{j \in I} P_{ij} \log(P_{ij}), \quad (2)$$

where P_{ij} is the probability that two pixels in I , separated by a displacement vector d , will have grayscale pixel values i and j . The generalization is $T_{entropy} = \max\{S_R, S_G, S_B\}$ where $S_{R,G,B}$ is the entropy of the R , G or B co-occurrence matrix over the tile at ξ . In effect $T_{max\ entropy}$ measures the randomness of the RGB distribution. In order not to discriminate between the two scan directions, a displacement vector that compares pixels at different warps and wefts, $d = (1, 1)$, is used.

The second measure is a simple statistical measure, $T_{stat} = \max\{\sigma_R^2, \sigma_G^2, \sigma_B^2\}$ where $\sigma_{R,G,B}^2$ is the variance of the R , G or B values over the tile at ξ . This measure also favors inhomogeneity and has the advantage that it is efficiently computed. However, T_{stat} does not discriminate between different scales of texture.

Figs 3-6 show two images and their *max entropy* maps. The images of these texture maps of have been equalized in order to highlight differences. Fig. 3 shows the tech-tile for a 5 second multiphonic saxophone tone. This was produced by mapping the complete tone, recorded in 8 bit mono, onto a single tile so that $R = G = B = \text{sample} + 128$, and preserving the warp/weft scan of Fig 1. A multiphonic is a technique whereby a single fingering produces multiple pitches. Since the pitches are harmonic, the samples vary smoothly in time, evident in the shallow texture of Fig. 3, and from the largely uniform texture map of Fig. 4. (The non-uniformities are due to edge effects.)

Figs 5 and 6 show part of the textile of Fig. 2 and the corresponding map. The texture of the textile image is far coarser than the multiphonic tone, showing greater inhomogeneity with a noticeable change in texture between the top right and bottom left of the image. The *max variance* maps of Figs 3 and 5 are broadly similar to the *max entropy* maps. Texture maps can be used as a look-up table for $T(x)$, hence saving a costly computation, but precision is lost due to the conversion of the measure into pixel values.

5 Experiments

Experiments were performed on images of a sunset, a calm seascape, a Eucalyptus tree, the Jefferies textile, recorded saxophone and voice, and synthetic images of pure



Fig. 3. Techtile of a multiphonic saxophone tone



Fig. 4. *Max entropy* texture map of Fig. 3

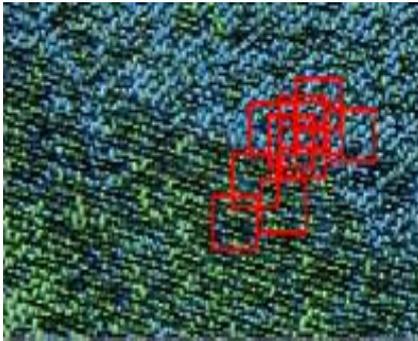


Fig. 5. 595 x 413 part (top left) of the textile of Fig. 2. The red squares show the last ten techtiles rendered by the granulator

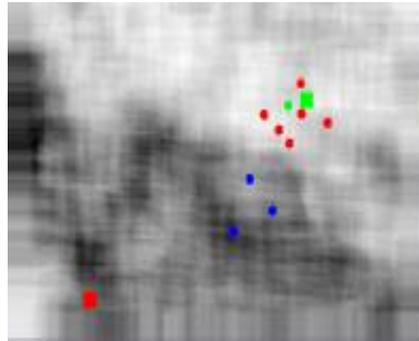


Fig. 6. What the swarm sees when flying above the textile of Fig. 5. Particles (red and blue discs) position attractors (boxes) on regions of high entropy.

tones, white noise, color rainbows and an image with an island of noise centrally placed on a constant color background. Some of these images and sonic tech-tiles are available for download at www.timblackwell.com.

To begin, a single tech-tile, chosen from any of the images by a mouse click, was rendered into sound. The size of this tile varied from 1 x 1 (which is too brief to be audible) up to the size of the entire image. Tiles below about 40 x 40 are heard as clicks with no discernable timbre. Tiles larger than 64 x 64 were used in the following trials since these are capable of probing both sonic and image micro-texture.

Images with areas of blended color, such as the reddening sky around a setting sun produce quiet, pulsating sounds. This is because the warp/weft scan picks up neighboring pixel values that are only gradually changing along a color gradient. As the scan doubles back and retraces the color gradient in the opposite sense, the sample values match this oscillation and harmonic tones at frequencies around $44100/(64 \times 2) \approx 345$ Hz are audible.

When the tile contains edges, the color discontinuity corresponds to a sample jump, so that the scan produces discontinuous waves. If the tile is placed over rough visual texture, for example when placed over the branches of the Eucalyptus image, a loud buzzing and a lesser noisy component is audible. When placed over the leaves on the same image, corresponding to finer texture, the sound contains more noise. The Jefferies textile produces a gentler and less buzzy sound, with more rattle and noise.

A sonic composition of some 106 seconds results from representing the Jefferies textile as a single tile. Although slight, changes in noise spectrum are audible against a background rattle (which has a rather pleasing loom-like hum). Regions of different visual texture produce intervals of varying sonic texture, which confirms that a single large tech-tile preserves large-scale structure.

Textural improvisations were also produced by allowing a ten particle swarm to fly over the image. Attractors were deposited and observed to move to regions of high micro-texture (statistical or entropic). Figs 5 and 6 show a typical configuration during one such improvisation. The swarm flies over the texture map of Fig. 6, where white pixels correspond to tiles of high micro-texture. The red particles of this figure can see an attractor (green box) and the green particle is currently being interpreted. The blue particles are searching for an attractor. There is an old attractor (red box) which has not yet been consumed. Fig. 5 shows the ten tech-tiles sampled from the swarm.

The influence of each attractor shrinks with each particle visit until the attractor evaporates completely. The perception decay rate λ was varied in a number of trials. Larger rates produce more diversity in the population and a more varied sonic output but the surrounding texture is not explored in any detail and the swarm wanders aimlessly above the image. Finally, it was confirmed that a swarm is able to discover the textured region in the island-of-noise image, developing pockets of sonic micro-texture between short intervals of silence.

6 Conclusions

Domain-specific definitions of texture are hard to pin down, but they all refer to a literal tactile meaning of the term. Analogous meanings can be investigated with mappings from *actual* texture onto the domain-specific texture. Although such mappings are arbitrary since they cross three modalities – touch, vision and hearing - they should preserve large-scale structure and possess a degree of “transparency”. Transparency means, in this context, that an intuitive link can be established between, for example, roughness of actual texture and roughness of sound in a sonic texture.

The tech-tile map, inspired by the construction of textiles, establishes just such a map. Experiments on a number of synthetic images, images of natural texture and a textile

image confirm that this map preserves structure and relates qualities such as smoothness (glowing sky during a sunset ↔ quiet, pulsating sounds), roughness (branches, twigs ↔ harsh buzzing sounds) and fine detail (Jefferies textile, leaves ↔ rattles and noise).

It is suggested that a swarm/attractor system in conjunction with a sound granulator can be used to develop a textural improvisation of indeterminate length. The improvisation itself arises from the exploration of the image for regions of high micro-texture. Mechanisms for attractor creation, movement and death, and particle forgetfulness, ensure that the improvisation is sonically diverse.

Swarm Tech-tiles is still in development and future research topics includes the investigation of more sophisticated texture measures (for example using Fourier analysis to quantify harmonicity and noise) and the use of cameras to extract 3D data from a textile. This would close the gap between our tactile experience of texture and the mapping onto sound. Finally, an exciting prospect is to use eye tracking equipment to extract information about how a user views a textured surface. Attractors can then be positioned accordingly and the viewers will be able to hear what they are seeing.

References

1. Athineos, M. and Ellis, D.: Sound Texture Modeling with Linear Prediction in both Time and Frequency Domains. Proceedings of ICASSP-03, Hong Kong (2003)
2. Bailey, D.: Improvisation: its Nature and Practice in Music. The British Library Sound Archive, London (1992)
3. Bonabeau, E., Dorigo M. and Theraulaz T: From Natural to Artificial Swarm Intelligence. Oxford University Press, New York (1999)
4. Blackwell T: Swarm Music: Improvised Music with Multi-Swarms. Proc. AISB '03 Symposium on Artificial Intelligence and Creativity in Arts and Science (2003), 41-49
5. Blackwell, T. and Young M.: Self-Organised Music. Organised Sound 9(2), Cambridge University Press (2004) 123-136
6. Coltrane, J. Ascension. Impulse! Records (2000)
7. Howard, J.: Learning to Compose. Cambridge University Press, Cambridge (1990)
8. Jain, R., Kasturi, R., Schunck, B.: Machine Vision. McGraw-Hill, New York 1995)
9. Jefferies, J: Uniform and Laundry (Restaged I). Produced at the Centre for Contemporary Textiles, Montreal (2003)
10. Kennedy, J. and Eberhart, R: Swarm Intelligence. Morgan Kauffman (2001)
11. Macy, L. (ed): 'Texture', Grove Music Online (Accessed 3 November 2004), <http://www.grovemusic.com>
12. Reynolds, C. 1987. Flocks, Herds, and Schools: A Distributed Behavioural Model. SIGGRAPH '87 21(4) (1987) 25-34
13. Roads, C: Microsound. MIT Press, Cambridge, MA (2001)
14. Wishart, T. On Sonic Art. (Revised Edition.) Harwood Academic, Amsterdam (1996)