# SKETCHING CONCATENATIVE SYNTHESIS: SEARCHING FOR AUDIOVISUAL ISOMORPHISMS IN REDUCED MODES

**Augoustinos Tsiros, Grégory Leplâtre, Michael Smyth**

Centre for Interaction Design
Edinburgh Napier University
Edinburgh, United Kingdom
a.tsiros@napier.ac.uk, g.leplatre@napier.ac.uk, m.smyth@napier.ac.uk

## ABSTRACT

This paper presents a prototype allowing the control of a concatenative synthesis algorithm using a 2D sketching interface. The design of the system is underpinned by a preliminary discussion in which isomorphisms between auditory and visual phenomena are identified. We support that certain qualities of sound and graphics are inherently cross-modal. Following this reasoning, a mapping strategy between low-level auditory and visual features was developed. The mapping enables the selection of audio units based on five feature data streams that derive from the statistical analysis of the sketch.

## 1. INTRODUCTION

The work presented in this paper pertains to the development of multimodal systems that allow for novel means of interacting with sound processing/synthesis algorithms. The main goal of such systems is to offer new creative opportunities to composers and sound designers. These systems commonly rely on the use of metaphors or associations between visual and auditory objects. However the question of how to approach audio-visual association and the development of such metaphor systems is not easy to answer. This is mainly due to two reasons: (i) the differences in the neurophysiology of auditory and visual perception, and (ii) the difficulties to fully comprehend the nature, the complexity, and the processes involved in conscious experience, which results in a limited understanding of the relationship between neurophysiology and subjective experience. However it should be emphasized that in the process of designing audio-visual associations, beyond the differences of auditory and visual perception and the difficulties that subjective experience introduces there is an area of convergence. This overlap originates in the observation of phenomena in the natural environment which can be experienced through a number of sensory modalities. Hence a number of similarities in typological and conceptual classes that describe form and structure seem to translate across auditory and visual modalities. This paper supports that consideration of these properties, as conceptual spaces shared by multiple forms of sensing, could have salient roles in designing audio-visual or in fact any other multi-sensory associations.

Drawing on this understanding, we analyse similarities between descriptions of sound and that of a sketch. In order to discuss the similarities in typological classes of organization we adopt the term *Isomorphism*. The term *Isomorphism* has been used in a number of disciplines such as mathematics, biology, and psychology to describe similarity of form and structure as the etymology of the word suggests. It is a composite word which is consisting of the prefix -iso from the Greek word *isos* which mean equal, and the suffix -morphism from the Greek word – *morphé* which means shape or form. Isomorphism is considered here from a Gestalt point of view "*which is not a strict structural isomorphism i.e. a literal isomorphism in the physical structure of the representation; but merely a functional isomorphism i.e. a behaviour of the system as if it were physically isomorphic*" [1:92]. We examine the domains which audio-visual features inhabit (e.g. spectral, temporal, textural, spatial), the behaviour they describe (e.g. energy, distribution), and their affordances (e.g. possibility for transformation). We show how this analysis was put into practice to define the design of an interface that enables interaction with a variable number of concatenative synthesis modules (i.e. *CataRT* developed by the IRCAM/IMTR research group, see [2]), through the act of sketching, video demonstrations of the tool are available at: *http://inplayground.wordpress.com/morpheme/*

## 2. PROBLEMS RELATED TO AUDIO-VISUAL ASSOCIATION

Are there intrinsic characteristics that suggest how audio-visual association should be made? The technologies required to develop audio-visual and other multi-sensory association are available. The search for commonly accepted principles for audio-visual correspondence has been pursued since ancient times [3,4,5], and it is still significant today. However the problem of how multi-sensory associations should be approached remains.

It has been argued that physical attributes of visual and auditory perception do not provide a clear system of correspondence [5, 6]. The differences in the affordances of auditory and visual perception in terms of psychophysics are well known [7, 8]. Moreover the issue of subjective interpretation and particularly with regards to meaning and semantics are a serious obstacle in forming a commonly accepted theory of correspondence [9]. Achieving an exact isomorphic visual representation of sound without any element of subjective decision-making is almost impossible [10]. This suggests that associations across media are subjective, predetermined only by cultural fashions, practices and preferences.

According to the above, it could be said that forming a general theory for multisensory association is merely hopeless, a utopia. The major issue appears to be the difference between the functioning of the brain and the sensory system (i.e. the neurophysiological level) and that of the mind (i.e. phenomenological, experiential level) [11]. As Nagel states, in order to overcome this problem we should set aside temporarily questions related to the relationship between mind and brain and identify new approaches for studying experience. He suggested that we should devise new methods based on objective phenomenology *"i.e. to quantify the structural features of the subjective experience in objective terms, without committing to any particular neurophysiological theory of perceptual representation"* [11:382]. While it has been argued that such reductionist approaches are unlikely to capture fully what is perceived at an experiential/conscious level [12, 13], such methods might allow us to capture a number of structural characteristics of subjective experience in an objective form which in turn could allow for a low resolution representation of a phenomenon to be constructed [12: 167].

## 3. VISUAL ARTEFACTS

### 3.1 Sketching and thinking

Our ability to express our thoughts and ideas in visual terms cannot be underestimated. As soon as a few marks are made on a piece of paper the brain reacts to them. These reactions develop naturally based on the experience of a lifetime of visual associations and a continuous effort of perceptual mechanisms to identify form, structure, and meaning. Auditory perception is not different in this respect [7]. During this process, unexpected thoughts occur about where and/or how to make the next actions. Drawing is about seeing, and envisioning is about capturing and exploring a series of marks that can inspire imagination, observation, and thought [14]. For example in the context of design usually the starting point is a fuzzy or vague idea. This idea can be expressed through a sketch which might be a free drawing or a more diagrammatic form. This enables further exploration and refinement of the idea through the process of drawing or sketching. So it could be said that visualizing concepts through sketching has an inherent experimental and speculative nature. The power of sketching lies in its ability to remove/abstract technical complications related to the implementation of a design or an idea, which in turn allows the designer to focus on conceptual development [15, 16]. It is a speculative medium that enables complexity and varying degrees of resolution to be captured. A number of studies have explored sketching as: (i) a model of interaction, and (ii) a medium for conceptual design in both computer music and HCI [17, 18, 19, 20, 21].

### 3.2 Visual artefacts & Music

Evidence show that composers, particularly in electroacoustic music, use sketching and drawing as a medium for conceptualisation of ideas at different stages of the composition process, [21]. These artefacts are created usually to represent micro, meso, and macro structures of a sound and/or a musical composition and they usually depict information related to the activity of sound in the temporal, spatial and spectral domains. It has been proposed that musical sketches, unlike architectural and product design sketches, can be very idiosyncratic and vague in nature [22]. Even in traditional music notation, the score is not an accurate representation of the music. It is a rather loose representation that leaves space for interpretation by the musicians. Weaknesses can be found on all forms of notation or representation. However what matters is neither the metaphors used nor the phenomena they signify *"but what lies behind them, and what we must create by means of these symbols"* [23]. As mentioned earlier metaphors provide a medium through which ideas and designs can be realized, communicated, and refined. Metaphors work through tension between opposing concepts (e.g. big-small, sparse-dense), dialectic relationships, and difference. As Coyne states, the tension is between sameness and difference. In metaphor we detect sameness in spite of the overwhelming presence of difference. The suitability of a metaphor system depends on context [24:296] For instance as sound qualities increasingly become of equal importance to the organisation of notes within the process of music composition, traditional forms of music notation progressively become inadequate in analytical terms. This raises the question of the appropriateness of a metaphor for a particular application.

### 3.3 Types of representation

Symbolic representation methods (e.g. music notation, writing) can provide a well coded system that can support significant analytical accuracy, but they are usually very inaccessible and difficult to read. Iconic representations (e.g. waveform, spectrogram) link some degree of perceived qualities of a sound to that of graphics. Iconic representation can be relatively easy to read and comprehend [25]. To paraphrase Smalley, descriptive and conceptual tools which can relate and organise sounds and structures, could be a valuable compositional aid as composers nowadays have a vast array of sounds at their disposal. Smalley pointed out that the importance lies in forms of notation which are more concerned with *"spectral qualities than actual notes, more concerned with varieties of motion and flexible fluctuation in time than metrical time, more concerned with sounds whose source are mysterious or ambiguous rather than blatantly obvious"* [10: 109]. A number of studies emphasise that defining approaches for visual representation of sound is extremely important for analytical, compositional and pedagogical purposes [10, 26, 27, 28, 29], but also for transformation of sound. However, the significance for the development of such systems of correspondence is not limited to computer music but concerns a wider range of disciplines.

# 4. HIERARCHICAL STRUCTURES AND ISOMORPHISM

## 4.1 Manifold Experiences

Holtzman claimed that the possibility of isomorphisms across concepts becomes possible due to similarities of organisation of experiential constructs across different hierarchical systems [30:267]. He describes deep structures and surface structures, where deep structures are organised elements that give rise to higher level constructs. This forms a continuum or cascade, where this higher level construct become the deep structures to another surface structure. Similarly according to gestalt psychology, organised forms in perception are called gestalten. Gestaltens are multidimensional entities, patterns, qualities consisting of a variable number of attributes, and signify concepts which are not fully described / reduced to their constituent parts [31]. For example, mechanical motion enables sound, organized sound enables notes, that in turn allows musical sequence to be constructed, which enables the composition of a piece of music. Hence, there is a hierarchical aspect to relationships between concepts and structures and there seems to be an interdependence across levels.

Based on this, it could be said that the higher the level of organisation, the more difficult it becomes for multi-sensory representations to act in an isomorphic manner. For instance, low-level organisational features of a vocalised word or a painting, such as densities, energy, dynamics can be represented more easily across modalities than higher level constructs, such as the semantic content of a word or a painting. In the latter, the problem of subjectivity of interpretation becomes more prominent. Moreover the understanding of hierarchical structures could be useful when designing metaphors for multiparametric control, as many parameters can be abstracted under a higher level construct. This method to be successful would require modelling of the coupling relationship, and the dependencies between the lower level attributes.

## 4.2 Audiovisual reduction and Isomorphism

Hearing and seeing are in a number of ways different to listening and watching. Listening and watching involve some form of observation and interpretation of the received signals. The observation could have as focus: (i) the content and/or the cause of the stimulus such as causal, and semantic listening, and (ii) the phenomenological observation of the intrinsic characteristics of the sensory stimulus (e.g. mode of reduced listening). In reduced listening, sound no longer acts as the vehicle/signifier that signifies another object or concept, but rather it is the signified itself [32]. It could be argued that this reduced approach could be applied to vision as well. For example, looking at a painting of a landscape, one can focus on low-level qualities such as tonality, texture, shapes or focus on higher-level objects like birds, trees, cultural and emotional connotations. We support that reduced type observation and interpretation might have a salient role to enable audio-visual associations to be approached in a less subjective manner, as certain types of isomorphisms of concepts and their substance develop through our embodied experience of our physical environment.

These similarities originate in the observation of phenomena in the natural environment which can be experienced through a number of sensory modalities. Isomorphic perceptual organisation is essential as there has to be a complementary mental representation of the features that describe the physical world, independently of which sensory modality is used. For example, the concept of texture resides in a sensory experience that has been learned through vision, sound and touch [33]. This suggests that the typology of perceived qualities of textures might be independent from modality. Wishart, on the phenomenology of sound image, referred to deep structures as: *"Various oppositions and transformations which constitute basic elements of human thinking"* [34:167]. Identifying these overlapping areas is very important for both intuitive audio-visual associations and interaction with sound synthesis and processing parameters.

## 4.3 Multimodality and perspectives from the neuroscience

Perceptual organization is a process that is involuntary and active. It occurs in an attempt to construct a faithful mental representation of the stimulus received by the external environment. Consequently it would be naive to think that perceptual processes do not interact and integrate amongst the stimulus which arrives from a number of sensory inputs in order to construct that mental representation [35]. It has been argued that studying sensory modalities in an isolated manner could only be justifiable if the stimuli of sensory modalities are interpreted independently to that of other sensory input [36]. At different times, signals from different sensory modalities may have a confirmatory relationship or discrepant relationship and cross modulation across sensory signals is far greater than one would expect, [36, 37, 38, 39]. These studies provide evidence which suggests that overall audition tends to be dominant over vision in temporal aspects of perception and vision tends to be dominant over audition in spatial aspects of perception. These interactions occur mainly when exposure to the audiovisual stimuli is synchronous. They appear to be implicated in a number of low level temporal, spatial, syntactic, and descriptive cross-modulations. These findings have changed the notion of modularity of perceptual processing and provide adequate evidence for a shift to more "*interactive and integrative paradigm of perceptual processing*" [40]**.** This research aims to explore concepts related to low level structural features of perceptual organization for the design of audiovisual interaction with sound synthesis and signal processing parameters. This approach might have the potential to enhance our abilities to comprehend, and interact with sound and its parameters among other information. The studies discussed in this section also raise questions about how multimodality might affect our perception of sound if interaction through the use of well designed isomorphic representation is enabled.

## 5. MORPHEME INTERFACE

### 5.1 Purpose

*Morpheme* allows the control of a concatenative synthesis module (see [2] for more information on concatenative synthesis) through the act of sketching on a digital canvas using a set procedural brushes. The aim is to identify isomorphic models for feature matching between the qualities of the audio units and that of the sketch. This is achieved through the calibration of a procedural brush, which is used as the input for the retrieval algorithms. Through the iterative process of sketching, control over the sound is attained. Tapping into the act of sketching and its phenomenal capacity as a medium to conceptualize and explore ideas and designs, we attempted to use a set of metaphors to facilitate the exploration of an audio corpus.

We have made a conscious decision to focus on the analysis of the artefact (i.e., the sketch) for the interaction with the selection algorithm, as opposed to the history of the construction of the sketch (i.e. path stroke coordinates, and settings). This decision was made considering that the artefact is what is important and what the practitioner sees and works with. This is not to say that the expressive qualities and the real-time aspects of the gestures are not important, but rather that the focus is in identifying ways to capture what is important about the gesture on the artefact, possibly through physical modelling. We consider that focusing on the analysis of the sketch would make it easier to interpret what is perceptually relevant for the performer in visual terms, as oppose to the history of the construction which leads to a way more fragmented account of what is on the sketch. Moreover trying to reconstruct the sketch from its construction history was deemed to difficult. The interface presented here aims to extend the compositional and sound design potential of concatenative synthesis. Interesting results emerge in terms of interaction, multimodality, and sound synthesis,

*Morpheme* was developed using the graphical programming environment Max MSP. The implementation of concatenative synthesis that *Morpheme*s the user to control, works by segmenting a number of audio files into small units. The units are then analysed, tagged with the analysis data, and stored in a database. Synthesis is accomplished by recombining audio units from the database based on an input feature data stream. In the context of this project, the data stream results from statistical analysis of the pixels of the canvas. Based on the data from the analysis, retrieval of audio units is achieved. Video examples of the interface can be found at: *http://inplayground.wordpress.com/morpheme/* .

### 5.2 System Architecture

The diagram in Figure 1 illustrates the architecture of the control system for five concatenative synthesis unit selection modules, all of which connect to the same audio corpora. The digital canvas where the sketches are drawn is divided into five areas corresponding to each selection module. Each section's pixels are analysed using a combination of statistical methods (i.e. mean, variance, kurtosis, and feature analysis). Therefore, the statistical data that belong to each section are used to select audio units only within a pre-defined range of the corpus spectrum, which is relative to the one fifth of the overall spectral content of the corpus.

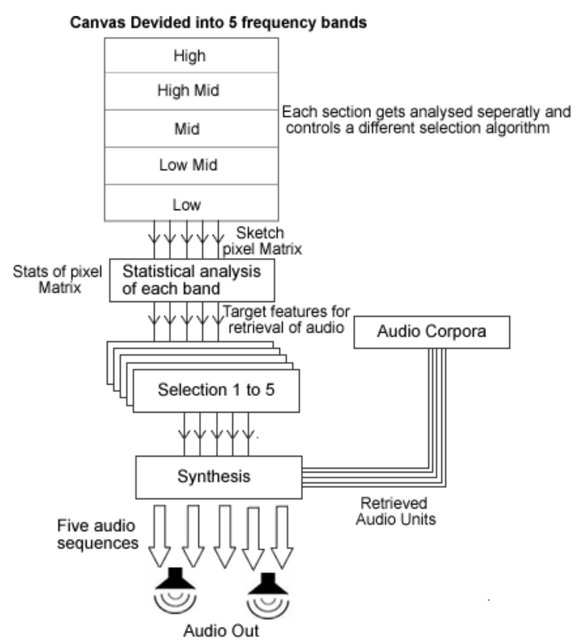


**Figure 1** . System architecture for retrieval of a variable number of audio-units.

The main motivation for dividing the canvas in five sections was for convenience of the analysis, more interactive approaches to the analysis of the canvas are proposed in the future work section. Figures 2. and 3. show screen-shots from the *Morpheme's* graphical interface.

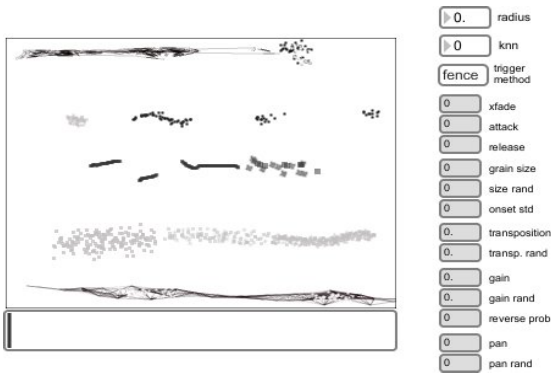


***Figure 2*** *. The graphical interface of Morpheme*

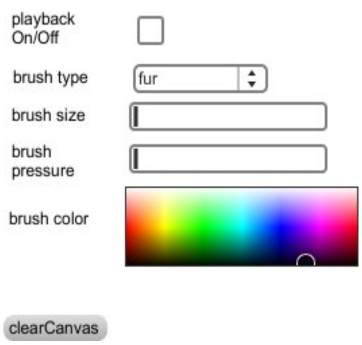


**Figure 3** . The parametric controls of the brush

### 5.3 Association for audio-unit selection

This section discusses the rationale based on which association between features of audio-units of the corpus, and that of sketch has been accomplished. The canvas is scanned vertically stepping through each column of the pixel matrix from left to right. Following the score conversions, the x-axis represents time and the y-axis represents frequency. The features used for the retrieval of audio units include periodicity, spectral flatness, loudness and spectral centroid. The features extracted from the sketch include brightness, density, path centroid, and variance. Tables 1 and 2 illustrate the features used for associations, the domains these attributes inhabit, and their affordances. As it will be shown, this type of analysis might help the development of intuitive audiovisual mapping.

| **Canvas features** | | |
|---|---|---|
| **Parameters** | **Domain** | **Affordances** |
| Value | Texture energy | Dark-Bright |
| Density | Texture distribution | Dense-Sparse |
| Path | Spatial distribution of texture | Continuous-Discontinuous/ Straight-Curved/ |
| Variance | Spatial and Textural distribution | Invariant-variant |

**Table 1.** Analytical cataloguing of canvas features in terms of parameters, domains and affordances.

| **Audio-unit selection features** | | |
|---|---|---|
| **Parameter** | **Domain** | **Affordances** |
| Periodicity | Spectral and Temporal distribution | Periodic-Non-periodic |
| Spectral Flatness | Spectral distribution | Symmetry-Asymmetry |
| Loudness | Energy | Quiet-Loud |
| Spectral Centroid | Spectral distribution | Low-High/ Variant-Invariant |

**Table 2.** Analytical cataloguing of concatenative synthesis features in terms of parameters, domains and affordances.

Identifying the domains which each parameter inhabits and its affordances can be informative with regards to audio-visual representation, as it might expose any shared phenomenological basis between sound and visual features. For example, in this context, we know that the x-axis is used to illustrate time and the y-axis spectra. This suggests that all the features of the concatenative synthesis which are related to the temporal domain will have to use x-axis as their medium for representation, while all features related to spectra will have to use the y-axis as their medium for representation. Moreover taking a deeper look at the nature of the parameters in relation to the domain they inhabit, we can see that in both tables there are some recurrent themes such as energy, distribution and barycentre of distribution. Brightness and loudness are indicators of energy. Density and spectral flatness indicate morphological aspects of texture and spectral distribution. Path and spectral centroid both relate respectively to the perceived barycentre of texture and spectral distribution. Variance and periodicity are both parameters which depend on two domains. Sound periodicity depends on temporal and spectral organisation, while graphical periodicity depends on both coordinate axes and the texture organisation. Looking further at the affordances of the features can also be informative as to how sound and graphic descriptors could be associated. Table 3 illustrates the associations between audio and visual features which were developed based on the analysis described above. Figure 4 illustrates four examples that show some extreme cases of how different statistical distributions of the pixel data are associated to the features of the audio units in the database.

| **Feature Associations** | |
|---|---|
| **Graphic descriptors** | **Audio descriptors** |
| Variance x-axis | Periodicity |
| Variance y-axis | Spectral flatness |
| Pixel mean Brightness | Loudness |
| Regression performed horizontally on Feature analysis data | Spectral Centroid |

**Table 3.** Associations between graphical and audio-units features.

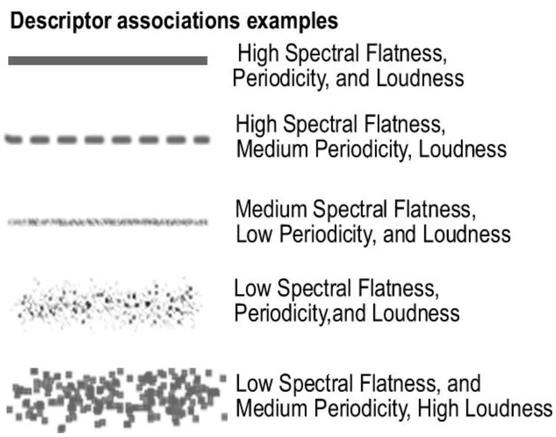


**Figure 4** . Five examples of the interpretation of the mapping association between visual and concatenative synthesis retrieval descriptors.

### 5.4 Feature Mapping

Once an audio corpora has been analysed, the minimum and maximum values of the audio units related to the features of periodicity, spectral centroid, loudness and spectral flatness, are used for scaling the data that derive from the statistical analysis of the canvas. Variance across the x-axis is estimated as we are stepping through the matrix of the canvas, from left to right, and column after column. This is achieved using a computer vision algorithm that measures the rate of change of the pixels by comparing them to the pixel configuration of the last frame. The data are then used as a target feature for retrieval based on periodicity. Hence, the less variance across the x-axis there is, the more periodic the retrieved units will be, while variance across the y-axis is used as a target feature for the retrieval of audio-units based on spectral flatness. Hence the less the variance across the y-axis the flatter the frequency spectrum around the barycentre. The average brightness of the pixels vertically in each of the five areas of the canvas is used as a target feature for retrieval based on loudness. The computer vision library Open.CV (for more information see [41]) is used for feature analysis and feature analysis is performed on each one of the five areas of the canvas. The feature data at high resolution tend to line up across the barycentre of the pixel distribution on the canvas. Then a linear regression is performed on feature plot. The curve values are stored in a matrix and are used as the target feature for spectral centroid.

### 5.6 Future Work

We are currently considering the development of algorithms for (i) blob detection (ii) feature analysis of each blob, and (iii) multiparametric control of the concatenative synthesis parameters. The aim is to use two layers of analysis. The low level statistical analysis described above will be used for the retrieval of audio units, while a second layer which analyses features at blob level will be used for the control of the synthesis parameters. The visual features we intend to analyse through the development of the algorithms include continuity, variance, opacity, size and roundness inherent in each blob. The synthesis parameters which we would like to control include: grain rate and rate randomness, grain size and size randomness, transposition randomness, and x-fade. A fuzzy classifier and a set of rules might be used to handle the data derive from the blob analysis and to dynamically generate mapping strategies.

The interaction with *Morpheme* is currently achieved using a stylus and a graphic tablet as an input device which can be considered quite limiting but expressive. Future versions will be implemented for multi-touch surfaces. Moreover the current system does not provide instant audio feedback when a brush stroke is made which is a shortcoming of the current version and could be seen as inconsiderate towards the temporal nature of sound. As mentioned earlier, the prototype currently scans the canvas from left to right. A number of alternative approaches have been identified that would allow practitioners to explore the canvas in a non-linear and real-time manner. This could involve playback based on a radius around a pointer (i.e. finger) or the area between a number of points (i.e. multiple fingers). Zooming in different areas of the canvas could also provide an interesting interaction model for playback of the sketch. Indeed, on each scaling transformation, a complete new view can be achieved, which might result in a very different sonic outcome. Moreover sliding the canvas at a certain speed and having a static time cursor is another possibility.

The sketching algorithm that we currently use enables the configuration of procedural parameters of the brush, both in qualitative and quantitative manner. However the control of these parameters is currently achieved through the use of standard graphical controls such as sliders. This means that when a gesture is performed on the canvas, the only data that is exploited is the path of the gesture. An algorithm could be developed to enable control of the procedural parameters of the brush (i.e. size, pressure) based on the physical dynamics of the gesture (i.e. speed, pressure, acceleration, directionality). This could enable to capture the nuances and the communicative aspects of the gesture in the artefact (i.e. the sketch). Hence this approach is expected to provide a finer control of the synthesis parameters. Finally, we plan to conduct a number of participatory studies to explore and test different approaches for (i) the association between audio and visuals, (ii) multiparametric strategies for mapping, and (iii) appropriate interaction models for the exploration of the canvas, and the corpus.

## 6. CONCLUSIONS

When comparing auditory and visual perception at the level of physical signals and psychophysics, one becomes increasingly aware of the differences between the two. Through a theoretical analysis we explored the possibilities for multisensory association, based typological qualities of form and structure which relate to low-level descriptors of auditory and visual phenomena found in physical environment. This suggests that a number of typological classes of perceptual organization are modality independent. We support that the identification of these relationships could be vital to designing intuitive multi-sensory associations that can be understood as opposed to learned. These shared areas of perceptual organisation were explored for the design of an audio-visual interface for intuitive interaction with a concatenative synthesis module.

Based on our experience of the interaction using the *Morpheme* interface, we believe that an intuitive audio-visual mapping for the selection of audio-units has been developed. Through the act of sketching, practitioners can express compositional intentions by means of a perceptually meaningful set of metaphors (i.e. position, path, density, and variation). We feel that the act of sketching is well suited to the exploration of an audio corpora, through the triadic iterative process of act, reflection, and production sound can be synthesised. It could be said that *Morpheme* facilitates and promotes an exploratory approach, and does not require users to have any prior knowledge or special training other that using a pointing device. However having an understanding of the way the synthesis method and the mapping works, as well as of the content of the audio corpus certainly helps in attaining finer control over the outcome sound. Moreover the content of the corpus also plays an important role in achieving intuitive audiovisual mapping. For instance if the corpus holds audio segments from a white noise audio signal, then there is no periodic sound to be selected. Consequently drawing a sparse or a dense line would result in the same sonic output. The approach and the interface described in this paper offer the prospects to extend the compositional and sound design potential of concatenative synthesis.

## Acknowledgments

The authors would like to acknowledge the IIDI (Institute for Informatics & Digital Innovation) for their financial support. Also we would like to thank the IRCAM research centre for sharing the *CataRT* system.

## REFERENCES

[1] Köhler W., The Task of Gestalt Psychology. Princeton NY. Princeton University Press, 1969.

[2] Schwarz D., Beller G., Verbrugghe B. and Britton S., Real-Time Corpus Based Concatenative Synthesis with CataRT. Intl. Conf on Digital Audio Fx, 2006.

[3] Campen C. V., Artistic and Psychological Experiments with Synesthesia, The MIT Press, Leonardo, Vol. 32, No. 1 (1999), pp. 9-14

[4] Campen, C. V., The Hidden Sense: Synesthesia in Art and Science, Cambridge: MIT Press, 2008. Chapters 8 & 9.

[5] Berman C., Synesthesia and the Arts. MIT Press, Leonardo, Vol. 32, No. 1 (1999), pp. 15-22.

[6] Hertz P., Synesthetic Art-An Imaginary Number. The MIT Press. Leonardo, Vol. 32, No. 5, p. 399-404, 1999.

[7] Bregman A. S., Auditory Scene Analysis:The perceptual organization of sound. MIT press Cambridge MA, 1994 .

[8] Chion, M., Audio-Vision: Sound on Screen. Columbia University Press, 1994 .

[9] Stevens S. S., On the Theory of Scales of Measurement". Reprinted from SCIENCE, Vol. 103, no. 2684, p. 677-680, 1946.

[10] Smalley D., Spectromorphology: Explaining sound-shapes. Organised Sound: Vol. 2, no. 2. Cambridge University Press: 107-126, 1997.

[11] Lehar S., Gestalt isomorphism and the primacy of subjective conscious experience: A Gestalt Bubble model. Behavioral and Brain Sciences, 26 , pp 375-408, 2003.

[12] Nagel T., What is it like to be a bat? In Readings in Philosophy of Psychology, Vol. 1, pp.159-168, 1980. Cambridge, MA Harvard University Press

[13] Harth E., Art and Reductionism. Journal of Counsiousness Studies, 11, 111-116, 2004.

[14] Olin L., More than Wriggling your wrist or your mouse: thinking seeing and drawing. In Drawing /Thinking: Confronting an electronic age. Taylor & Francis, 2008. Chapter 6.

[15] Grubbs C., Drawing life drawing ideas. In Drawing /Thinking: Confronting an electronic age. Taylor & Francis, 2008. Chapter 7.

[16] Barron E., Drawing in the digital age. Drawing /Thinking: Confronting an electronic age. Taylor & Francis, 2008. Chapter 8.

[17] Nelson P., The UPIC system as an instrument of learning. Organised Sound, 2, Cambridge University Press: 35-42, 1997.

[18] Thiebaut J. B., Sketching music: representation and composition. PhD Thesis, Queen Mary, University of London, 2010.

[19] Farbood M., Kaufman H., Jennings K., Composing with Hyperscore: an intuitive interface of visualizing musical structure. Proceedings of the International Conference on Computer Music-ICMC, 2007.
[20] Buxton B., Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann Publisher Inc., 2007.
[21] Healey, P. and Thiebaut, J. B., Sketching Musical Composition. Proceedings of the Cognitive Science Society conference, Nashville, USA, 2007.
[22] Goldschmidt, G., The Dialectics of Sketching, Design Studies. V. 4, N. 2, 123—143, 1991.
[23] Karkoschka, E., Notation in New Music: A Critical Guide to Interpretation and Realisation, trans. Ruth Koenig. New York: Praeger Publishers, 1972. In [27].
[24] Coyne R., Design Information in the Postmodern Age. MIT Press Cambridge MA, 1995.
[25] Couprie P. Graphical representation: an analytical and publication tool for electroacoustic music, Organised Sound: Vol. 9, no. 1. Cambridge University Press: 109 -113, 2004.
[26] Schaeffer, P., Traité des objets musicaux. Paris: É ditions du Seuil, 1966. In see [29, 32].
[27] Patton K.., Morphological notation for interactive electroacoustic music. Organised Sound: Vol. 12, no. 2. Cambridge University Press: 123-128, 2007.
[28] Blackburn M., Composing from spectro-morphological vocabulary : proposed application, pedagogy and metadata. Electroacoustic Music Studies, 2009.
[29] Thoresen L., Hedman A., Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer's typomorphology. Organised Sound: Vol. 12, no. 2. Cambridge University Press: 129-141, 2007.
[30] Holtzman S. R., Digital Mantras. MIT Press Cambridge MA, 1994.
[31] North W., The Handbook of Semiotics. Indiana University Press, 1995.
[32] Kane, B., L'Objet Sonore Maintenant: Pierre Schaeffer, Sound Objects and the Phenomenological Reduction. Organised SoundCambridge University Press, 12(1): 15–24, 2007.
[33] Smalley D., Space-form and the acousmatic image. Organised Sound, V.12 n.1, Cambridge University Press: p. 35-58, 2007.
[34] Wishart, T., On Sonic Art. Taylor and Francis Group, 1996.
[35] Ward J., Huckstep B.,& Tsakanikos E., Sound-Colour Synaesthesia: To What Extent Does It Uses Cross-Modal Mechanisms Common To Us All? Department of Psychology, University College London, Cortex, 42, 264-280, 2006.
[36] Shimojo, S. and Shams, L., Sensory modalities are not separate modalities: plasticity and interactions. Elsevier Science LTD, 2001.
[37] Marks, L. E., On cross-modal similarity: auditory–visual interactions in speeded discrimination. J. Exp. Psychol. Hum. Percept. Perform. 13, 384–394, 1987.
[38] Sekuler, R., Sekuler A. B.,Sound Alters visual motion perception. Nature, 385:308, 1997.
[39] Meylan, R. V., Murray M. M., Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. Elsevier Inc. NeuroImage 35 244-254, 2007.
[40] Shams, L., Iwaki, S., Chawla, A., Bhattacharya, J., Early modulation of visual cortex by sound: an MEG study. Science Direct, Neuroscience Letters 378, 76-81, 2005.
[41] Palletier J. M., OpenCV computer vision for Jitter: http://jmpelletier.com/cvjit/.