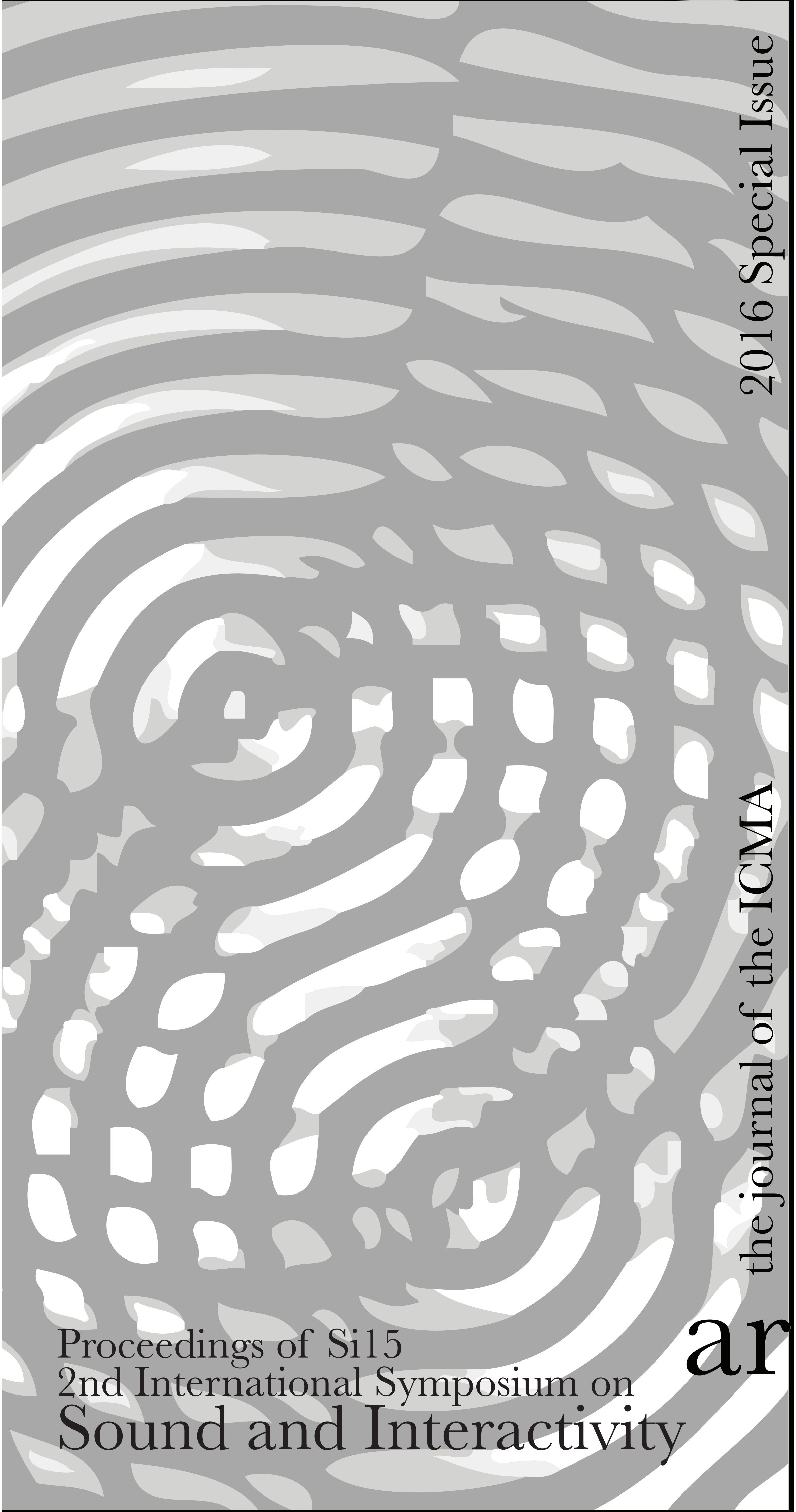


array

the journal of the ICMA

2016 Special Issue

Proceedings of Si15
2nd International Symposium on
Sound and Interactivity

Soundislands

Proceedings of the 2nd International Symposium on Sound & Interactivity

20-21 August 2015, Singapore

Part of Soundislands Festival, 18-23 August 2015

ADM SCHOOL OF ART, DESIGN AND MEDIA 10 YEARS 2005 - 2015

CENTRE FOR LIBERAL ARTS AND SOCIAL SCIENCES

Istituto Italiano di Cultura

NANYANG TECHNOLOGICAL UNIVERSITY

ArtScience Museum MARINA BAY SANDS

NATIONAL ARTS COUNCIL SINGAPORE

BOLLYWOOD VEGGIES SINGAPORE

Asia New Zealand Foundation

creative nz ARTS COUNCIL OF NEW ZEALAND TOI AOTEAROA

Introduction to the Special Issue by the Editor

Christopher Haworth

christopher.p.haworth@gmail.com

I am delighted to present this **Special Issue of *Array*: Proceedings of Si15, the 2nd International Symposium on Sound and Interactivity.** *Array* is the online publication of the International Computer Music Association (ICMA, URL http://computermusic.org). As well as hosting news, reports, reviews, keynote speeches, and other material pertaining to the annual International Computer Music Conference (ICMC), *Array* periodically covers events that will be of interest to the computer music community at large: hence this special issue, which marks the second *Array* publication for the year 2016. I am grateful to PerMagnus Lindborg and Suzy Styles for their excellent work in editing and compiling this interesting and rich collection, and hope that this special issue spurs further new territories for *Array*. Expanding and enriching *Array* depends, as always, on input from you, the readers, so please do get in touch with any ideas, big or small, for future issues: themes, articles, concerts to review... this is your publication.

Contents

Introduction

Invited Keynotes

Aesthetics

Perception

Technology

Sound Art

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 1-4.

Introduction to the *Special Issue: Proceedings of Si15* by the Guest Editors

PerMagnus Lindborg [*], Suzy J. Styles [†]

[*] Interactive Media, School of Art, Design, and Media

[†] Division of Psychology, School of Humanities and Social Sciences

Nanyang Technological University, Singapore

permagnus@ntu.edu.sg

1. Special Issue

Welcome to this **Special Issue of *Array*: Proceedings of Si15, the 2nd International Symposium on Sound and Interactivity.**

The articles in the present issue originated in the **Si15 Soundislands Festival**, which was held in Singapore 18–23 August 2015. The festival events included five invited artist performances, two scientific keynotes and two days of proceedings, a commissioned sound installation, an afternoon of public talks, an internet panel, two pedagogic workshops, a concert with young performers, and more than fifty artworks and scientific papers in numerous forms and formats selected from an open call (http://soundislands.com/si15).

We are thrilled to present 20 articles, by 31 authors, emanating from Si15. The articles have been extended and thoroughly revised for this special issue of Array. They cover a range of topics related to aesthetics, perception, technology, and sound art. We hope that you will enjoy the fruits of the authors' labour and therein discover many a stimulating thought.

2. Theme of Si15

While starting as separate ideas, the fates of Soundislands Festival 2015 and the Si15 Symposium gradually intertwined, and eventually joined under the theme " ~sound:senses". This somewhat cryptic notation expresses the intersection between the arts of sound and the sciences of psychology. The swung dash (~) indicates a dependency and the colon (:) specifies a co-variation between variables. Thus, the focus is on the interaction between physical sound and human senses. Yet, one might interrogate: *"What does the interaction denote? What is on the other side of this function*?"; and the reply must be: *"You!"*

Sound is a medium that acts as a glue between the senses. Hearing is deeply intertwined with vision and movement, and beyond that, researchers and artists are mapping out cross-modal associations with touch, smell, feel, and taste, leading to new and fascinating multimodal designs and experiences, as well as a veritable coral reef of artworks. Much remains unknown as to how sensory integration works. Perhaps intuitively, artists rely on the principle of stimulus and sensation to optimise designed experiences; this is the focus of the science of psychophysics. Throughout history, technical means were developed to extend human capacities in all sensory modalities, both receptive and expressive: recent enabling technologies include film, games, virtual reality, data perceptualisation (e.g. visualisation, sonification, physicalisation), and immersive, smart architecture. Contemporary art represents a stream of possibilities, yet beneath a surface of flamboyance, some artworks might have sprung from confident intuition rather than theoretically grounded and empirically validated research.

Sound artists and musicians have always been inclined towards intermediality (cf. ancient rites, dance, opera, cinema). The first computer artworks were musical compositions. In numerous fields of application, sound integrates with other media, often in a supportive yet essential capacity. Sound and music have provided fertile grounds for research in perceptual psychology and computer science. The fluency of sound, its invisible and intangible nature – and because it can only happen inside time – might be key to an understanding the human condition: brain, body and soul.

The art-science loop needs to be closed. There are many questions to which answers can only come through an interdisciplinary approach. For example: *Can designers profit from a deeper involvement in the perceptual sciences? Can researchers use art and design more convincingly as hypothesis-generators or test-beds for experiments? Can artists tell better stories by distinguishing between innate sensory interactions and culturally acquired yet subconscious interaction experiences? Can a line be drawn between normal and pathological synaesthesia'?*

3. Articles in the Issue

The first three articles are based on the invited keynote papers and the opening address at Si15 Soundislands Festival.

Vibeke Sorensen is Chair of the School of Art, Design, and Media (ADM), which hosted several of the events at Soundislands Festival. In her opening address, "Sound, Media Art, and the Metaverse", she sets the tone by reviewing synaesthesia as a concept to understand transmodal multimedia art and design forms. **Charles Spence** is head of the Crossmodal Research Laboratory at Oxford University. His article, "Sound Bites & Digital Seasoning", probes recent research involving the sense of taste, in particular how it is affected by concurrently listening to music. **Stefania Serafin**, director of the Multisensory Experience Lab which at Aalborg University Copenhagen. As outlined in her studio report, "Sonic Interactions in Multimodal Virtual Environments", the Lab is set up for research in audio-visual and haptic simulations.

After the keynotes, articles are organised under the headings of aesthetics, perception, technology, and sound art.

Aesthetics

The performing arts scene in Singapore has been growing strongly over the past decade. Art is generally seen as an important component in the development of a national identity, and scholars look both to local and external practices in trying to explain what is currently happening. **Joseph Tham's** contribution "Beyond Sound in Sound Art: Society and Politics in the Art of Yasunao Tone and Akio Suzuki", contextualises performance art in Japanese post-1950s society and its interdependency with Western sound artists. In general, the influence of Japanese sound art in Singapore has been substantial. From an equatorial vantage point, looking north is also key to the work by **Natalie Alexandra Tse** and **Andy Chia**. Their article "Performing Sounds Using Ethnic Chinese Instruments with Technology: SA" takes their own practice as a starting point to examine Singaporean contemporary art politics and socio-cultural perspectives.

Perception

Suzy Styles is director of the Brain, Language and Intersensory Perception Lab (BLIP Lab), in the division of Psychology at NTU in Singapore. In "The Language of Dance: Testing a Model of Cross-Modal Communication in the Performing Arts", she discusses how multisensory processing may lie at the heart of some forms of artistic expression, in the context of an investigation into crossmodal matching between spoken syllables such as 'bu' and 'ki' and physical expressions of sound in dance. Members and collaborators of the BLIP Lab also contributed reports testing perceptual linkages between sounds and other senses: **Nora Turoman** asks "How well do Humans Capture the Sounds of Speech in Writing?", and in her article describes their 'science in action' exhibition at Soundislands Festival, where visitors at the ArtScience Museum matched vowel sounds to graphic symbols in a playful way. **Joel Lim**, in "Super-Normal Integration of

Sound and Vision in Performance", delves into crossmodal correspondences involving 'guitar faces' people make when playing air-guitar while listening to music; this was demonstrated in an audience participatory event at the Museum. **Nan Shang** explores the associations between visual shapes and vowel sounds articulated in Mandarin Chinese through "An Implicit Association Test on Audio-Visual Cross-Modal Correspondences". **Shao-Min Hung** and **Po-Jang Hsieh** fro Duke NUS Graduate Medical School explore "Pre-conscious Automaticity of Sound-Shape Mapping", in an experiment testing early crossmodal binding between the sounds of written words and abstract shapes.

Two additional contributions round out the Proceedings' section on perception: **Rachel Chen**, presently based in France, contributes an article called "Colors in Silence – The Journey of a Synesthete in Understanding the Senses". It is a case-study probing specific crossmodal bindings between musical pitch, colour, and texture experienced by one individual. US-based **Yago de Quay** works with technologically enabled performance art. In "Same Time, Same Place, Keep it Simple, Repeat: Four Rules for Establishing Causality in Interactive Audio-Visual Performances", he outlines a framework to analyse how audiovisual and spatial cues may establish the illusory perception of a common cause.

Technology

The three articles in the technology section describe projects exploring new methods for artistic expression in performance and installation. A central topic in each case is how to develop strategies for parameter mapping in human-computer interaction. **Stefano Fasciani** is a researcher in audio technology, previously in Singapore and now in Dubai. In "Interactive Computation of Timbre Spaces for Sound Synthesis Control", he pursues an approach to voice-controlled synthesis based on machine-learning techniques. Fasciani regularly uses his system in stage performance, for example at Si13 (http://soundislands.com/si13). **Jingyin He** is originally from Singapore and now based in New Zealand. Together with **Jim Murphy, Ajay Kapur,** and **Dale A. Carnegie**, he writes about their ongoing practice-based research on "Parametrically-Dense Motion Sensing Devices and Robotic Musical Instruments". Finally in the section on technology, **Rafael Ramirez** and his team from Barcelona consisting of **Sergio Giraldo** and **Zacharias Vamvakousis** present their approach to "Brain-Computer Music Interface for Music Expression". Focussing on emotion and using only low-cost EEG equipment, Ramirez's group takes creative music performance as a springboard to develop systems with applications in therapy.

Sound art

The five articles on sound art are all based in the authors' own practice, which they performed and demonstrated at events at ADM, The Arts House, and ArtScience Museum. From Japan, **Yoichi Nagashima** describes in "Assembling Music" a series of pieces created by hacking and building electronics from scratch. His compatriots **Shumpei Tamura** and **Yasuo Kuhara** explain the construction of an inventive interactive installation piece in "Spandex Shoji Synthesizer Transforming Elastic Interaction into Images and Sounds". **Dirk Johan Stromberg** and **Robert Casteels** certainly win the trophy for most catchy paper title: "Having a Ball With the Sphere". It aptly describes a custom-built instrument for audiovisual performance, showcased in their recent multimedia performances in Singapore. Working as an audiovisual design duo, based in Melbourne, **Paul Fletcher** and **Mark Pollard** introduce their project on site-specific installations and performances in "Resonating Spaces". Finally, the article by **PerMagnus Lindborg** and **Joyce Beetuan Koh** describe their generative sound installation "When We Collide", presented at Soundislands Festival. The approach was collaborative, employing audio material by the authors and four commissioned composers: Dirk Stromberg, Stefano Fasciani, Seongah Shin, and Andrián Pertout.

4. Acknowledgements

Si15 Soundislands Festival was organised in close partnership with ArtScience Museum. The Si15 2nd International Symposium on Sound and Interactivity received core funding

from the Centre for Liberal Arts and Social Sciences (CLASS) at Nanyang Technological University (NTU). The events were supported by School of Art, Design (ADM), National Art Council, Bollywood Veggies, Italian Cultural Institute in Singapore, and New Zealand Art Council. We gratefully acknowledge peer support from Asia Computer Music Project and International Computer Music Association.

Keynote speakers were interaction design specialist Stefania Serafin and crossmodality researcher Charles Spence. Invited artists were audiovisual performer Ryoji Ikeda, film-maker and choreographer Daniel Belton, and electronica duos CLUBbleu (Felix Leuschner & Julia Mihàly) and Black Zenith (Brian O'Reilly & Darren Moore).

The core organisation team consisted of PerMagnus Lindborg, Suzy Styles, Joyce Beetuan Koh, and Stefano Fasciani. Invaluable contributions to producing concerts and installations were made by Ross Williams, Dirk Stromberg and Yong Rong Zhao.

Our institutional partners, represented by Honor Harger, Anna Salaman, Nina Ernst, Stacy Lim, Ivy Singh-Lim, Veronica Manson, Luke Kang Kwong Kapathy, Vibeke Sorensen, and Michael Walsh, made the events possible in the first place. Also, many thanks to NTU staff, including Lucas Jodogne, Poh Zhuang Yi, Hong Bee Kuen, Lim Pheng Yew, Muhammad Mustajab Bin Mohamad, Lau Kheng Hock, Shukor, Bharat Singh, Michelle Tan, and Magdalene Lim for their contributions, and many others at Nanyang Academy of Fine Arts, The Arts House, Bollywood Veggies, and ArtScience Museum.

Importantly, warm thanks to our students Liew Kongmeng and Nurul Syarfiquah, who together with Valentia Chew, Sandra Soh, Issac Ting, Wong Jinpei, and Woon Fei Ting, who provided much assistance in running the events.

We are indebted to Si15 session chairs Sara Lenzi, Stephen Lim, Naotoshi Osaka, Randall Packer, Mark Wong, and Lonce Wyse, who also acted as double-blind peer reviewers of the open call submissions; in this effort they were joined by Renick Bell, Anders Elowsson, Emma Frid, Nicolas Jaquot, Joleen Loh, Antonio Pessotti, Seongah Shin, and Lindsay Vickery. Thank you all.

Last but not least, we would like to express our gratitude to all the authors for their patience during the editing process and conscientious work in revising the manuscripts. As editors, we hope that, through the authors' efforts, the merits of each and every work in this Issue is plain to see, and that any editorial oversight will not cloud the overall impression.

We wish the readers a great time with the Special Issue of Array: Proceedings of Si15. We hope you will join us for the future installments of the Soundislands Festival and International Symposium on Sound and Interactivity: Si17 and beyond!

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 5-8.

Sound, Media Art, and the Metaverse

Vibeke Sorensen

School of Art, Design, and Media, Nanyang Technological University

vsorensen@ntu.edu.sg

Abstract

The author discusses the fundamental role of sound in global cultures, introduces related perceptual and scientific phenomena such as synesthesia, and provides an overview of their relationship to the development of transmodal multimedia art and design forms emerging at the intersection of physical and digital media, including wireless technologies, social networks, and ubiquitous computing. The article is an edited version of the introductory address given at the opening event of *Si15 Soundislands Festival*, on Wednesday 19 August 2015, at the Arts House, Singapore.

Keywords: synaesthesia, neuroscience, visual music, animation, multimedia, wearable technologies

Sound fills space and our hearts and minds. It expresses our deepest feelings and emotions, and is often used for important rituals and to set the mood for special occasions. It has been refined for use in theatre, films, multimedia and experience design for museums, shopping malls, restaurants, homes, and even our bodies with wearable technologies. It is immersive and migratory, extending into the air, literally and metaphorically. It is like architecture, or *liquid architecture,* as we are inside it. It is spatial and temporal, like walking through a building or swimming. And sound is mathematical, a physical phenomenon that we can measure in frequency, amplitude, and so forth. Today with malleable digital media, when all measurable phenomena can be digitized, sound can be translated into and across a wide range of sense-based media forms blending boundaries. Sound can be visualized and images become sounds, temperature become color, CO2 pitch (as in the case of my own projects with 'singing' plants), and more. The electromagnetic spectrum contains waves and data that our perceptual system did not evolve to perceive, but now we literally can listen to. New technologies give us access to realities we did not know existed, and more than that, the ability to perceive in new ways *other* realities so as to gain deeper understanding of, and empathy *for* others. For me, this is the most important outcome. Our music and art can bring heart to the mind.

One important element is *synaesthesia,* a well-known phenomenon wherein an impression in one sensory modality arouses impressions in another. It has been studied both in the sciences, such as psychology and neuroscience, and in art and design. It has been explored creatively by artists for centuries, including the celebrated Russian composer Alexander Scriabin (1871 – 1915) who scored for colored lights in his 1910 musical masterpiece, *Prometheus: The Poem of Fire.*

With technologies that were invented to converge the senses, such as sound-film projection and computer animation[1], whole new

---

[1] John Whitney, Sr (1917-1995) widely known as the "father of computer graphics" was an American experimental animator and composer whose vision of unifying music and moving images through the common foundation of wave theory led him to the first use of computers for moving images in the 1950s. See Whitney, John (1980). "Digital Harmony: On the Complementarity of Music and Visual Art", McGraw-Hill.

genres of experimental art were pioneered that led to huge new industries. This includes Oskar Fischinger (1900 – 1967), who gave 'visual music' its name and developed Disney's greatest masterpiece *Fantasia* (1940). Visual music pioneers led the way to "light shows" in the 1960s, and MTV in the 1980s. Today almost everyone has a music app with graphics running on their laptops as screen savers.

Sound technology in particular, including mechanical and electrical, often integrated multiple senses. The ideas informing it go back to Pythagoras of Samos (Greece, c. 570 – c. 495 BC) who described mathematical relationships between musical pitch and geometry, and span millennia that followed and include Sir Isaac Newton (1643 – 1727) who theorized correspondences between colors in light and notes in a musical scale. There is also the French mathematician Father Louis-Bertrand Castel (1688 – 1757) who in 1725 famously created the "ocular harpsichord," a color organ made of candles, colored glass, and moving curtains that whenever a key was pressed opened to reveal a colored light behind it. And there is Thomas Edison (1847 – 1931) who invented the phonograph and the movie camera. We can add many 20th century artists, composers, and researchers such as Max Matthews (1926 – 2011) who worked at Bell Laboratories (1950s - 1980s), IRCAM (1970s), and Stanford University (1985 - 2011). There have been other great researchers in academia, industry and companies like Yamaha in Japan, including people participating in this very festival.

Science, music, and art have been having an inspired conversation for thousands of years, striving to make the invisible visible and the inaudible audible, and provide us with a way of experiencing, making sense of, and exploring the unknown: that which we know exists but that we still do not understand. Artists, composers, and technologists are working at the edge of understanding by exploring and discovering our universe through the temporal materiality, and increasingly the immaterial rematerialized, of their art forms.

Today, the idea of connecting the senses, realizing the centuries if not millennia old dream of unifying the arts, or "gesamtkunstwerk"[2], sits at the foundation of media art. Sound is at the centre. Sound and music, as in opera, have always incorporated space, time, performance, and multiple modalities. Today, through the powerful vehicle of data and information systems, the potential is vastly expanded. It is extended further with the blending of scientific knowledge of how the brain works, or neuroscience, with our distributed, wireless, networked, and pervasive technologies that permeate our increasingly connected digital metaverse.

Signals from the brain and body activate nerve connections that elicit feelings and emotions, patterns of neural pathways that give rise to the formation and playback of memories. These patterns are constantly reconfiguring themselves as we move, creating new pathways and patterns that with each new experience mix memory and sensation into an ongoing stream of consciousness, and a continuously updated narrative that we engage and contemplate. We are in a continual process of creation, always predicting and inventing the future based on experience of the past and present, so as to be able to move and survive in a constantly changing world that surrounds us. How the brain creates a mental model of the world through sensory impressions that somehow conform to it, is still not understood. It is called the "binding problem."

Again, it is artists and composers who are shedding light on this problem through their creative work and artistic research with media. They are asking questions about representation and whether or not the models are "correct," how to work imaginatively and sensitively with them, and what those models could do to help us understand and anticipate the complex changing environment. Sound, computer music, and media art are merging with environmental and social science as an emerging form of *bio-social-art.* They are shedding light

---

[2] "Gesamtkunstwerk" is German for "total work of art," a term first coined by philosopher K. F. E. Trahndorff in 1827 and subsequently used by opera composer Richard Wagner in two essays written in 1849: "Art and Revolution", and "The Artwork of the Future", which further disseminated the term.

on how groups of people and other living things such as plants and animals communicate with each other and their environment, how they relate, and even how they cooperate to survive and evolve. [3]

Composers, sound and media artists have catalyzed the development of new technologies for communication because we exist in time and, as speakers in this festival including *Si15 Soundislands Festival* Chair PerMagnus Lindborg, Nora Turoman, Suzy Styles, Bani Haykal, and Daniel Belton reflect upon in their presentations, because our languages are temporal. In fact "artists-in-labs" is a proven strategy for technological innovation. Bell Laboratories, mentioned earlier, brought composers and artists into their sound research labs starting more than 50 years ago with the expectation that in creating their own pieces they would accelerate the development of tele-communications technologies. This same Research and Development (R&D) strategy was widely adopted in the 1970s - 2000s in the computer graphics and multimedia industry including NASA, Xerox PARC, Microsoft, and many others.

Language exists in time and reflects upon our existence individually and collectively. All of our senses influence each other and shape our thoughts. Narrative is the meaning we give to sequential sensory experience, the stored impressions that are the patterns of neural connections that we *re-member* to create that exquisite tapestry that is our on-going story. Aroma in particular stimulates memories and enhances perception in other sensory domains. This is being explored scientifically and used creatively in multimodal art. We are just now starting to understand how it works, how to develop it to improve life, and provide heightened experiences of greater sensitivity, enjoyment, awe, and hopefully also compassion.

We know this is possible from our daily experience of life, including the delicious food we eat and the exciting relationships being discovered between taste and sound, which *Si15* Keynote Speaker Charles Spence is exploring as the head of the Crossmodal Research Group at the University of Oxford. We also know it from our experience of exquisite rainbow colors observed in the refracted light of the sun, in ethereal textures that swirl around us in high dynamic range spatial sound, and in sumptuous physical and dreamlike virtual materials that we touch and play with esthetic refinement. We express our mood in food, music and clothing every day. With wearable technologies, we can correlate our feelings and expressions with data from our bodies. We can 'tune' our bodies to our internal and external environment through sound and music using sensors and biofeedback. This can be used for medical purposes, letting us know that when we feel good, it really is good. Evolutionary biologists tell us that we "like" certain kinds of images (and presumably also sounds) because in the past those things they represent helped us to survive. A template seems to be stored in our collective DNA. My view, then, is that we are moving increasingly from "natural" selection to "esthetic" selection.

There are so many exciting new works of art for us to experience, contemplate, and learn from that exist in time and have at their core, sound and music.

From an exploration of that which has come before (i.e. history) while embracing a fundamental respect for life, and the individual and the social in an approach that emphasizes the humanization of technology and a conception of networks not just as technological but "e-cological" (a network not just of things but of *living* things), to the mysteriously embodied and sentient environment (a network of the tangible and intangible), we always return to the question of representation and narrative. What are we saying and doing, and how do our ideas and actions affect and reflect the worlds we inhabit? We are in an era of big data, and each of us is unique. And like the "butterfly effect" in a huge orchestra, each person is playing a part. Each sound artist, composer, and multimodal researcher in this festival will

---

[3] The Second International Congress on Animal Computer Interaction was held in Johor, Malaysia in November 2015 as part of Advances in Computer Entertainment (ACE) and in association with ACM SIGGRAPH and ACM SIGCHI. See URL http://animalcomputerinteraction.org/index.html

shed light on their inspired and unique way of feeling and knowing, thinking and telling. We join them in exploring the miracle of the mind and mysteries of the universe. It is a wondrous journey.

We are grateful to all of the artists, and the organisers and hosts, especially the core organising team which includes PerMagnus Lindborg, Suzy Styles, Joyce Beetuan Koh and Stefano Fasciani for creating such an intriguing and relevant event. Special thanks to Andrián Pertout, Seongah Shin, Dirk Stromberg and Daniel Belton. I also want to join the hosts in thanking everyone who provided support. The School of Art, Design and Media (ADM) at Nanyang Technological University Singapore is proud to be a part of this festival. We are in our 10th year in 2015, which we are celebrating together with SG50. I wish to thank all of our participating faculty and staff, especially the ADM AV, IT and administrative team, for their hard work and contributions. ADM, our iconic heart shaped building is also a kind of island, rising from waves of green to the sky, and joined by vital bridges that you and this festival are making to the entire planet. We unite in the celebration of the flowering of the creative spirit here in Singapore, a vibrant and highly connected global media arts Nation, tonight, throughout this festival, and long into the future.

Thank you for your attention, and I wish everyone an enjoyable Festival.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015 pp 9-15.

Sound Bites & Digital Seasoning

Charles Spence [*]

[*]Head of the Crossmodal Research Laboratory, Department
of Experimental Psychology, University of Oxford, UK
charles.spence@psy.ox.ac.uk

Abstract

Although often considered as the forgotten flavour sense, what we hear – be it music, soundscape, or product sound – influences what we taste. For instance, loud noise has been shown to suppress our ability to taste sweetness, while enhancing the taste of umami. Furthermore, many studies have demonstrated sensation transference effects, whereby what we think about what we hear (and the ideas/concepts primed by such music or soundscapes), can be transferred to whatever we happen to be tasting. The emotions that can be induced by music can also influence the experience of taste. In this presentation, though, I want to take a closer look at the specific way in which what we hear influences what we taste: In particular, the focus will be on the latest research showing that the crossmodal correspondences between music and tastes, textures, aromas, and flavours can be systematically used to direct a listener's attention to certain elements within the tasting experience. I will demonstrate how chefs, sound designers, culinary artists, brands, and psychologists are becoming increasingly interested in modifying the taste of food and drink through sound – think of it as digital, or sonic, seasoning. I will also stress why contemporary gastronomy and sensory apps offer a rich opportunity both to advance our theoretical understanding in this area, and also to impact consumer behaviour more generally.

Keywords: multisensory, gastronomy, digital seasoning

1. Introduction

Although often considered as the forgotten flavour sense (Spence et al., 2011), what we hear – be it music (Spence & Wang, submitted b), soundscape (Knapton, 2015), or product sound (Spence, 2015a; Zampini & Spence, 2004, 2005) – influences what we taste. So, for instance, changing the sound we make when we crunch on a crisp or apple significantly changes perceived crispness and freshness (see Demattè et al., 2014; Zampini & Spence, 2004). Meanwhile, loud noise has been shown to suppress our ability to taste sweetness, while enhancing the taste of umami (Spence, 2014; Spence, Michel, & Smith, 2014a; Yan & Dando, 2015). Music and soundscapes can also induce a change in the mood or emotion of the listener (Juslin & Sloboda, 2010; Konečni, 2008; see also Crisinel & Spence, 2012b), and this too has also been shown to affect certain aspects of taste perception (see Spence, 2015b, for a review). Soundscapes such as the sound of the sea, as served at Heston Blumenthal's The Fat Duck restaurant in Bray have also been shown to enhance the pleasantness of a matching seafood dish while having no effect on perceived saltiness (see Spence et al., 2011).

Furthermore, many studies have demonstrated sensation transference effects (Cheskin, 1957). This is where what we think about what we hear (and the ideas/concepts primed by such music or soundscapes), can be transferred to whatever it is we happen to be tasting (e.g., North, 2012; Spence & Wang, submitted b). For instance, Adrian North has demonstrated how playing music that is heavy and powerful (think Carmina Burana by Carl Orff) can bring out the heavy and powerful notes in wine. In other words, there are a number of ways in which what we hear can influence what we think about what we taste (see North & Hargreaves, 2008; Spence et al., 2011, for reviews).

2. The crossmodal correspondences

In this presentation, though, I want to take a closer look at the role that crossmodal correspondences between music and tastes, textures, aromas, and flavours (e.g., Belkin et al., 2007; Crisinel & Spence, 2010, 2012a, b; Deroy et al., 2013; Knoeferle et al, 2015; Knöferle & Spence, 2012; Mesz et al., 2011; Spence, 2011, 2012) may play in systematically directing a listener's attention to certain elements within the tasting experience (Gray, 2007b; Spence & Wang, submitted b). Crossmodal correspondences refers to the initially-surprising connections that many of us appear to make between seemingly unrelated attributes, features, or dimensions of our sensory experience. So, for example, most people associate sweet tastes with round shapes and high-pitched sounds – think tinkling wind chime or piano sounds. By contrast, most people will associate bitter tastes with angular shapes, low-pitched sounds, and brass instruments. It is hard to point to where such associations may have come from, hence why they initially seem surprising.

In my talk, I will argue against the notion that synaesthesia provides a useful way of thinking about such connections between music and taste (Knapton, 2015; Rudmin & Cappelli, 1983; Sachse-Weinert, 2012, 2014; see also Spence & Wang, submitted a), seductive though it may be to do so. Importantly, the correspondences appear to be shared across large groups of people whereas the concurrents that are such a key feature of synaesthesia tend to be idiosyncratic (though, of course, the synaesthete is also subject to the crossmodal correspondences just like the rest of us). The correspondences may be learnt from the statistics of the environment or may reflect some more affective form of matching (Palmer et al., 2013; Parise et al., 2014).

A large body of empirical research shows that sweetness tends to be matched with sounds that are higher in pitch, with the sound of the piano, with music that is legato in articulation, and with consonant harmonies (Bronner et al., 2012; Mesz et al., 2011). By contrast, sourness tends to be matched with very high pitch sounds, fast tempo, and dissonant music instead (Bronner et al., 2012; Mesz et al., 2011).[1] Bitterness is matched with sounds that are lower in pitch and more likely to be brassy (e.g., Crisinel et al., 2012; Wang et al., submitted).

Further information concerning the crossmodal correspondences that have been documented in this area comes from the results of a series of experiments conducted by Crisinel and Spence (2010, 2012a). The participants in these studies had to pick a musical note (one of 13 sustained musical notes (from C2 (64.4 Hz) to C6 (1,046.5 Hz) in intervals of two tones), and pick a class of musical instrument (piano, strings, wind, and brass) to go with a variety of basic tastes, and with each of 20 of the key aromas (presented orthonasally) commonly found in wine (including almond, apple, apricot, blackberry, caramel, cedar, dark chocolate, cut hay, green pepper, honey, lemon, liquorice, mushroom, musk, pepper, pineapple, raspberry, smoked, vanilla, and violet).

The participants were seated in front of a virtual keyboard that allowed them to play each one of the 52 possible sounds (i.e., 4 instruments x 13 pitches) in order to find the best match. The results demonstrated that for a number of the tastes and aromas, the participants were consistent in terms of the notes and instruments that they felt went especially well together. So, for example, fruity notes such as apricot, blackberry, and raspberry were all matched with higher (rather than lower) musical notes, and with the sounds of the piano and often also woodwind instruments, rather than with brass or string instruments. In contrast, lower pitched musical notes were

---

[1] Back in 1855, Hector Berlioz suggested that the sound of the oboe had an 'acid-sweet voice'

associated with musky, woody, dark chocolate, and smoky aromas, bitter tastes, and brassy instruments.

Of course, just because stimuli are judged as matching does not, in-and-of-itself entail that playing the matching music, soundscape, or sound/chord would necessarily influence the tasting experience. However, that being said, the evidence that has been published to date does indeed suggest that such crossmodal effects are normally found when matching (or corresponding) stimuli are presented together. And what is more, the subjectively rated goodness of the match between what we hear and what we taste tends to correlate with how much we report enjoying what we are tasting (Spence et al., 2014b; Wang & Spence, in press b).

A growing body of empirical research now shows that our experience of many different food and drink products can be modified changing the music or soundscape that people listen to. To date, studies have been conducted with everything from cinder toffee through chocolate and fruit juice (Crisinel et al., 2012; Reinoso Carvalho et al., in press; Wang & Spence, in press a, submitted a), and from beer through wine, whisky, and vodka (Holt-Hansen, 1968, 1976; Rudmin & Cappelli, 1983; Spence et al., 2013, 2014c; Velasco et al., 2013; Wang & Spence, in press b, submitted b).

By far the most research has, though been done on the crossmodal matching of music with wine (see Spence & Wang, submitted a, for a review). A growing body of empirical research now shows that sweetness, acidity, fruitiness, astringency, and length of the flavour sensation can all be modified by playing the appropriate musical selections (see Spence & Wang, submitted b, for a review). Hence, while some have been sceptical concerning music's ability to influence taste (just see Jones, 2012), there is now enough evidence to demonstrate just what a profound effect it really has.

Get the combination right and it is possible to deliver an experience like the following from James John, Director of the Bath Wine School, when Mozart's *Laudate dominum* is combined with Chardonnay: *"[...] Just as the sonant complexity is doubled, the gustatory effects of ripe fruit on toasted vanilla explode on the palate and the appreciation of both is taken to an entirely new level"* (quoted in Sachse-Wienert, 2012).

What is particularly interesting about the crossmodal influence of music and soundscape on taste and flavour perception is that the effects often appear to occur more-or-less instantaneously (Crawshaw, 2012; Spence & Wang, submitted c). No sooner has the music changed from major to minor say than the taste of the wine also changes.

My suspicion about what may be going on here is that the crossmodal correspondences between audition and taste/flavour serve to direct the listener's attention to one aspect of what can be a complex tasting experience. Just take the following quote from one sceptic describing his experience on finding that a change in music (by Clark Smith, a Californian wine maker and wine consultant) changes the perception of wine "*What seemed to be happening was not that we noticed new flavors when the music changed. Instead, the same flavor elements were there all along, but the music seemed to change the way we perceived them. Some music made us pay more attention to astringency, so we disliked the wine. With other music, we chose to ignore the oak and tannin, so we liked it more."* (Gray, 2007). From everything I have seen and experienced I would certainly wish to reiterate that point about sound merely accentuating, or suppressing, certain features that are already there in the tasting experience, rather than creating new tastes/flavours out of nowhere.

Now as to whether this crossmodal attention effect happens automatically, that is, in a stimulus-driven manner, or whether instead it is a voluntary (i.e., effortful) process of matching is not yet known. Should the former be the case then, of course, incidental music playing in the background might be expected to influence the taste of whatever we are drinking. If, however, the crossmodal matching is an effortful process then, perhaps, the music / soundscape will only influence our taste perception if we actively direct our attention to the music and consider the possible links to certain aspects of the tasting experience.

To date, most of the multisensory tasting events that have demonstrated an effect of music or soundscape on tasting have involved the participants/attendees being actively encouraged to try and make the link between their senses. Finding out whether this is necessary for these kinds of crossmodal effects to occur is certainly going to be an important question for future research to address (see Spence, 2015a). One final caveat to note here is that while we now know that music and soundscapes can change taste perception in the short term, no one has yet looked at how long-lasting the effects of 'sweet music' on taste perception may last.

It is perhaps unsurprising given what we have seen so far why a growing number of chefs, sound designers, culinary artists, brands, psychologists and even a few philosophers are all becoming increasingly interested in modifying the taste of food and drink by means of sound (e.g., Crisinel et al., 2012, 2013; Spence, 2014; Spence & Piqueras-Fiszman, 2014; Spence, Shankar, & Blumenthal, 2011; Spence & Wang, submitted c). Even British Airways launched a sonic seasoning soundtrack on their long haul flights last year, the idea being that those who were dining in the air could choose to listen in to music to match the taste of the food that they had ordered. Other chefs, like Grant Achatz of Chicago's Alinea fame is now starting to consider having a musician come into the restaurant to play something to accompany one of the dishes on his fabulous menu (see Ulla, 2011).

While much of the research in this area to date has utilized pre-existing music, this is not always ideal. Music tends to evolve over time, while the matching tasting experience may stay pretty much the same. Hence, it should come as no surprise that a growing number of composers and designers are now starting to create music and soundscapes especially to match a particular taste, aroma, or flavour (see Bronner et al., 2012; Crisinel et al., 2012, 2013; Knoeferle et al., 2015; Mesz et al., 2012). Ben Houge, the sound artist has made the sensible suggestion here that what one might actually want is something like the music in video games that stays forever the same until such time as the player gets to the next level when the music then evolves to match the. We have recently taken to testing the various musical solutions that have been developed to see which people find it easiest to match with specific tastes (Wang et al., submitted). What is more, we have also been able to demonstrate the cross-cultural meaning of certain of these musical selections (Knoeferle et al., 2015). So, for example, we have been able to demonstrate that music composed in Germany to match each of the four basic tastes (e.g., bitter, sweet, salty, and sour) can be decoded by participants in India almost as well as by those in Europe/North America. Excitingly, internet-based testing, as well as large-scale citizen science experiments are allowing us to collect large amounts of data in a very short space of time (see Woods et al., 2015, on the strengths and weaknesses of the internet-based testing approach). Such an approach can work well for those tastes/flavours that people are already familiar with.

Now while the majority of the work on crossmodal correspondences has focused on instrumental music/soundscapes, there is an interesting area to explore, moving forward, in terms of crossmodal correspondences between vocal attributes on the one hand and tastes, aromas, and flavours on the other (Simner et al., 2010).

Contemporary gastronomy and sensory apps (Crisinel et al., 2013; Jones, 2014; Spence, 2014; Spence & Wang, submitted, c) both offer a rich opportunity to advance our theoretical understanding in this area, and also impact consumer behaviour, more generally. I will discuss a couple of sensory apps that have been designed to match flavours/aromas with music that highlight the two approaches here, both the scientific and the more personal. The scientific approach is illustrated by the *Le Nez de Courvoisier* app (URL http://courvoisier.com/uk/le-nez-de-courvoisier-app/ [accessed 29th July, 2013]; Crisinel et al., 2013), and the more personal approach by the Krug Music Pairing app (https://www.krug.com/; Jones, 2014).

3. Conclusions

In conclusion, I firmly believe that the influence of what we hear on what we taste has the opportunity to transform our dining/drinking experiences in the years to come. There is also evidence to suggest that by combining audition with the flavour senses in the right manner, some truly extraordinary experiences can result (Knapton, 2015; Spence & Wang, submitted c). What is more, I see far more innovation coming out of the chefs and food/drinks brands than out of big business or anywhere else, hence making the rate of progress here much more pronounced than elsewhere in the field of multisensory experience design (see Spence et al., 2013; Velasco et al., 2013).

References

Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality. *Psychological Science*, 8, 340-342.

Berlioz, H. (1855). *A treatise on modern instrumentation and orchestration, dedicated to Frederick William IV, King of Prussia*. London, UK: Novello.

Bronner, K., Bruhn, H., Hirt, R., & Piper, D. (2012). "What is the sound of citrus? Research on the correspondences between the perception of sound and flavour." In *Proceedings of the 12th International Conference of Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, 23-28 July, 2012 Thessaloniki, Greece. URL http://icmpc-escom2012.web.auth.gr/sites/default/files/papers/142_Proc.pdf

Cheskin, L. (1957). *How to predict what people will buy*. New York, NY: Liveright.

Crawshaw, A. (2012). "*How musical emotion may provide clues for understanding the observed impact of music on gustatory and olfactory perception in the context of wine-tasting*." Unpublished manuscript.

Crisinel, A.-S., Cosser, S., King, S., Jones, R., Petrie, J., & Spence, C. (2012). "A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background." *Food Quality and Preference*, 24, 201-204.

Crisinel, A.-S., Jacquier, C., Deroy, O., & Spence, C. (2013). "Composing with cross-modal correspondences: Music and smells in concert." *Chemosensory Perception*, 6, 45-52.

Crisinel, A.-S., & Spence, C. (2010). As bitter as a trombone: Synesthetic correspondences in non-synesthetes between tastes and flavors and musical instruments and notes. *Attention, Perception, & Psychophysics*, 72, 1994-2002.

Crisinel, A.-S., & Spence, C. (2012a). A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses*, 37, 151-158.

Crisinel, A.-S., & Spence, C. (2012b). "The impact of pleasantness ratings on crossmodal associations between food samples and musical notes." *Food Quality and Preference*, 24, 136-140.

Demattè, M. L., Pojer, N., Endrizzi, I., Corollaro, M. L., Betta, E. Aprea, E., Charles, M., Biasioli, F., Zampini, M., & Gasperi, F. (2014). "Effects of the sound of the bite on apple perceived crispness and hardness." *Food Quality and Preference*, 38, 58-64.

Deroy, O., Crisinel, A.-S., & Spence, C. (2013). "Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes." *Psychonomic Bulletin & Review*, 20, 878-896.

Gray, W. B. (2007). "'Enter Sandman' with Cab? Road testing Smith's theories." *SFGate*, November 2. URL http://www.sfgate.com/wine/article/Enter-Sandman-with-Cab-Road-testing-Smith-s-3235605.php on 13/08/2015.

Holt-Hansen, K. (1968). "Taste and pitch." *Perceptual and Motor Skills*, 27, 59-68.

Holt-Hansen, K. (1976). "Extraordinary experiences during cross-modal perception." *Perceptual and Motor Skills*, 43, 1023-1027.

Jones, A. (2012). "A far from vintage show as Callow uncorks the clichés." *The Independent*, 1st November, 51.

Jones, S. (2014). "Krug uses music to communicate Champagne taste digitally." *Luxury Daily*, August 8th. URL http://www.luxurydaily.com/krug-uses-music-to-explore-differences-between-champagne-varieties/ on 09/08/2015.

Juslin, P. N., & Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications*. Oxford, UK: Oxford University Press.

Knapton, S. (2015). "Why sparkling wine sounds like beans falling on a plastic tray." *The Daily Telegraph*, May 2nd, 7.

Knöferle, K. M., & Spence, C. (2012). "Crossmodal correspondences between sounds and tastes." *Psychonomic Bulletin & Review*, 19, 992-1006.

Knöferle, K. M., Woods, A., Käppler, F., & Spence, C. (2015). "That sounds sweet: Using crossmodal correspondences to communicate gustatory attributes." *Psychology & Marketing*, 32, 107-120.

Konečni, V. J. (2008). "Does music induce emotion? A theoretical and methodological analysis." *Psychology of Aesthetics, Creativity, and the Arts*, 2, 115-129.

Mesz, B., Sigman, M., & Trevisan, M. A. (2012). "A composition algorithm based on crossmodal taste-music correspondences." *Frontiers in Human Neuroscience*, 6:71, 1-6.

Mesz, B., Trevisan, M., & Sigman, M. (2011). "The taste of music." *Perception*, 40, 209-219.

North, A. C. (2012). "The effect of background music on the taste of wine." *British Journal of Psychology*, 103, 293-301.

North, A., & Hargreaves, D. (2008). *The social and applied psychology of music*. Oxford, UK: Oxford University Press.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). "Music-color associations are mediated by emotion." *Proceedings of the National Academy of Sciences of the USA*, 110, 8836-8841.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). "Natural auditory scene statistics shapes human spatial hearing." *Proceedings of the National Academy of Sciences of the USA*, 111, 6104-6108.

Reinoso Carvalho, F., Van Ee, R., Rychtarikova, M., Touhafi, A., Steenhaut, K., Persoone, D., & Spence, C. (in press). "Using sound-taste correspondences to enhance the subjective value of tasting experiences." *Frontiers in Perception Science*.

Rudmin, F., & Cappelli, M. (1983). "Tone-taste synesthesia: A replication." *Perceptual & Motor Skills*, 56, 118.

Sachse-Weinert, M. (2012). *Wine & musik: 2+2 = 5* [Wine & music: 2 + 2 =5]. Vortrag im Rahmen der Ringvorlesung "Weinwissenschaft" an der Johannes Gutenberg-Universität Mainz im Sommersemester. Presentation given on 4th July.

Sachse-Weinert, M. (2014). "Voce:divino in München. Synästhesie in perfektion. [Voce:divino in Munich: Pure synaesthesia]." *Weinfeder Journal*, 43 (July), 34-35.

Simner, J., Cuskley, C., & Kirby, S. (2010). "What sound does that taste? Cross-modal mapping across gustation and audition." *Perception*, 39, 553-569.

Spence, C. (2011). "Crossmodal correspondences: A tutorial review." *Attention, Perception, & Psychophysics*, 73, 971-995.

Spence, C. (2012). "Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism." *Journal of Consumer Psychology*, 22, 37-54.

Spence, C. (2014). "Multisensory advertising & design." In B. Flath & E. Klein (Eds.), *Advertising and design. Interdisciplinary perspectives on a cultural field* (pp. 15-27). Bielefeld: Verlag.

Spence, C. (2015a). "Eating with our ears: Assessing the importance of the sounds of consumption to our perception and enjoyment of multisensory flavour experiences." *Flavour*, 4:3.

Spence, C. (2015b). "Music from the kitchen." *Flavour*, 4:25.

Spence, C., & Deroy, O. (2013). "On why music changes what (we think) we taste." *i-Perception*, 4, 137-140.

Spence, C., Michel, C., & Smith, B (2014a). "Airplane noise and the taste of umami." *Flavour*, 3, 2.

Spence, C., & Piqueras-Fiszman, B. (2014). *The perfect meal: The multisensory science of food and dining*. Oxford, UK: Wiley-Blackwell.

Spence, C., Richards, L., Kjellin, E., Huhnt, A.-M., Daskal, V., Scheybeler, A., Velasco, C., & Deroy, O. (2013). "Looking for crossmodal correspondences between classical music & fine wine." *Flavour*, 2:29.

Spence, C., Shankar, M. U., & Blumenthal, H. (2011). 'Sound bites': Auditory contributions to the perception and consumption of food and drink. In F. Bacci & D. Melcher (Eds.), *Art and the senses* (pp. 207-238). Oxford, UK: Oxford University Press.

Spence, C., Velasco, C., Vanne, M., & Hopia, A. (2014b). "Can you taste the music?" In A. Hopia & S. Ihanus (Eds.), *5D cookbook* (73). URL www.5Dcookbook.fi Seinäjoki, Finland: KUMURU-project.

Spence, C., Velasco, C., & Knoeferle, K. (2014c). "A large sample study on the influence of the multisensory environment on the wine drinking experience." *Flavour*, 3:8.

Spence, C., & Wang, Q. (J.). (submitted a). "Wine & music (I): On the crossmodal matching of wine & music." *Flavour*.

Spence, C., & Wang, Q. (J.). (submitted b). "Wine & music (II): Can you taste the music? Modulating the experience of wine through music and sound." *Flavour*.

Spence, C., & Wang, Q. (J.). (submitted c). "Wine & music (III): So what if music influences taste?" *Flavour*.

Ulla, G. (2011). "Grant Achatz plans to 'overhaul the experience' at Alinea." [Online], Available: http://eater.com/archives/2011/11/23/grant-achatz-planning-major-changes-at-alinea.php#more. [2 November 2013].

Velasco, C., Jones, R., King, S., & Spence, C. (2013). Assessing the influence of the multisensory environment on the whisky drinking experience. *Flavour*, 2:23.

Wang, Q. (J.) & Spence, C. (in press a). "Striking a sour note": Assessing the influence of consonant and dissonant music on taste perception. *Multisensory Research*.

Wang, Q. (J.), & Spence, C. (in press b). "Assessing the effect of musical congruency on wine tasting in a live performance setting." *i-Perception*.

Wang, Q. (J.), & Spence, C. (submitted a). "Unravelling the bittersweet symphony: Assessing the influence of crossmodally congruent soundtracks on bitter and sweet taste evaluation of taste solutions and complex foods." *Food Quality & Preference*.

Wang, Q. (J.), & Spence, C. (submitted b). Assessing the influence of the multisensory atmosphere on the taste of vodka. *Beverages.*

Wang, Q. (J.), Woods, A., & Spence, C. (submitted). ""What's your taste in music?" A comparison of the effectiveness of various soundscapes in evoking specific tastes." *i-Perception*.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). "Conducting perception research over the internet: A tutorial review." *PeerJ*, 3:e1058; DOI 10.7717/peerj.1058.

Yan, K. S., & Dando, R. (2015). "A crossmodal role for audition in taste perception." *Journal of Experimental Psychology: Human Perception & Performance*, 41, 590-596.

Zampini, M., & Spence, C. (2004). "The role of auditory cues in modulating the perceived crispness and staleness of potato chips." *Journal of Sensory Science*, 19, 347-363.

Zampini, M., & Spence, C. (2005). "Modifying the multisensory perception of a carbonated beverage using auditory cues." *Food Quality and Preference*, 16, 632-641.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 16-18

Sonic Interactions in Multimodal Virtual Environments

Stefania Serafin

*Department of Architecture, Design and Media Technology

sts@create.auu.dk

Abstract

In this paper we present the research and teaching activities performed in the Multisensory Experience Lab at Aalborg University Copenhagen. We also briefly describe the Sound and Music Computing Education started at Aalborg University in September 2014.

Keywords: sonic interaction design, multisensory interaction, sound and music computing

1. Introduction

Sonic interaction design is an emerging field recently defined as the study and exploitation of sound as one of the principal channels conveying information, meaning, and aesthetic/emotional qualities in interactive contexts. This field lies at the intersection of interaction design and sound and music computing (Franinovic and Serafin, 2013). At Aalborg University Copenhagen we examine sonic interaction from different angles.

Since we live in a multi sensorial world, sonic interactions become more meaningful when combined with simulations of other modalities. It is well known that sound can indeed complement, enhance or even substitute other senses.

Our research explores sonic interactions when combined with other senses such as haptic and visual feedback. This includes the physics based simulation of multimodal interactions together with evaluation of the user experience.

Recent applications have been focused on the field of virtual reality, where the focus on auditory feedback has been rather limited when compared, for example, to the focus placed on visual feedback or even on haptic feedback. Other applications have been in the field of cultural heritage, in order to use sonic interaction technologies to reconstruct and preserve musical instruments. In particular, thanks to a project supported by the Culture 2000 EU framework, we reconstructed and exhibited the devices and the music of the Rai Studio di Fonologia Musicale in Milan (see Novati and Dack, 2012).

2. MultiSensory Experience Lab

In the MultiSensory Experience Lab (see **Figure 1**) we research the integration of different senses by combining technology, computer simulations and user experience evaluation. The lab consists of three main spaces.

The larger space is used for multimodal (audio-visual-haptic) simulations, and contains a motion capture system (16 cameras motion Optitrack system by NaturalPoint), a nVisor SX head mounted display and Oculus head mounted display, a 24 channels surround sound system (Dynaudio BM5A), and several devices for haptic feedback. In addition, the lab contains an anechoic chamber and a 64 speaker wavefield synthesis system (see **Figure 2**).

Figure 1. The Multisensory experience lab.

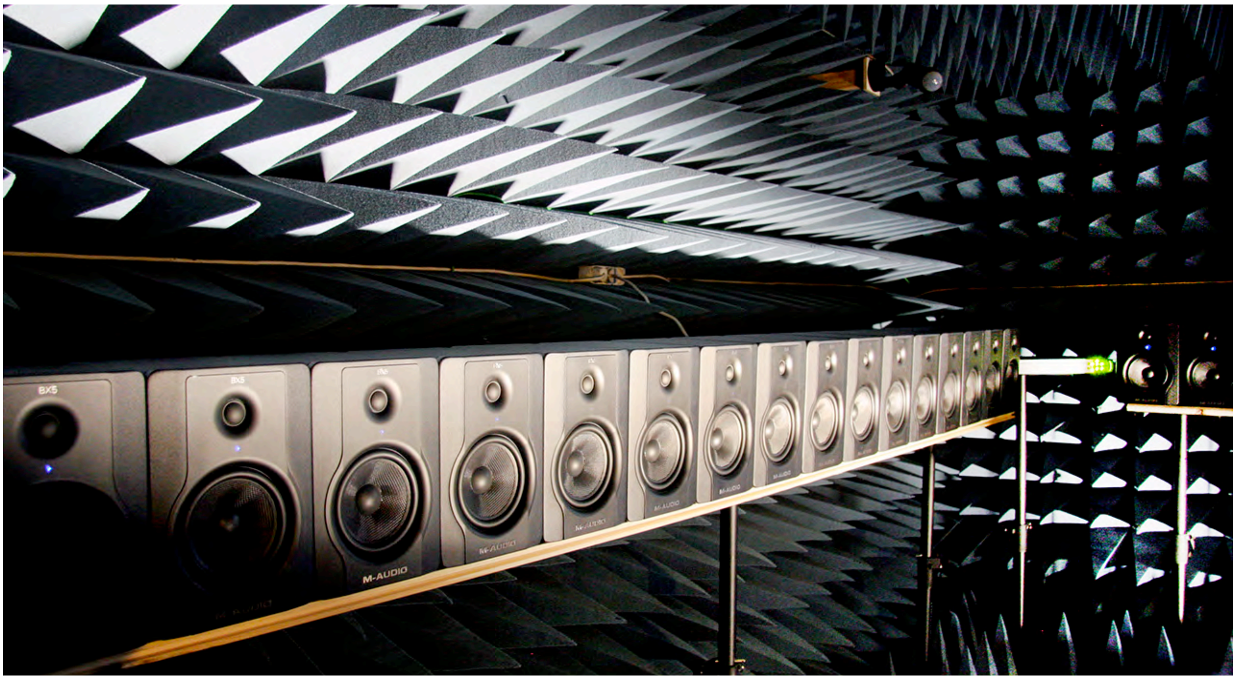


Figure 2. The wavefield synthesis lab.

3. The Sound and Music Computing education

From September 1st, 2014, Aalborg University in Copenhagen will offer a Master of Science in Sound and Music Computing. The Master of Science is a 2-year, research- based, full-time study programme, set to 120 ECTS credits. Its mission is to train the next generation of professionals to push forward the sound and music technologies of the new information society. By combining practical and theoretical approaches in topics such as computational modeling, audio engineering, perception, cognition, and interactive systems, the program gives the scientific and technological background needed to start a research or professional career. This program trains students on the technologies for the analysis, description, synthesis, transformation and production of sound and music, and on the technologies and processes that support sound and music creation.

More information can be found at URL http://media.aau.dk/smc

References

K.Franinovic and S.Serafin (2013). *Sonic Interation Design*. Mit Press.

M. M. Novati and J. Dack, (2012). *The Studio Di Fonologia: a musical journey 1954-1983, update 2008-2012*. Ricordi.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015 pp 19-26

Beyond Sound in Sound Art: Society and Politics in the Art of Yasunao Tone and Akio Suzuki

Joseph Tham Chin Pang

Crescent Girls' School, Ministry Of Education, Singapore

tham_chin_pang@crescent.edu.sg

Abstract

The history of sound art has been a case of displacement and misplacement. It has been displaced due to the nature of the medium – sound, a phenomenon that is heard and felt but not seen. In a visually dominated world like ours, this spells the relegation of the importance of sound and audio sensitivity in a human's perception, cognition and consciousness. It is misplaced as it is more often categorically subsumed under the other art form, music. Sound art is in fact one of the most ancient of art forms when the name of Greek god of wind, Aeolus, was used to name the Aeolian harp. Sound art has often been defined, in the late 20th century and early 21st century and by academics and critics, as a novel form of art. Once again it is another case of displacement and misplacement. The two case studies which will be discussed in this paper will foreground the social and political contexts of the artists as well as their responses to these socio-political conditions to demonstrate that sound art, just like any other art forms like paintings, music and theatre, is reflexive of the times of the creators, and more.

Keywords: sound art, Japan, politics

1. Sound Art

"Machines are designed not to make mistakes, in our behaviour we often make mistakes, so why not machines also. I add that reality to it. So it's not destruction but an addition."[1]

The history of sound art has been a case of displacement and misplacement. It has been displaced due to the nature of the medium – sound, a phenomenon that is heard and felt but not seen. In a visually dominated world like ours, this spelt the relegation of the importance of sound and audio sensitivity in a human's perception, cognition and consciousness. Sound art is misplaced as it is more often categorically subsumed under the other art form, music: it is either considered as modern composition, noise, free improvisation or electronica. Sound art is in fact one of the most ancient of art forms: the name of the Greek god of wind, Aeolus, was used to name the Aeolian harp. It is a wooden instrument well known in the ancient world in the West that is played not by a human agent but instead, via the force of nature, wind. The first instance of sound art is thus both a tribute to the powers of the supernatural as well as the revelling of nature.

Sound art has often been defined, in the late 20th century and early 21st century and by academics and critics, as a novel form of art[2]. Once again it is another case of displacement

---

[1] Yasunao Tone: sound artist, *Fluxus* artist, composer, and all-round provocateur.

[2] For more please see: Caleb Kelly (ed.) (2011), *Sound,* London/Cambridge, MA: Whitechapel Gallery/The MIT Press; Brandon LaBelle (2010), *Background Noise Perspectives on Sound Art*, London: Continuum Books; and Seth Kin-Cohen (2009), *In the Blink of an Ear – Toward a Non-cochlear Sonic Art*, London: Continuum Books.

and misplacement in the histories of sound art. For the former, sound art is posited as the use of sound to continue the pathway opened up by conceptual artists since the 1960s, albeit focusing on the sonic canvas, the unseen vibrations in the air[3], and basically its physical properties[4]. For the latter, it is usually described as apolitical or asocial as the emphasis is always on the sonic properties and its aesthetics and only that[5]. Traditional art criticism and art historical writing of the symbiotic relationships between artwork and the social and political milieu that germinate them are almost always left out. If any of its social and political landscape is mentioned, it is usually tacked conveniently under the general biographies of the artists without further studies of the artists' work with that of their socio-economic context.

The two case studies which will be discussed in this paper will foreground the social and political contexts of the Akio Suzuki and Yasunao Tone, as well as their responses to these socio-political conditions to demonstrate that sound art, just like any other art forms like paintings, music and theatre, is reflexive of the times of their creators. The two sound artists are from Japan, a country which due to its unique post-World War II history as well as its past before the Pacific War, but they exemplify two different takes on their socio-economic and political settings since the 1960s. Both artists were born and brought up during the most tumultuous period of time in 20th century Japan. Both grew up in the strangely schizophrenic era of initial post-War humiliation and extreme poverty and then, taking an abrupt about-turn swiftly to become one of the most powerful nations, economically, within a short span of ten years (late 1940s to late 1950s). The military antagonisms against the USA during the War from 1941 to the devastation of two Japanese cities from the deadliest man-made weapons ever, the atomic bombs in 1945, left a permanent mark in the national psyche of the Japanese people. The Americans then, due to Cold War politics, decided to reinstall Japan as quickly as possible as it was viewed strategically as a viable ally in the containment of the Communist states in Asia. This led to the subsequent sponsoring of the defeated nation economy to boost its drained capacity to that of global status by the 1960s[6]. The tension produced by such rapid and fortuitous turn-of-events for Japan was viewed apprehensively by many Japanese who experienced these rapidly changing developments.

On one hand the influx of Euro-American ideas and products fed the hunger of the Japanese to seek actively for something else to replace the failed militaristic nationalism of the previous decades, derived culturally and historically. Many saw the opportunities opened up in the post-war years as a great avenue to change Japan fundamentally from its conservative and conformist traditions[7]. On the

---

[3] In the 1960s and 1970s, sound art did not exist as a form as least in name; many examples cited by anthologies and history texts reveal most as sideway explorations from the then mainstream context of modern composition, so composers/artists like La Monte Young, Tony Conrad, Alvin Lucier, Robert Ashley, Charlotte Moorman/Nam June Paik and many other who worked with them or within their milieu are perceived as early exemplars of proto or pre-sound art practitioners. At the end, the notion of music or intermedia or Fluxus became the convenient categories which many sound art works of the artists were positioned then and retrospectively. See Alan Licht (2007), *Sound Art: Beyond Music, Between Categories*. New York: Rizzoli.

[4] Alvin Lucier's "I Am Sitting In A Room" (1969) and Robert Ashley's "Wolfman" (1964) are two classical pieces which are often cited in accounts and books.

[5] Please refer to: Brandon LaBelle & Steve Roden (eds.) (1999), *Site of Sound: Of Architecture & the Ear*, Los Angeles/Santa Monica: Errant Bodies Press/Smart Art Press; Christof Migone (2012), *Sonic Somatic*, Los Angeles: Errant Bodies Press and footnote 1. The relative avalanche of literature on sound art and its history since the late 1990s gave the false impression that sound art was an outgrowth of conceptual art or a new form of art to be documented and curated in art spaces.

[6] Julian Cope (2007), *Japrocksampler*, London: Bloomsbury, pp.23-40.

[7] Ibid., pp.41-72.

other hand, many also questioned the rampant consumerist-driven materialism brought in by the Americans as threatening the values and morals of the people. Their criticism also extended to the rampant consumerism in Japan. This dichotomy thus presented to artists, writers, and the two artists with the necessary fuel to channel their critique on the politics and society of Japan.

One interesting practice which is discernible in many of the Japanese art collectives and artists emerging in the 1950s and 1960s demonstrated clear debt to their Japanese culture – the emphasis on simplicity, interconnectedness of things and non-permanence which one can attribute to certain sensibilities of Japanese aesthetics but to Zen Buddhism and Shintoism as well. Key groups like Gutai, *Fluxus* in Japan and Mono-ha are just some of the more historically well-known examples[8].

The two following case studies will highlight the significance of such influences and traits commonly found in Japanese works and practices as well as pointing out how these artists channel some of these sensibilities through their choice of art medium – sound.

2. Back to Nature: A case study of Akio Suzuki

Born in 1941 in the then Japanese colony of Korea, Akio Suzuki, was repatriated back to Japan with his parents after the War in 1945. Due to the cultural import of Euro-American influences, Suzuki was exposed to the ideas of iconoclasts like John Cage and David Tudor and their then-revolutionary introduction of chance and conceptual bent to the arts[9]. In 1963, Suzuki was working in an architect's office and he became fascinated by an idea. He was examining the possibility of making a staircase that would not tire out the users but in fact would become a pleasure to them. Then it occurred to him that he could conceive of a staircase that would look like a musical stave and if he would to drop a ping pong ball down it, the sounds created should correspondingly be beautiful. This spurred him on to start on his first actions, which Suzuki coined Self Study Events[10]. He went to a train station in Nagoya and he tipped a dustbin down a staircase. The sounds created was horrible and he was in fact, looked at with disdain by the public as many of them picked up the rubbish which was scattered all over the base of the staircase.

It was perceived by other Japanese as an act of public nuisance and was considered inappropriate social behaviour. Undeterred, he threw down another bin, and was arrested by the police[11]. Suzuki's first work challenged norms and provoked thought in the parts of Japanese society that are particular with etiquette.

Similar to many Happenings which were taking place in the major cities of Japan due to the absorption of such creative but provocative acts critiquing the societies in the West, the authority and the crass consumerism many of these artists witnessed around them, Suzuki found the experiments (like the one described above) to be unsatisfactory, and he decided to conduct his own research to help boost himself intellectually and to help him understand the world around him better. This was, he hoped, to provide ideas for him to come up with more meaningful and more impactful artistic endeavours in the post-War years. He subsequently developed Throwing and Following. He would project sound into a space and he would follow the consequential resonance of it in the time and space. This opened him up to a better understanding of the objects around him and in turn he could use them more effectively to generate sounds in spaces[12]. He thus gradually grasped the interactivity between material, sonic properties of objects, and the resonances and echoes produced. This laid the foundation for his mature works in the 1970s.

In the 1970s Suzuki hand crafted one of his most signature sound art tools, the *Analapos*. It is built from two cylindrical resonators connected by long springs. The instrument ex-

---

[8] For more details see Alexandra Munroe (1994), *Japanese Art After 1945 – Scream Against The Sky*, New York: Harry N. Abrams.

[9] David Toop, "Akio Suzuki – Acoustic Trickster" in *The Wire* issue 231 May 2003, p.12.

[10] Ibid.

[11] Ibid.

[12] Ibid.

plores the reverberation of springs, the resonance of vessels and the transmission of signals between two physically linked terminals. The tool also reacts differently from space to space as the reverberation and resonances generated vary from one venue to another. The simplicity of the tools belies a few principal ideals of Suzuki: his subtle commentary on the hyper-modernisation of the Japanese society and the growing estrangement between individuals in such a materialist-driven, dog-eat-dog environment with the overemphasis on science, technology and human agency. Just like most of his sound art pieces/performance works, Suzuki wants to reclaim the human-ness of art by consciously interweaving human agency with the implements/instruments. He also wants to alert his audience/listeners to pay attention to their surroundings once again. He has since the 1970s increasingly performed in open or natural settings, inviting the audiences to join him in such back-to-nature creation of his sound art. This can perhaps be linked back to the Japanese religious and aesthetic reverence of nature but can also be interpreted as a veiled rejection of the modernist project of relentless human progression in its scientific and technological drive since the 19th century.

Coming back to the *Analapos*, its use in the wide-open natural context can reap surprising results as the tool reacts in indeterminate manners with the winds, the humidity, the openness and altitude of the space as well as the resonating human bodies in its vicinity[13]. No single event can be repeated. This reflects prominently the unpredictability of nature and the state of our world.

Suzuki also created site-specific events that required the audience to travel to a space close to his current residence in Tango, Japan, to immerse themselves in the piece with him, his installation space and the elements. Suzuki calls the venue his dream space and he constructed and coined it *Space in the Sun* (1988). The venue is found on a mountaintop in rural Tango under the direct purview of the sun. The site consists of two crumbling and identical 17 meters by 3.5 meters facing walls. The walls were built from 20,000 handmade sun-baked loam bricks. They are separated by a 7 meters floor space constructed from the same materials as the walls[14]. There are many "imperfections" found at the site such as the almost ubiquitous cowpats left there by the herds that visited the site daily. The howling winds that surge through and around the walls can be deafening. Once in a while, one can hear the odd birdcall, a lowing cow or the distant farm machinery. As one strolls through the site, the sounds come together to suggest a confluence of nature and man-make phenomenon around us even in such rural settings as Tango[15]. Akio Suzuki's intentions are thus not so much of a rejection of human agency and modernisation in a simplistic fashion but instead Suzuki is presenting to us a conundrum of the nature-versus-human essentialism of man. With regards to Suzuki's nature-based works, his good friend and writer in Japan, Shin Nakagawa, wrote, "The act of listening to Nature while at the same time destroying Nature forms an interesting contrast[16]." For one to reject human agency, in fact, still involves actively, the agency of man on the chosen site of return.

Akio Suzuki's sound art and works therefore stemmed from his understanding of the dilemma of modern Japan as well as modern society as a whole. This explains why his works are simultaneously parochial and universal in its message and therefore in the 1980s when he participated in *Documenta 8* (Kassel, 1987), he received rapturous response from all in attendance. While we go back to nature to embrace it we are still interfering with it. When one critiques the process of modernisation and its supposed ills, one needs to realise the inevitability of human agency on this planet. Suzuki's relatively understated comments on the world via his sound art is thus thought-provoking and they force us to delve deeper into the interconnectedness of things and to see beyond the black-and-white of things which most people are so conveniently latch-

---

[13] Ibid.

[14] Biba Kopf, "Global Ear: Tango/Sapporo" in *The Wire* issue 239 January 2004, p.14.

[15] Ibid.

[16] Ibid.

ing onto when critiquing the world around them[17]. He is thus not providing a didactic answer but through his works we can perhaps be triggered to contemplate a bit more: humans are maybe too egotistical to lay all blame on human agency as Nature often fights back or react to us in mysterious ways which we can never predict. Suzuki does not romanticise about nature; he is more concerned about the often ignored symbiotic relationship between the natural world and humans and he has been trying to remind us again and again through his works.

3. Taking Technology by its horns: The many tales of Yasunao Tone

Yasunao Tone was born in 1935 in Tokyo, Japan. He witnessed the atrocities of the war when the Allied planes carpet-bombed Tokyo on a daily basis towards the end of the war. The surrender of Japan after the dropping of the two horrifying atomic bombs and the subsequent events in Japan and its initial heinous existence in the first post-war years was experienced first-hand by Tone. However like many who grew up and were being exposed to the influx of Western modernist ideas and books in the 1950s, the cultural environment provided creative and cultural fuel for him and many to embrace modernism. Neither a trained musician nor artist, Tone went to Chiba University to study literature[18]. This did not stop him from exploiting the world around him for creative ends. He first dived into "music" making not of a conventional manner but that of a Surrealist and Abstract Expressionist type. According to Tone, "We thought... our improvisational performance could be a form of automatic writing, in a sense that the drip painting of Jackson Pollock was a form of automatic writing."

Tone formed the music group, Group Ongaku in the 1950s with his fellow enthusiasts and an archival CD/LP[19] of the group was released in recent years which shocked many as their pieces sounded like the works of equivalent outfits in Euro-America like AMM, MEV and Gruppo Di Improvvisazione Nuova Consonanza[20] (which featured renown film composer Ennio Morricone), even though Tone and company were earlier than them[21]. Group Ongaku's counterparts in the West were all formed in the mid to late 1960s and this provides ample evidence challenging the common historical narrative that the West has always been the avant-garde in modern arts and culture. He also linked up with the key artists of the Japanese art and music avant-garde in the late 1950s and 1960s: Takehisa Kosugi (who founded the acclaimed performance group Taj Mahal Travellers in the 1970s as well as serving his residence as music director of Merce Cunningham Dance Company later that decade), Yoko Ono, infamous Japanese performance art group Hi Red Centre and avant-garde composer Ichiyanagi Toshi (who amongst many, worked with John Cage and polymath poet, director and playwright Shuji Terayama[22]). He was involved in the Hi Red Centre anti-Olympic performance piece during the Tokyo Olympics of 1964, *Cleaning Piece*, which saw him and other participants and posed as street cleaners in the business district of Ginza, stopping traffic for three hours and in fact receiving cooperation from the unknowing police[23]. Tone was working and collaborating

---

[17] Akio Suzuki, "Akio Suzuki" in Yang, Yeung (ed.) (2010), *Pocket: 1 – Around*, Hong Kong: Soundpocket, p.82.

[18] Alan Licht, "Yasunao Tone - Random Tone Bursts" in *The Wire* issue 223 September 2002, p.31.

[19] The recordings were edited from two sessions of music-making: the first took place on May 8 1960 and the second session happened on September 16 1961.

[20] AMM, a British group, and MEV, a largely expat American collective in Rome, were formed in the mid-1960s in Britain and Italy respectively while Gruppo started in 1964 in Italy.

[21] Licht, *The Wire* issue 223, p.31.

[22] William A. Marotti, "Sounding the Everyday: the Music group and Yasunao Tone's early work" in Various (2007), *Yasunao Tone: Noise Media Language*, Los Angeles: Errant Bodies Press, pp. 13-33.

[23] Cope, pp.54-57.

actively in a milieu that demonstrated rabid interdisciplinary intent and even border-erasing outcomes that had clear resonance with the utopian and counter-cultural fervour felt by so many artists in both Japan and the West during the 1960s.

Similarly, Suzuki was also a key member of the *Fluxus* movement in Japan and he contributed tapes and scores to *Fluxus* founder George Maciunas and he was involved in the performances of *Fluxus* pieces in Japan in the 1960s and in New York later. His penchant for irreverence prompted him to comment in recent years that *Fluxus* was only properly historically meaningful from 1962-66, and that while the works are being exhibited in the museums today, to him it meant that these works are being stripped of their original intentions and functions of being handled and played by both the artists and the audiences[24]. The same spirit gleaned here is similar to Tone's artistic purpose and the conceptualisation of his sound art that he started creating in the 1960s. His disdain and questioning of modern technology in the post-World-War-II world and the associated techno-fetish of modern society to any new fanged media/technology basically served as fodder for his deconstruction and re-conceptualisation of some of these technological products in his sound art.

Despite Tone's key role in the various contemporary art fields since the 1950s, he was a relative unknown figure until recently. This was due to the fact that his works from the 1950s to the 1970s were largely real-time happenings and performance pieces with little documentation. It was also due to the fact that Tone was residing far away from the world cultural capitals (i.e. the West) that he suffered from the more European-American biased focus on *Fluxus*, the Happenings and related art movements in Euro-America by critics and historians in those years. Only when he went to New York in 1972 he became more open to the idea of documenting his works, particularly with the advent of improved recording technologies in sound and video at that time.

Tone's most famous work is his mid-1980s "wounded CDs". The direct inspiration of this phase of his works was a book entitled *Science Seminar For The Familiar*[25]. The part which captured Tone's imagination was a chapter on digital recording, which highlighted to him the error correction programme in CD players which kicks in if a one is misread as zero in the machine's binary codes. Since the error creates a totally different sound, unknown of, he then embarked on a path to override the error correcting system. Eventually, a friend suggested putting adhesive tape with many pinholes in it on the CD itself[26]. The initial impressions one gets upon hearing the outcomes when these "wounded" CDs are played in the machines are that of typical CD skipping/malfunctioning due to surface blemishes. The actual mechanism behind, however, is more than that. As Tone shared, "... numbers are altered so it becomes totally different information. The Scotch tape enables me to make burst errors without significantly affecting the system and stopping the machine[27]." Tone's interventions in the process of the readings of CDs are actually acts of creative noise. By disrupting the machines that are supposed to be the state of the art technology back in the 1980s when the world was moving from the analogue to the digital in many fields of everyday science, Tone successfully questioned and re-inscribed the supposed prescribed workings of an increasingly ordered contemporaneity of the triumphant neo-liberal capitalist paradigm.

Tone wants to interrogate the triumphalism of new scientific and technological "breakthroughs" in the post-War years. With every new product such as this, the media and the corporations together will seemingly snap into a frenzy to valorise the "invincibility" of the product. The fascination with the new and technologically clued-in perpetuates wave after wave of consumer craze to replace existing goods to embrace the new blindly. For example, when the 12-inch vinyl album appeared in the market in the 1960s, it swiftly overtook the

---

[24] Licht, p.31.

[25] Licht, p.32.

[26] Ibid.

[27] Ibid.

previously dominant 7-inch vinyl single format. And when CDs were introduced into the market in the 1980s with the promise amongst many, of its longer playtime, the needlessness to flip the disc over after 20 minutes and of course the purportedly indestructible nature of the CD sent the consumer markets all over to discard the vinyl format to the small shiny plastic discs. Tone's experiments question this notion head-on: are new technologies necessary and necessarily better? He seems to be probing a deeper issue of that of an uncritical celebration of contemporary scientism and modernisation around the world.

To bring his irreverence and critical interrogation for technology up-to-date, Tone ventures into interrogating the next big thing in recent years – MP3 files. In 2009, Tone started collaborating with a team of the New Aesthetics in Computer Music (NACM) at Music Research Centre at the University of York in UK in 2009[28]. Tone wished to develop new software based on the disruption of the MP3. Originally Tone wanted to explore the intervention of the MP3 as reproducing device by interfering with the interface across its main elements (the ones and zeros), the compression encoder and decoder. However, the initial outcomes were unsatisfactory, as they did not much affect the sonic output of the playback of the CDs. Then Tone and his team discovered that by corrupting the sound file in the MP3 format, it would lead to the generation of error messages which could be utilised to assign various lengths of samples automatically. By combining the different playing speeds of the samples, it could produce unpredictable and unknowable sound. To further enrich the software, Tone's team also incorporated other possible elements into the software like flipping stereo channels and phasing ranges that would produce different pitches and timbres[29].

Tone used the software as a performance art tool several times at the MRC at the University of York. He has since also performed in public in Kyoto, Japan, in May 2009 as well as New York City in 2010 and many more subsequent events lined up. He has also produced a CD documenting his performances of his pieces using the said software (**Figure 1**).

Figure 1. *Yasunao Tone's MP Deviations #6+7 CD Cover.* Editions Mego/Peter Rehberg; used with permission.

While the conceptualisation and eventual production/utilisation of the software might be as irreverent as Tone's wounded CDs but it is consistent with Tone's subtle but consistent socio-political critique and his artistic harnessing of the latest consumer-based technology. Even though the first converting, uploading and sharing of MP3 files of existing CDs into the internet in the mid-1990s began more as an act of wilful rebellion, Tone wanted to examine the by-now ubiquitous format on which contemporary life relies upon – the ones and zeros head on. In fact, the MP3 format not only allows the ease of transfer and sharing of data but it is almost physically empty except for its physical carriers and vehicles like computer hard disks, mobile phones and the Cloud. Like most other consumer technologies, MP3 (or other extant formats such as MP4) seems to be used by people merely as a medium of consumption of everyday life for entertainment, work and connecting with others. Most do not see the (often) possible misuse and abuse of technologies to break out of the 'big corporation' push to leverage on such formats and churn out products endlessly.

Tone and his team recognise the potential of the MP3 to be distorted and reconfigured for creative purposes in music and performance art fields. He wants to demonstrate that modern society can and should break out

---

[28] Yasunao Tone, in sleevenotes to CD album, *MP3 Deviations #6+7*.

[29] Ibid.

of the passive consumption of technological products and question and thus create using the seemingly static functions of modern technology. Tone's art projects belie the belief that reclaiming the role of an individual and conversely the society to not be enslaved by their new gadgets and toys but to interrogate them and abuse them innovatively.

Yasunao Tone is a humanist who is concerned with not only the society and politics of Japan or his immediate surroundings. It is always an on-going search and re-creation for a more considered human perception and re-think of the relevance and role of science and technology today. His message and agenda is about alerting us to be always circumspective about the impact of technology on all human life at the end of the day.

4. Conclusion

Both Yasunao Tone and Akio Suzuki are very interested in the tension of humanity, technological advancement and the natural world we all live in today. The recent nuclear and natural disasters in 2011 in Japan were an alarm for all: despite all our advancement in the sciences, we are still at the mercy of nature as well as our own human-driven scientific follies. Suzuki shows us the way back to nature, but he does not romanticise it. In fact, he highlights the irony between humankind embracing and destroying nature concurrently in whatever we do. Tone on the other hand, leads us into the core of new scientific products and deconstructs them to reveal to us the blind spots we have towards technology. Their projects are not about the local society and its parochial views and politics but they are actually about the society and politics of all human kind. In other words, it is about something more universal and all-encompassing.

References

Adlington, Robert (ed.) (2009), *Sound Commitments: Avant-garde Music and the Sixties*. New York: Oxford University Press.

Cope, Julian (2007). *Japrocksampler*. London: Bloomsbury Publishing.

Kelly, Caleb (ed.) (2011). *Sound*. London & Cambridge, MA: Whitechapel Gallery & The MIT Press.

Kim-Cohen (2009), Seth. *In the Blink of an Ear: Toward a Non-cochlear Sonic Art*. London: Continuum Books.

Kopf, Biba. "Global Ear: Tango/Sapporo"*The Wire* issue 239 January 2004 14. London: The Wire, 2004.

LaBelle, Brandon (2010). *Background Noise Perspectives on Sound Art*. London: Continuum Books.

LaBelle, Brandon and Steve Roden (eds.) (1999). *Site of Sound: Of Architecture & the Ear*. Los Angeles & Santa Monica: Errant Bodies Press & Smart Art Press.

Licht, Alan (2007). *Sound Art: Beyond Music, Between Categories*. New York: Rizzoli.

Licht, Alan. "Yasunao Tone – Random Tone Burst" in *The Wire* issue 223 September 2002 30-33. London: The Wire, 2002.

Migone, Christof (2012). *Sonic Somatic: Performance of the Unsound Body*. Los Angeles: Errant Bodies Press.

Munroe, Alexandra (1994). *Japanese Art After 1945: Scream Against the Sky*. New York: Harry N. Abrams, Inc.

Suzuki, Akio. *K7 Box* (CD). ALM Records/Kojima Recordings, Inc. 2007.

Suzuki, Akio. *Odds and Ends* (2xCD). Hören 2002.

Tone, Yasunao. *MP3 Deviations #6+7* (CD). Editions Mego 2011.

Tone, Yasunao. *Yasunao Tone* (CD). Asphodel 2003.

Toop, David. "Akio Suzuki – Acoustic Trickster" in *The Wire* issue 231 May 2003. London: The Wire, 2003.

Toop, David. "On Location – Akio Suzuki: London School of Oriental and African Studies UK" in *The Wire* issue 218 April 2002. London: The Wire, 2002.

Various (2007). *Yasunao Tone: Noise Media Language*. Los Angeles: Errant Bodies Press.

Various (2010). *Around by Soundpocket 1* Programme Catalogue. Hong Kong: Soundpocket.

Yang, Yeung (ed.) (2010). *Pocket: 1 – Around*. Hong Kong: Soundpocket.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015 pp 27-34.

Performing Sounds Using Ethnic Chinese Instruments with Technology: SA

Natalie Alexandra Tse, Andy Chia

SAtheCollective@gmail.com

Abstract

From a first-person research method, using the perspective of Singaporean experimental Chinese music group SA, this paper discusses the questions of motivation for incorporating technology with ethnic Chinese music instruments from a socio-cultural perspective; how this incorporation of technology is performed and has affected the perception of sound performance for the musicians; what were the initial challenges and how the challenges were overcome.

Keywords: Ethnic Instruments, Sound Technology, SA

1. Introduction

This paper is born out of the necessity to address several questions about SA, an ethnic Chinese instrumental trio utilizing the instruments of the *dizi* (flute), *guzheng* (zither) and percussions such as the *dagu* (drum), along with electronic effects pedals. The inquiry addresses a range of issues, from identity and motivation to performative, artistic, and technical reflections.

Some fundamental questions that we had reflexively asked ourselves are: *Who are we? What do we do? Why do we do what we do?* In relation to our music and performance, we have also questioned: *How has our artistic choice of incorporating the use of technology changed our perception of sound in our performance as an ethnic Chinese musician? How has it changed the compositional process? How has it developed the instruments? How has it changed the interaction between the musicians in the group?*

As members of the group, (Natalie being the *guzheng* instrumentalist and Andy the *dizi* instrumentalist of SA), we are continuously engaged in a form of action research (more common in the field of education than performing arts, but which we find highly applicable to the performing arts as well) which demonstrates reflective practice. This paper also aims to explore the phenomenon of SA, taking a first person subtextual phenomenological approach (e.g. Vallack, 2010; Valera & Shear, 1999).

Before we move on to discuss the aforementioned questions, it is important to give a brief introduction to SA. Having founded SA in 2011, we have been the constant members of the group, while Cheryl, our third permanent member who plays Chinese drums, percussion and drum kit joined the group in 2013. The group's portfolio emphasizes the performance of 'sound', as opposed to 'music'. The essence of the group could be seen in the following section of our biography:

> *"For SA, each instrument is merely the start of an exploration in sound. Every original composition is a bold experiment that combines traditional elements with modern techniques such as live looping. Triggered live, every performance is a risk that the band takes to interpret the question of identity."*

To date, SA has released one debut EP in August 2013, debuted their first ticketed concert in 2015 at the Esplanade Recital Studio (Esplanade, 2008), performed locally and internationally, having recently completed our France-China-Macau-Hong Kong tour, as documented on our various social media platforms (e.g. Facebook and Weibo).

In answering the questions of our identity and motivation, the formation of SA cannot be discussed in isolation from its socio-political-cultural context, which will be addressed in the following section.

2. What was the Motivation for SA to Incorporate Technology into Ethnic Music Making?

As we questioned ourselves as performing artists why we have consciously chosen to incorporate the use of technology such as electronic effects pedals, and microphone techniques as our tools of performance, our responses reflected not only personal motivations but motivations driven by social circumstances as well.

On a personal level, all 3 of us as musicians feel that we no longer wanted to be bound by our classical training and practice. The notion of performing set, composed pieces, practicing our own parts to technical perfection to be put together within an orchestra was no longer our individual aesthetic philosophy. Whether as individuals or as a group, we are seeking new possibilities in sound and art.

In conversations with Cheryl, she said, "all of us have been classically-trained and were very used to the classical musical environment...a composer writes a song, you get your parts in the form of a score and rehearsal begins... there is hardly any room for creative ideas or improvisation. So i guess we kind of got sick and tired of that... there's hardly any space to express yourself through your music, it's always someone else's composition/ idea that you are trying to put across. The interesting thing about playing in SA is that it is an avenue for us to express our ideas and ideals, using our instruments / electronics as a tool, so it's always an open book... we are always inspired by different things at different times, so the music is ever-changing (at least i hope!)" (Ong, 2015).

On a socio-cultural level, SA's aesthetic philosophy may not be entirely agreeable with certain directives put in place via the National Arts Council, which has further motivated the group to seek new artistic and sonic possibilities, which led to the incorporation of technology. This motivation emerged from negotiating dilemmas in Singapore's Chinese music scene, which falls under National Arts Council's *Traditional Arts* (TA) sector.

In Singapore, the TA sector falls under a specific and separate category from the rest of the art forms of music, dance or theatre under the National Arts Council (NAC), as evident in the *Performing Arts Masterplan* released in 2014 (National Arts Council, 2014). As can be seen from the appended documents following the master plan, the TA sector can be seen to be adhering to dominant national ideologies such as the Chinese, Malay, Indian and Others (CMIO) construct (Chua, 2003; 1995 & Siddique, 1990), which clearly delineates TA to be Chinese, Malay, and Indian respectively. In this national directive for the performing arts sector in Singapore, it is also specifically stated that "NAC will be undertaking a research study that will explore how certain TA forms were introduced, evolved and gained traction in their respective communities" (National Arts Council, 2014). This shows that there is a special focus that NAC has placed on what they have defined to be TA.

The master plan articulates examples of Chinese music groups, as follows:

- Amateur orchestras and instrumental ensembles (e.g. *guzheng* ensembles) in schools and community centres.
- Professional, i.e. the Singapore Chinese Orchestra.
- Amateur/semi-professional groups such as Ding Yi Music Company and Siong Leng Musical Association

It is interesting to note here that through these examples, the NAC has undoubtedly, whether consciously or not placed a definition of what Chinese music should be, under the TA sector. One fundamental similarity between these examples is that the nature of performance leans towards classical (orchestras and ensembles) and ancient, traditional genres (Siong Leng Musical Association), thus putting contemporary or developing genres that use ethnic instruments, such as SA, neither explicitly in the TA sector nor in the music sector under NAC.

Interestingly, an article by the "Straits Times Life!" in 2014 featured several other Chinese instrumental groups - including SA - aside from Chinese orchestras and instrumental ensembles, which NAC's master plan did not explicitly acknowledge. Therein lies a dilemma that SA has negotiated with and emerged from to motivate our current artistic and sonic direction, which forms our Identity.

Firstly, it is ironic that Singapore, an immigrant country, places much emphasis on the development of her traditional cultural art forms based on the host country (i.e. China) as opposed to recognising what had been born out of Singapore itself. The Chinese orchestra is a 20th century construction, born out of the need to compete with Western, symphonic orchestras (Wong, 2005 & Jones, 1995). Yet, the NAC has identified the Orchestra as the only professional form of Chinese music practice in Singapore. In our opinion, this competes with the idea of developing Singapore's traditional art forms, as other embodiments of Chinese music, such as *jiang nan si zu* ("as in the silk-and-bamboo music of the Shanghai literati") (Jones, 1995) or even Cantonese music are not considered. Needless to say, the original intentions and significance of Chinese music, such as its ceremonial/ritualistic functions, or its function as court music (Jones, 1995) had not been considered as well.

This dilemma presented in the national directives set forth by the NAC has made SA question the social function and value of traditional Chinese music, and the pursuit of an ethnic Chinese music identity that not only allows creative expression and imagination, but is also relevant and representative of ourselves in current times.

All three musicians in SA were born in the 1980s, children to parents belonging to the Baby Boom Generation (1946-1964) after World War II (Ministry of Social and Family Development, 2015 & Roy, 2014).

Throughout the process of nation building, there have been many national policies that Singaporeans have been subjected to. This brings us to the 2nd dilemma that we as ethnic Chinese musicians in SA face. The introduction of the "Speak Mandarin Campaign" in 1979 (Speak Mandarin Campaign, 2015) also marked the demise of Chinese dialects in Singapore. In a socio-political move to unify the Chinese community in Singapore "to use more Mandarin and less dialect" (Speak Mandarin Campaign, 2015), also resulted in a loss of ethnic Chinese cultural identity, and resultantly, a loss of musical vocabulary and genres in Singaporean ethnic Chinese music.

A majority of Singaporean Chinese trace their ancestry to Fujian and Guangdong. The dialect groups of Hokkien, Teochew, Cantonese, and Hakka make up a large component of the population (Lee). However, the percentages of Singaporean Chinese who speak Chinese dialects have dropped from 30.7% in 2000 to 19.2% in 2010 (Department of Statistics, Ministry of Trade & Industry, Republic of Singapore). This implicates a loss of traditional musical styles, such as the 'silk-and-bamboo' genres common in southeast China (Jones, 1995) as mentioned earlier. With a diminishing knowledge or identification with such styles, Singaporean Chinese music practitioners resonate with composed pieces for the orchestra rather than the musical vocabulary of traditional genres that traces back to our ancestry.

Thus, the motivation for SA to seek new artistic and sonic possibilities was also a realization that while we may be able to replicate the forms and techniques of Chinese musicians from mainland China (who have largely been subjugated to the "conservatoire style" [Jones, 1995]), we will never be able to replicate the essence of Chinese music at its origin, as we neither speak their language (in terms of dialects) or live in their environment. With the influences that we have experienced as Singaporean Chinese, we were motivated to seek sonic experiences that we could call our own.

Thus, through reflexive conversations amongst ourselves and retrospectively as individuals, we are thus able to identify that SA made an artistic, conscious choice to incorporate the use of technology in our music-making process because we did not agree with the national direction which TA was taking, and we believe in our own creation and expression that undoubtedly have foundations in mainland Chinese music, but which we should not blindly abide.

3. How do we Perform Sounds using Ethnic Chinese Instruments with Technology?

In order to understand the nature of SA's sound and performance, the following section illustrates how we use technology with our ethnic instruments in performance.

Referring to the **Figure 1**, all three musicians in SA use a similar signal chain to execute our performance. Firstly, using miking techniques unique to each instrument, a dry signal of the instrument is routed through a sub-mixer. Then, through the auxiliary channels of the sub-mixer, the dry signal is output to the effects pedals which each of us use. These effects vary according to each one of us, which include various different combinations of effects processors that modulate time in terms of delay, vary our pitch and harmony, as well as modulate the notion of space through reverberation/resonance. In addition, it gives us the ability to create multi-layered tracks through live looping. Next, the wet signal from our effects pedals is routed back to the same sub-mixer. Finally, the summation of the wet and dry signals routed through the sub-mixer is then sent to the sound engineer's main mixing desk.

As mentioned, the effects pedals we employ allow us to modulate time, pitch and space, as well as create different layers through live looping. This allows us a magnitude of sounds that we can manipulate in our performance, which was never possible previously with our ethnic instruments. Although this may not be new to the western musical world and the incorporation of technology with acoustic instruments had already been widely discussed, (e.g. Arias, 1998; Emmerson, 1998 & Friedrickson, 1989) it is fairly new for ethnic Chinese musicians, with little or no research documentation in the area.

Undeniably, the incorporation of technology in this manner has changed our performance and perceptions in various aspects. As Arias (1998) states, "Technology propitiates a fundamental rethinking of music and rehearing of sound, paving the way for a musical praxis that can be tailored specifically to any particular individual and social context while remaining almost free of any predetermined cultural traits."

For SA, the incorporation of technology has definitely changed the way we think about our music-making, and the way we hear sounds. Though we are using instruments specific to our ethnic identity, technology also allows us to retain our ethnic Chinese identity, while creating new sonic experiences that are free of these "predetermined cultural traits" as mentioned by Arias (1998).

Our new perception towards sound is no longer limited to how we have been classically trained as ethnic Chinese musicians. As mentioned earlier, our 'traditional' practice dictates the playing of set, composed pieces according to the score, and to practise that piece to technical perfection, perhaps at the same time, to deliver the musicality that the piece dictates. We had been used to replicating a piece according to the way in which our teachers have taught us. However, technology has given us the space to explore.

In the traditional performance of fixed repertoire, we are often limited to linear melodies, especially with the *dizi.* With the incorporation of technology, the *dizi* is now able to play chords using a harmonist pedal; or, with a looping device, the instrument can now create layers of notes to form chords. Traditionally, in an ensemble setting, each musician has only one part to play, but with the use of the "looper", one percussionist can now sound like an entire percussion ensemble.

Though there have also been developments techniques and instrument making in Chinese instrumental music, the possibilities in which technology offers us is tremendous. For example, although the modern development of playing techniques on the *guzheng* now incorporates the use of the left hand heavily to create accompaniments for right hand melodies (or interchangeably), as opposed to linear melodies on the right and pitch bending techniques on the left hand previously, the *guzheng* is still highly limited by its tonality. Though the fixed tuning method is utilized in contemporary compositions for the instrument, the *guzheng* is limited to only one key with each tuning or repertoire. Any change in tonality requires the shifting of the movable bridges on the instrument. With effect pedals that modulate pitch, the potential for the *guzheng* in SA to perform harmonies simply by striking one string, or even arpeggios, now surpasses its traditional limitation of tonality. This reflects Arias' (1998) statement "in which the methods and aesthetics are not intrinsically bound to the tradition that produced it."

The possibilities on each of our instruments in SA have also highly altered the way in which we compose music. Whereas in the past, one of us would write a piece of music out in notations using Sibelius with individual parts for each one of us to subsequently rearrange, we now create new repertoire simply by coming together and jamming. As opposed to looking at a notation and imagining the sound and harmony of a piece of music, the sonic possibilities which we are now able to produce has surpassed what is able to be represented visually on a traditional staff notation score. Our sonic imagination has gone beyond what can be documented in writing. Our song writing process has evolved from notation writing to selecting and developing specific segments from our audio recordings during jamming sessions. Each one of us in SA knows our parts by heart, and would be almost impossible to replicate in its entirety by another musician. This also allows our live performances to by highly dynamic in nature – each performance is unique because we are not bound by score, and our repertoire allows for a high degree of freedom in improvisation.

The interaction amongst the three of us in SA is also no longer limited to one musician playing one part in a piece of music. Previously, as traditional classical Chinese musicians, we can always anticipate what our fellow musicians in the ensemble or orchestra is going to play, and how it is going to sound. With SA, every music-making session is exciting and no longer anticipatory. There are creative expressions in response to the sounds

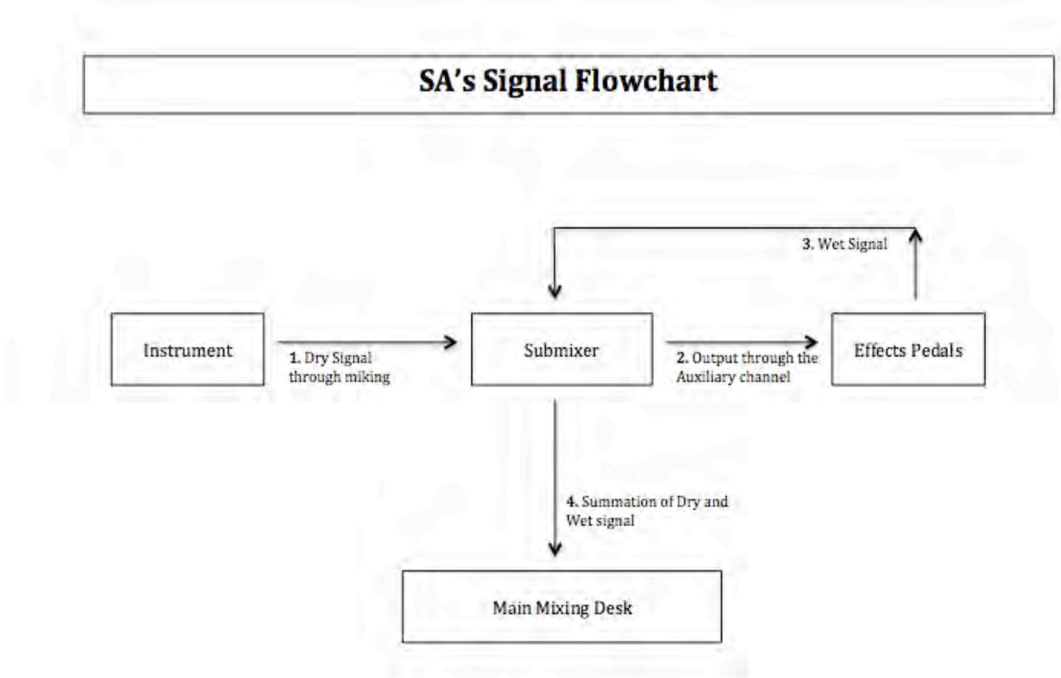


Figure 1. SA's signal flowchart.

we hear as a trio rather than as individual musicians. An interesting change for the incorporation of technology in performance within SA is that our sound engineer has become a paramount factor and thus member of the group. This is related to the challenges that SA had faced in the incorporation of technology, which shall be discussed in the next section.

4. What were SA's initial challenges in incorporating technology?

In conversations with our sound engineer, Yew Jin, he said, "for traditional ethnic instruments to play on live stages which involves big amplification using big scale speakers to project on, to project out, the whole workflow is designed based on western instruments, so, some of the for example, microphones, may not be entirely suitable for pick-up" (Lee, 2015). As a result of the use of western sound technology theories and concepts on ethnic Chinese acoustic instruments, some of the challenges which SA faces are – an imbalance of wet and dry signals, sound monitoring issues, and feedback issues. Though these may be issues common to any musician who use instruments (be it electric guitar or bass) with effects pedals, they are real issues that SA faces. In addition, unlike instruments such as the electric guitar or bass, we are working with acoustic instruments. These issues affect our performance levels. Thus, the execution of a successful performance for SA is also dependent to a certain extent on our sound engineer, who needs to know our repertoire and sound well.

For example, in the process of looping, we may sometimes be unable to achieve the same volume between our wet and dry signals. As SA uses acoustic instruments, when we stop playing live after we have executed a loop, it results in a drop in volume levels when we stop playing live. This also results in difficulties when we as musicians monitor sounds on stage during a live performance.

Sound monitoring had proved to be highly challenging during our initial phases of experimenting with technology in our performances. As mentioned above, the imbalance in volumes between loop levels and live sound playing affects the monitoring of the loops during performance. Due to the gentler, softer nature the *guzheng*, it was common during the initial period where it was challenging to monitor the sound of the *guzheng* on stage during performance. These are issues that affect our performance levels during a show.

In addition, the inaudibility that we feel results in the increase of volume on stage monitors. In return, this results in feedback issues, as the stage sounds are too loud, and the various microphones on stage pick up the stage sounds, resulting in feedback loops. This problem has been solved through the use of in-ear monitoring (IEM). Instead of using stage monitors, we sometimes use IEM to monitor our own sound, as well as the band's sound. However, the use of IEM reduces the level of musicianship, as we sometimes become too consumed in the direct sounds input. This results in a lower level of engagement with one another, as well as with the audience.

The nature of our instruments as acoustic, ethnic instruments was a challenge in itself. For the *dagu*, it is highly dependent on weather as it is made of animal skin. Thus, the sound of the instrument varies with temperature and humidity levels, and is not within the control of the musician. For the *dizi*, the resonant frequency membrane (*dimo*) interferes with the primary tone that is produced. This proved to be challenging as it interfered with the use of modulation pedals. The long and big sound box of the *guzheng* is problematic as it frequently resulted in feedback loops in the chamber of the body of the instrument.

Some of these challenges were solved by simple modifications to the instruments. For example, instead of using the *dimo* on the *dizi*, a tape over the membrane hole solves the issue of resonance that interferes with the use of modulation pedals. More recently, we have customized *dizis* without the membrane hole as well. For the *guzheng*, we used shorter, travel-sized instruments available in the market and modified its soundboard by smoothening its surface, to achieve consistencies in the height of its movable bridges. We also increased the height of its saddle to balance the tension of the strings from the saddle, to the movable bridges of the instrument, over a smaller sound box. Thus, the reducing its susceptibility to feedback loops. For the *dagu*, an alternative to animal hide needs to be sought, but it will also be a negotiation which our drummer Cheryl will have to make on a personal level as it will affect the sound and thus her performance levels.

As mentioned earlier, the incorporation of technology in SA's music-making changes the way we perceive sound. Such a paradigm shift towards sound perception required us to have an open mind about possibilities, which may at times run contrary to our traditional, classical training. Conversations with our mentor from the *Noise Singapore Music Mentorship Programme*, Randolf Arriola pointed us in the direction whereby there was a need for a paradigm shift in our perceptions towards sound. He said, "there was the need to identify the motivations and find the right balance between sonic and performance aesthetics, novelty and dynamics and innovative sound design and tasteful application that function as creative tools to enhance rather than simply to ornament or overwhelm the integrity of artistic musical composition" (Arriola, 2015). In the initial stages of incorporating technology into our performance, we had to be mindful not to lose our musicianship, but instead, enhance our performance through contemporary techniques.

5. What's next?

Performing sounds on ethnic Chinese instruments with the incorporation of technology has given an opportunity for SA to question our aesthetic philosophies and ideals, which have helped clarify our identity, motivations and focus on sound perception. It has resulted in classically trained musicians playing a fixed repertoire to being able to participate in a more improvisatory form of performance that revels in creative sound-making. We are constantly experimenting different ways to explore sound, with further insights on sound, and building our knowledge on how to use technology to enhance our performances (such as modulating sounds simply by manipulating our sub-mixers and effects pedals through feedback loops, with or without our instruments), and also how to perform using extended and unconventional acoustical techniques. Rather than employing conventional methods and techniques, we now have the freedom to explore different ways to create and perform sound. The limitless possibilities only give us further potential to inquire deeper into sound art and performance.

Moving forward, much research needs to be conducted in the areas of organology for ethnic instruments (not limited to Chinese) for efficient use with technology. Having written this paper adopting a first-person approach is also limiting where further socio-political and cultural issues could be discussed in terms of the use of ethnic instruments with technology, and further theoretical frameworks in the area of sound and technology could also have been employed for a deeper understanding of the subject matter.

6. Conclusion

In essence, this paper has addressed the following issues: the motivations for SA as an ethnic Chinese instrumental trio to incorporate technology from personal and social perspectives; how SA performs with ethnic instruments and technology and how this have affected the perception of sound, composition and relationship between the musicians; and finally, what initial challenges emerged and how some of them were resolved.

References

Arias, Ricardo. (1998). "From the Margins of the Periphery: Music and Technology at the Outskirts of the West – A Personal View." In *Leonardo Music Journal 8, pp.49-54*.

Arriola, Randolf. (2015). Personal Communication. 17 May 2015.

Chua, Beng Huat. (2003). "Singapore – Multiracial harmony as public good". In

Ethnicity in Asia. Ed. Mackerras, Colin. London: RoutledgeCurzon.

Chua, Beng Huat. (1995). "Culture, Multiracialism and National Identity in Singapore. " *Working Paper.* Singapore: National University of Singapore Department of Sociology.

Department of Statistics, Ministry of Trade & Industry, Republic of Singapore. (2010). *Census of Population 2010 Statistical Release 1 Demographic Charactersitics, Education, Language and Relision.* Department of Statstics Singapore.

Emmerson, Simon. (1998). "Acoustic/Electroacoustic: The Relationship with Instruments". In *Journal of New Music Research 27(1-2), pp. 146-164.*

Esplanade, (2008.) "SA – The Pursuit" *What's On?* Retrieved 13 may 2015, from http://www.esplanade.com.sg/whats_on/programme_info/latenite_sa_the_pursuit/index.jsp

Facebook. (2015). *SA.* www.facebook.com/SAtheCollective

Frederickson, Jon. (1989). "Technology and Music Performance in the Age of Mechanical Reproduction". In *International Review of the Aesthetics and Sociology of Music 20(2), pp.193-220.*

Hass, Jeffrey. (2013). "Chapter Two: Studio Gear". *Introduction to Computer Music, Volume One.* Indiana University, Jacobs School of Music, Centre for Electronic and Computer Music. Retrieved 18 May 2015, from http://www.indiana.edu/~emusic/etext/studio/chapter2_effects.shtml

Jones, Stephen. (1995). *Folk Music of China – Living Instrumental Traditional.* United States: Oxford University Press.

Lee, E.F. Edmond. (n.d.) "Profile of the Singapore Chinese Dialect Groups. Social Statistics Section." In *Statistics Singapore Newsletter.* Singapore Department of Statistics.

Lee, Yew Jin. (2015). Personal Communication, 18 May 2015.

Ministry of Social and Family Development. (2015)."Chapter 1 Demographic Realities." *In Report on the Ageing Population.* Ministry of Social and Family Development, Singapore, Committee on Ageing Issues. Retrieved 11 August 2015 from http://app.msf.gov.sg/Portals/0/Summary/research/Chapter%201%20-%20Demographic%20Realities.pdf

National Arts Council, Singapore, (2014). *Building Blocks for a Culture of Creation: A Plan for the Performing Arts. Annex E – Traditional Arts Plan.* Retrieved 10 December, 2014 https://www.nac.gov.sg/art-forms/pa-masterplan

National Arts Council, Singapore, (2014). *Appendix I – Traditional Performing Art Forms in Singapore. Overview of Active Traditional Performing Art Forms in Singapore.* Retrieved 10 December, 2014, https://www.nac.gov.sg/docs/schemes/8-appendices-i_vi.pdf

Nylon Singapore. (2014). "Nylon Singapore Exclusive: SA(仨)".

Ong, Cheryl. (2015). Personal Communication. 11 May 2015.

Roy, S. (2014). "Baby Boom Generation in Singapore and its Impact on Ageing." *International Science Index 8(3).* Retrieved 11 August 2015, from http://waset.org/publications/9997944/baby-boom-generation-in-singapore-and-its-impact-on-ageing

Siddique, Sharon, (1990). "The Phenomenology of Ethnicity: A Singapore Case Study." In *Understanding Singapore Society.*Ong, Jin Hui, Tong Chee Kiong and Tan Ern Ser. Eds. Singapore: Times Academic Press, 1997.

The Straits Times Communities, (2014). *The Pipa Quartet: Rocking Out on the Pipa.* Retrieved 10 May 2015, http://www.stcommunities.sg/entertainment/music/st-review/pipa-quartet-rocking-out-pipa

Singapore Chinese Orchestra. (2015). *Singapore Chinese Orchestra.* www.sco.com.sg

Siong Leng Musical Association. (2014). *About Us.* www.siongleng.com

Speak Mandarin Campaign. (2015). *About the Campaign.* http://mandarin.org.sg/en/about

Vallack, Jocene. (2010) "Sub textual Phenomenology: A Methodology for Valid, First-Person Research" *The Electronic Journal of Business Research Methods 8(2), pp.109-122.* Retrieved 13 May 2015, from www.ejbrm.com

Varela, Francisco J., & Shear, Jonathan. (Eds.) (1999) "First-person Methodologies: What, Why, How?" In *The View From Within – First Person Approaches to the Study of Consciousness.* Exeter, UK: Imprint Academic.

Weibo. (2015). *SAtheCollective.* www.weibo.com/SAtheCollective

Wong, Samuel, S.M. (2005). *Qi – An Instrumental Guide to the Chinese Orchestra.* Singapore: TENG.

array

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 35-42.

Si15

The Language of Dance: Testing a Model of Cross-Modal Communication in the Performing Arts

Suzy J. Styles

Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

suzy.styles@ntu.edu.sg

Abstract

Integration between the senses is an intrinsic part of the human condition. Many forms of artistic expression make use of these sensory alliances, for example, the expression of rhythm and melody in dance. To test whether performers can effectively communicate information from one sensory modality (hearing) into another (vision), we asked one experienced dancer to perform dance-motions to the sounds of meaningless speech, and asked junior dancers to guess which dance motions were produced in response to which sounds. The junior dancers were substantially better than chance in the guessing task, suggesting that the dance performer successfully captured acoustic information about the identity of the speech sounds in her motions. We also found that dance experience did not predict performance in the task, suggesting that sensory congruence may not be learned through practice, but may be shared among the general population. However, a subset of dancers were much better than the main group, suggesting that sensory congruence may be differentially distributed through the population. This fits well with a model in which the strength of sensory connectivity differs across the population, and in which the creative arts attracts those individuals for whom the intrinsic links between the senses are experienced more powerfully.

Keywords: cross-modal perception, performing arts, sound symbolism

1. Introduction

Information floods our senses simultaneously, giving us an intensely multi-modal experience of our world. These sensory co-activations allow us knowledge about how animals with louder, lower-pitched calls have larger bodies (Morton, 1977), and that lowering certain resonant vocal frequencies generates dominance gestures (e.g., Chimps, ‘ooh ooh’ dominance versus ‘ee ee’ submission calls: Ohala, 1994).

The implicit understanding that certain kinds of sensory information ‘go with’ others is the tip of an evolutionary iceberg known as ‘cross-modal correspondences’. For many years, researchers in the psychological sciences have experimentally demonstrated that high pitches are typically mapped to smaller, paler, spikier shapes, positioned higher up in visual space (Evans & Treisman, 2010; Spence, 2011), effects that are shown by chimpanzees (Ludwiga, Adachid, & Matsuzawa, 2011), as well as by human infants (Mondloch & Maurer, 2004; Walker et al., 2010). These findings demonstrate that connections between the senses are evident from the beginnings of human development, and perhaps most importantly for the creative arts, these correspondences are shared.

This means that sensory relationships for one person should more-or-less align with sensory relationships for the community at large: Our sensory system provides a platform for communication *across the senses*.

One off-shoot of cross-modal processing arises in the domain of language. Despite the ability of language to use arbitrary sounds to encode abstract meanings (de Saussure, 1911:1959), there is a growing body of evidence showing that certain speech sounds 'go with' certain sensations in other modalities. For example, when given two novel names for two novel objects, (one curvy, one spiky), the majority of people have a strong preference to match up the curvy shape with 'round-sounding' word-forms like 'bouba' or 'maluma', and spiky shapes with 'hard-sounding' word-forms like 'kiki' or 'takete' (Köhler, 1929:1947; Ramachandran & Hubbard, 2001).

This bouba/kiki task has been replicated in a variety of languages, for participants at various ages, including infancy (e.g., Bremner et al., 2013; Ozturk, Krehm, & Vouloumanos, 2013; Spector & Maurer, 2013). As such, sensory congruence for speech sounds forms part of the constellation of shared inter-sensory mappings.

Where sensory integration differs dramatically from the general population, we tend to think of this as dis-orderd. For example, people who experience 'synaesthesia' may spontaneously see colours and shapes when they listen to music, or experience colours when they see particular letters in printed text. The particular sensory experiences differ between individuals, but within an individual the sensations remain stable throughout adulthood (for overview, see Cytowic & Eagleman, 2009). Interestingly, synaesthetes tend to be over represented in the arts (Rothen & Meier, 2010).

Ramachandran and Hubbard (2001) proposed that synaesthesia might exist at the extreme end of normal sensory 'connectivity' or cross-modal integration, suggesting that inter-sensory connectivity is so strong it generates a kind of automatic sensory 'cross-talk'. This model of sensory congruence suggests that although we are biologically endowed to experience sensory connectivity, the extent of functional connectivity differs across the population; with dis-regulated hyper-connectivity at one end of the spectrum (i.e., synaesthesia), and dis-regulated hypo-connectivity at the other (e.g., sensory integration deficits).

Recent evidence from brain imaging studies has indeed suggested that people who show greater sensitivity to linguistic sound symbolism also exhibit greater activation in left superior parietal cortex, a brain region associated with cross-modal integration, as well as hallmarks of greater functional connectivity in the left superior longitudinal fasciculus (Pirog Revill, Namy, DeFife, & Nygaard, 2014). This finding lends support to the idea that differences in brain structure and function underpin differences in the strength of sensory integration across a population.

At the extreme end of the integration spectrum, several studies have also shown differences in brain structure and function for synaesthetes, as compared to the normal population (for overview, see O'Hanlon, Newell, & Mitchell, 2013). At the other end of the spectrum, people with autism show lower sensitivity to the bouba/kiki effect (Oberman & Ramachandran, 2008; Occelli, Esposito,

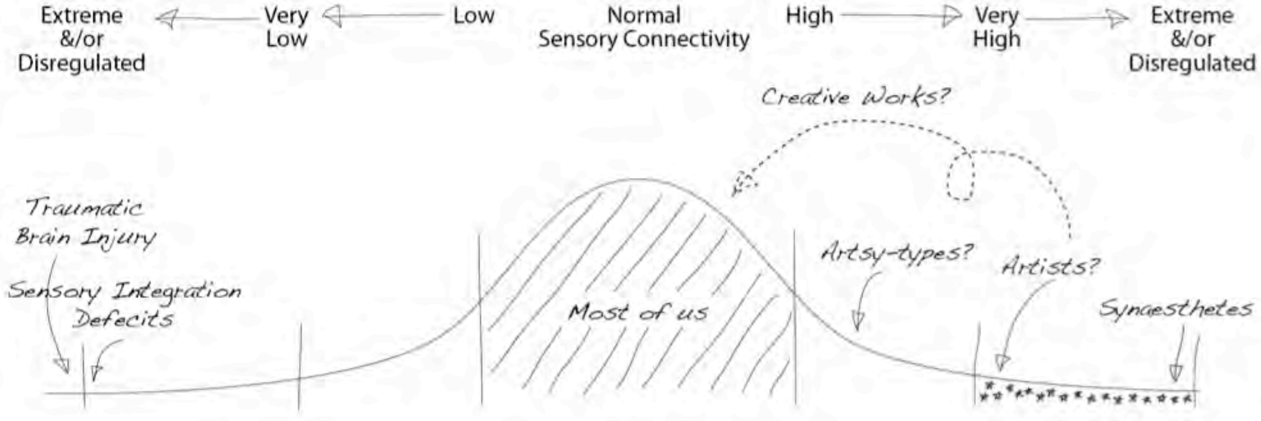


Figure 1. A population model of sensory integration strength. Normal distribution curve showing 1, 2, and 3 standard distributions from mean.

Venuti, Arduino, & Zampini, 2013), and this may be a marker of sensory integration deficits which are frequently reported as co-occurring with autistic traits. Taking these sources of information together, the idea of individual differences in sensory integration across the population seems well supported by the current evidence. A model of population differences in sensory connectivity is sketched in **Figure 1** (previous page).

This leads us to draft a model of sensory integration strength and creative arts across the population. In his public lectures (2003), Ramachandran has suggested that artists may be those in a community with greater sensory connectivity, which provides them with an enhanced ability to generate novel sensory metaphors (e.g., a sweet melody, a dark expression, angry brush strokes), but since their sensory congruence is part of the system shared by the population, these novel metaphors can be readily interpreted by the majority.

If we follow these suggestions to their logical conclusion, we can formulate a model of how certain creative arts relate to sensory integration strength, which gives rise to some interesting predictions: Firstly, there should be more consumers of art than creators, as the majority of us are endowed with sufficient sensory congruence to appreciate a well-made sensory alliance, but perhaps not the depth of sensory integration to imagine a novel relationship we have not yet encountered.

Secondly, there should be a sector of the community who are more deeply inter-sensory than most – they may be 'artistically inclined' but perhaps not as deeply engaged as artists (see **Figure 1,** 'Artsy types'). These people would be predicted to get more enjoyment than normal out of well-matched cross-modal experiences, and may actively seek them out: They may visit galleries, sing in choirs, dance in an ensemble, or perform in an orchestra, but they might be less inclined to create, conduct, choreograph, compose, or improvise.

Thirdly, there should be a corresponding sector of the community for whom sensory integration is lower, and may simply be less compelling. People in this group may be 'artistically disinclined', and have little interest in the non-literal components of art (e.g., "I like pictures where you can tell what it's supposed to be", "My three year old can drawn better than that").

Finally, we should predict that when an artistic expression 'works,' it is because the creative artist is able to communicate something (a percept, an emotion) through their chosen medium, in such a way that an audience is able to 'get it'.

The current study

Here, we used the medium of dance to see whether an artist (an experience dancer) could translate sensory information from one modality (sound) into another (sight), by generating novel dance movements to abstract sounds of speech. If our dancer can create motions which tap into sensory mappings shared with the general community, then other viewers of her motions should be able to extract relevant information about the sounds she had heard. If, on the other hand, her motions are too idiosyncratic (e.g., too synaesthetic), or don't contain sufficient sensory information (e.g., too weakly integrated), they might not contain sufficient shared sensory information for her audience to identify the sounds. In this small-scale investigation, we worked with junior dancers at a local high school.

2. Method

To make our speech sounds highly contrastive, we picked two of the most commonly reported syllables from the sound symbolism literature: /bu/ and /ki/ (e.g., D'Onofrio, 2014; Ramachandran & Hubbard, 2001). To give the dancer a more dynamic performance range, and to give viewers two types of acoustic information to base their judgements on, each syllable was produced in three rhythmic styles: a *short*, single articulation, a *long*, single articulation, and a *staggered* string of five short articulations, in which each short articulation was the same length as the short condition, and the total string was the same length as the long articulation (**Table 1**).

Table 1. Speech stimuli: two syllable types produced in three rhythmic patterns.

		Rhythm		
Speech		short	long	staggered
syllable	bu	bu	bu~~~~	bu bu bu bu bu
	ki	ki	ki~~~~	ki ki ki ki ki

A Singaporean bilingual speaker of English and Chinese produced each speech sound in each rhythm condition three times, using an animated gif as a timer. The clearest token of each sound type was selected for use in the test.

Speech-to-dance: motion generation

An experienced adult dancer was invited for a video recording session. She had several years of experience in Hip Hop and Street dance, with a current specialisation in Pop-n-lock. A single recording session was conducted at Nanyang Technological University.

The dancer was asked to produce one novel dance gesture for each of the six recorded sounds. The purpose of the experiment was explained, so that the dancer understood that later viewers would be guessing which movement was triggered by which sound. She was also instructed to use only meaningless, abstract motions based on the sounds she heard, rather than trying to depict meaningful shapes (e.g., tracing a letter shape). The dancer wore plain clothes appropriate for street dance. To remove the possibility that the dancer might inadvertently signal the identity of a speech sound with her face, the dancer wore a plain white mask.

All the speech sounds were played before video recording began, so that the dancer could familiarise herself with the sounds. One dance motion was recorded for each of the sounds. The dancer began each motion in the same position, and returned to that position at the end of her motion. She was allowed to review each recording, and re-take any video as many times as she liked. When the dancer was satisfied her motions expressed the sounds she had heard, the videos were trimmed into clips (see **Figure 2**), and edited to remove audio.

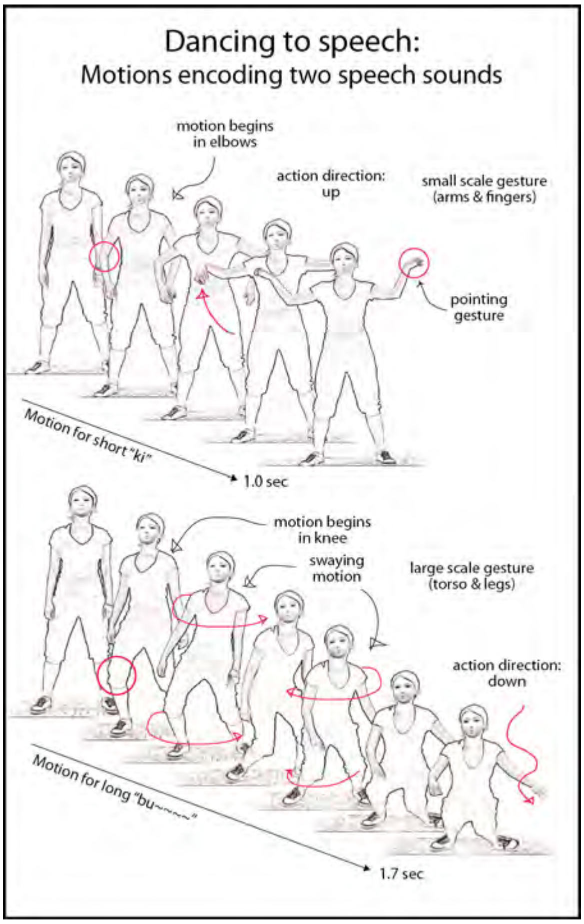


Figure 2. Illustration of two dance gestures produced in response to speech. Gesture durations measured from first frame of motion to last frame before returning to starting position.

Dance-to-speech: cross-modal matching

Participants in the cross-modal matching study were tested in three group sessions. Each dance motion was projected in front of the group, in silence, in a random order. Participants guessed which dance motion was produced in response to which speech condition, and marked their answers on a paper form. Before beginning the projection, participants had a chance to hear what each of the speech tokens sounded like. At the end of the experimental session, participants filled out a survey about their dance backgrounds. The whole procedure lasted a few minutes.

Junior dancers involved in dance classes at Nanyang Girls' School took part in the study after obtaining parental consent. The 20 participants were aged 13-18 years, and had dance experience ranging from 5 to 14 years. There were no reported hearing problems, and normal (or corrected-to-normal) vision. The study was conducted at the regular meetings of three dance classes held at the school. The junior dancers had no contact with our experienced adult dancer. The procedure was approved by the NTU Institutional Review Board, in collaboration with Nanyang Girls' School.

Predictions

We predicted that if our dancer was able to capture something about the sounds of speech in her motions *and if other students of dance share a similar understanding of the links between sounds and motions*, then junior dancers should be better-than-chance at guessing which motion was produced in response to the different speech sounds, even if their dance backgrounds were different, and even though abstract, meaningless speech is not normally used as a stimulus for dance.

Since the rhythm classes differed substantially in their large-scale temporal structure, and rhythm is typically incorporated into dance performance, we predicted that participants might be better at matching dance motions to rhythm categories, than they would be at matching dance motions to speech sounds (which differ in their fine spectral and temporal detail).

We also predicted that if acoustic properties of the different speech sounds shared a sensory congruence with the rhythm classes they were produced in, that some of the motions would be more guessable than others, because they represented a kind of superstimulus, in which each feature of the articulation enhanced its identity, making it easier to generate a holistic motion for the sound.

Finally, since dance experience generates functional changes in the way dancers' brains respond to the motions of other dancers (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005), we predicted that experience might modulate gesture/sound congruence, giving more experienced dancers better performance in the guessing task. On the other hand, even though dancers use gesture/sound congruence when they perform, it is possible that they generate motions which have a shared visual congruence for everyone, and practical experience in dancing may have little influence on how well dance viewers are able to match motions to sounds.

3. Results

Since each dance motion could be matched to any of the six speech stimuli, participants had a one-in-six chance (18%) of guessing correctly if they were picking purely at random – effectively the same as rolling dice. We found participants were significantly better than chance at picking the correct dance motion, with an overall accuracy of 56% ($t(19)=5.8$, $p<.001$, $d=1.21$ **Figure 3, Left**)

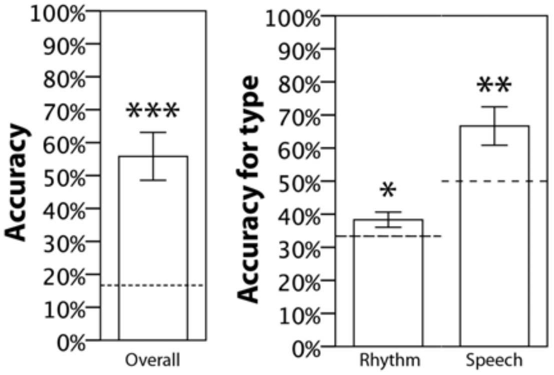


Figure 3. Average number of correct guesses per participant (**Left**), and shown separately for different acoustic features (**Right**). Chance values are dashed lines. Error bars +/- 1SE, ***p<.001, **p<.01, *p<.05.

We also looked separately at accuracy for rhythm, and accuracy for speech sound (**Figure 3, Right**). When allocating a dance motion to a rhythm category (regardless of which speech sound), participants were correct 38.0% of time, and this was significantly better than the chance value of one-in-three ($t(19) = 2.18$, $p < 0.05$, $d = .48$). When allocating a dance motion to a speech sound (regardless of rhythm), participants were correct 67.0% of the time, which was significantly better than the chance value of one-in-two ($t(19) = 2.87$, $p \leq 0.01$, $d = .64$). The effect sizes suggest that participants found it easier to guess the fine detail of the speech than the large scale detail of rhythm.

Stimulus Differences

When we compared the percentage of correct guesses for the different videos, some motions were clearly more guessable than others (**Figure 4**). Short 'ki' and long 'bu' were the most guessable (70%), followed by short 'bu' and staggered 'ki' (55%), with the worst performance for long 'ki' (45%) and staggered 'bu' (40%).

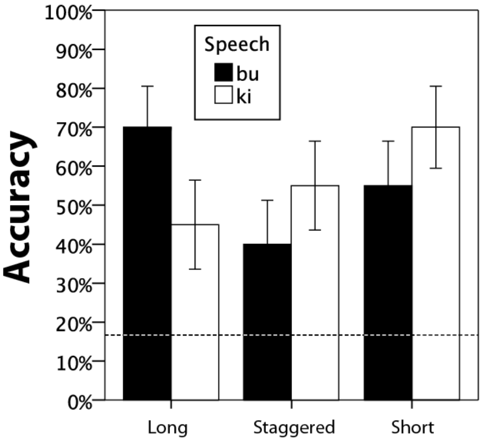


Figure 4. Average number of correct guesses for each stimulus. Chance values are dashed lines. Error bars +/- 1 SE.

As evident in the illustration of the dance gestures (**Figure 2**), 'ki' motions were short, jerky, upward motions, made with arms and hands, while 'bu' motions were slower, swaying, downward motions made with the whole torson and legs moving downward. This pattern fits well with the idea that within a single sensory modality, sensory alliances between different features of a signal can enhance the identity of the stimulus itself, acting as a kind of super-stimulus. For 'ki' the duration of the short rhythm was the best match for a sound typically mapped to small, jagged shapes (and here, small, jerky motions), while for 'bu' the long, smooth rhythm was the best match for a sound typically matched with large curved shapes (and here, large, smooth motions).

Individual differences across the population

Since we were interested in whether dance experience might influence how well junior dancers could guess the symbolism of these actions, we compared cumulative years of dance experience with overall task scores. **Figure 5** shows that there was no clear relationship between years of dance experience and performance in this guessing task (ρ(20) = .19, non-significant). This suggests that the performance we observed for the group taps into a general system of cross-modal correspondence shared by the general community (see **Fig 1**, "Most of us"), regardless of the duration of dance experience.

However, it is also interesting to note the cluster of six participants who correctly identified all six dance motions. This group is interesting for two reasons: Firstly, participants chose not to guess the same sound twice, meaning there was a one-in-21 chance of guessing all six motions correctly. This means, from our group of 20, we should expect one participant to guess all dances correctly purely by chance. Instead, we see almost a third of our sample achieving this high score.

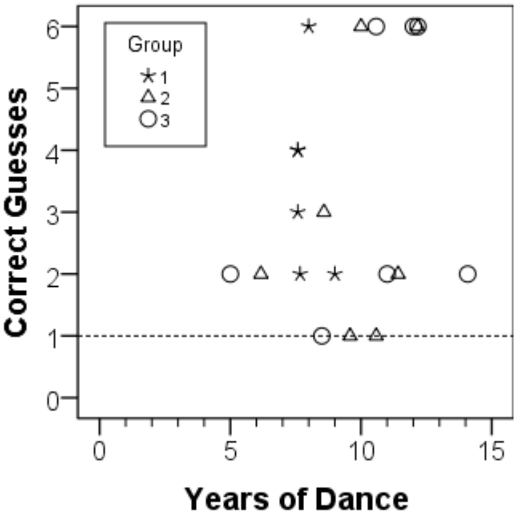


Figure 5. Years of dance experience and number of correct guesses for each participant, with different year groups shown separately. Chance value marked with dashed line.

Secondly, there is something of a performance gap between the response of these six, and the spread of responses across the rest of the group. This bimodal distribution suggests that we observing two groups – a general population with a normal spread of scores, and a separate cluster of high performers. These six could represent a cluster of high-sensory integration individuals, as proposed in the population model of sensory congruence: these could be artists in-the-making.

4. Discussion

In this small-scale test, we observed that an experienced dancer was able to convey sensory information about meaningless speech sounds to an audience using abstract dance gestures, in such a way that performance in the guessing game was above chance. This means that something about the sound structure was encoded in the detail of our dancer's motions, despite speech being a non-standard stimulus for dance.

The fact that the dancer's audience was able to extract relevant information is evidence for shared cross-modal abstractions between the dancer and her audience, and this did not appear to be driven by experience in the medium of dance.

Furthermore, we observed that the rhythm and the identity of a sound interact in making a stimulus 'more danceable', and/or making a video more guessable: when the internal features of the sounds 'go together' the stimulus works better in tests of this kind.

Within our group of junior dancers, we observed two separate groups: the majority displayed a normal distribution scoring from one-to-four correct guesses, and a separate cluster of six perfect scores. Even in this small-scale test, this clustering provides partial support for a model in which the majority of participants share more-or-less agreement of sensory interactions, but a smaller proportion of the population may feel these sensory interactions more deeply – and these people may be the ones who actively seek out the arts (hence, being over represented in the group).

However, since all of our junior dancers were taking part in dance as a school activity, it is possible that they have chosen dance for a variety of reasons other than their own 'feeling' for sound and motion. It therefore remains to be seen whether these results are representative of a more general population (including adults and non-dancers), and whether people with higher-than-normal rates of sensory integration are similarly over-represented among adult dancers.

Acknowledgements

Congratulations and wild applause go to junior collaborators Chearles Ming Huey Lisa and Farm Yi Wen Gina, whose enthusiasm for dance brought this project to life. They collected the data as part of the Singapore MOE's Humanities and Social Sciences Research Programme (HSSRP), with the support of their Teacher mentor Ong Lee Hua, at Nanyang Girls' High School. Thanks also to Lim Li Rong, who danced for us, and to Clara Chan for assisting with the filming.

This project was supported by a Nanyang Assistant Professorship Grant to SJS: *The Shape of Sounds in Singapore*.

References

Bremner, A. J., Caparos, S., Davidoff, J., Fockert, J. d., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition, 126*, 165-172.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex, 15*, 1243-1249. doi: doi:10.1093/cercor/bhi007

Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is Indigo Blue: Discovering the brain of synesthesia*. Cambridge MA: MIT Press.

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm *Language and Speech, 57*(3), 367-393.

de Saussure, F. (1911:1959). *Course in general linguistics* (1959 ed.). New York: Philosophical Library.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 1-12.

Köhler, W. (1929:1947). *Gestalt psychology*. New York: Liveright Publishing.

Ludwiga, V. U., Adachid, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans *PNAS, 108*(51), 20661-20665.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective & Behavioral Neuroscience*, *4*(2), 133-136.

Morton, E. W. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, *111*, 855-869.

O'Hanlon, E., Newell, F. N., & Mitchell, K. J. (2013). Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia. *Frontiers in Psychology*, *4*(755), 1-13.

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience*, *3*(3-4), 348-355. doi: DOI: 10.1080/17470910701563681

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The takete–maluma phenomenon in autism spectrum disorders. *Perception*, *42*, 233-241. doi: doi:10.1068/p7357

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 325-347): Cambridge University Press.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*(114). doi: http://dx.doi.org/10.1016/j.jecp.2012.05.004

Pirog Revill, K., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain & Language*, *128*, 18-24.

Ramachandran, V. S. (2003). The Emerging Mind. *Reith Lectures, BBC Radio Four*. Retrieved from http://www.bbc.co.uk/radio4/reith2003/

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(123-34).

Rothen, N., & Meier, B. (2010). Higher prevalence of synaesthesia in art students. *Perception*, *39*, 718-720.

Spector, F., & Maurer, D. (2013). Early sound symbolism for vowel sounds. *i-Perception*, *4*, 239–241. doi: dx.doi.org/10.1068/i0535

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception and Psychophysics*, *73*, 971-995.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science*, *21*(1), 21-25.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 43-44.

How well do Humans Capture the Sounds of Speech in Writing?

Nora Turoman, Suzy J. Styles*

Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

*suzy.styles@ntu.edu.sg

Abstract

A large body of research on connections between sensory modalities has shown that deep connections exist between sound and vision, such that people have a tendency to associate certain sounds with certain visual properties, including line-drawn shapes. While recognising the role of written language in audio-visual associations, previous research has largely considered written language a potential source of bias rather than a means of gaining deeper insights into underlying audio-visual associations. We looked to ancient and unfamiliar writing systems spanning recorded human history, to explore whether humans have tried to encode certain characteristics of speech sounds in the letters they created to represent them. Our findings have revealed that modern humans can correctly identify unfamiliar letters at levels higher than would be obtained by chance, and that scripts which encode a particular sound with a particular set of visual characteristics tend to have more correct guesses. This suggests that humans share certain correspondences between sound and sight, which transcend both geographical space and historical time. The present multisensory demonstration aims to provide an interactive experience of the powerful connection between sounds and written letters through a series of activities integrating vision, audition, touch and imagination.

Keywords: Sound symbolism, cross-modal processing, evolution of language

1. Experimental investigation

In the field of linguistic sound symbolism, there has been growing recognition that humans tend to match particular speech sounds to particular visual shapes (e.g., Sapir, 1929; Köhler, 1929; Ramachandran & Hubbard, 2001). Previous research has regarded sound-shape associations in writing as a potential source of bias. We have, in a series of experiments, found letters to be a fertile site for investigating links between vision and audition.

In three experimental investigations, participants saw pairs of ancient and unfamiliar letters representing /i/ (the 'ee' vowel in 'feet') and /u/ (the 'oo' vowel in 'shoe') derived from a systematic review of 'The World's Writing Systems' (Daniels & Bright, eds., 1996).

Results showed that modern participants were able to correctly guess the identities of ancient and unfamiliar letters at levels above chance. We also found that scripts with the highest numbers of correct guesses encoded certain sounds using certain visual characteristics (e.g., curvature, width, complexity and enclosure for /u/) while scripts with the lowest numbers of correct guesses did the opposite (e.g., straightness, tallness, simplicity, less enclosure for /u/) (c.f. D'Onofrio, 2014). See **Figure 1**.

The findings suggest that humans over geographic space and historical time have encoded something about the sounds of speech in the letters they choose to represent them, and that humans share correspondences between vision and audition which transcend space and time (Turoman & Styles, in preparation).

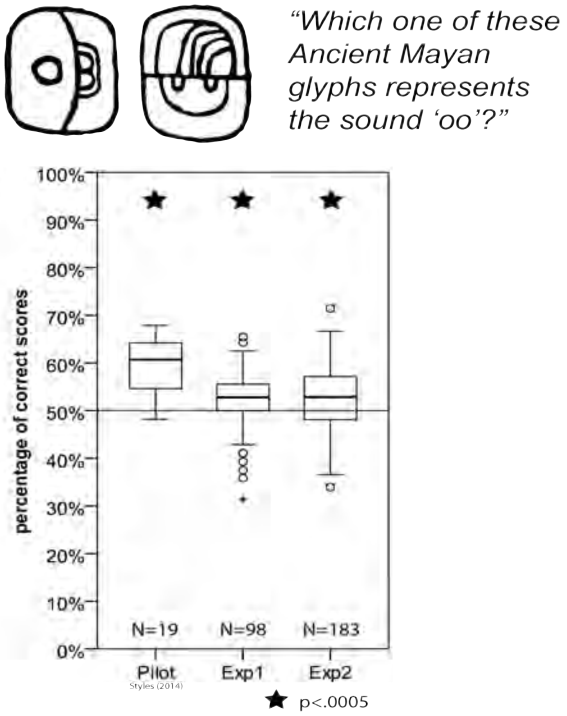


Figure 1. Example question and results

2. A demonstration

To provide an experience of the relationships between vision and audition encoded in writing systems of the world, we created two interactive multisensory demonstrations at *Si15 Soundislands Festival*.

Letter guessing game. Gallery visitors were invited to guess which letters represent which sounds in ancient and unfamiliar scripts, in an activity derived from the experiments mentioned above. Tactile objects depicting contrasting letter forms on opposite surfaces will be placed in the demonstration space, and visitors will be invited to arrange the tokens according to the sounds they think 'go best' with the letters. Depicted letter pairs included some of the most and least 'guessable' scripts obtained in the original experiments (above), allowing visitors' decisions to be compared with previously collected responses. Completed arrangements were compared with an answer key on-the-spot.

Collaborative glyph generation. Gallery visitors were be invited to take part in a collaborative glyph generation exercise, in which they can created their own letters, by drawing novel shapes that they feel best represented particular speech sounds.

The participants listened to a series of audio recordings of the following vowel sounds: /i/ (the 'ee' vowel in 'feet'), /u/ (the 'oo' vowel in 'shoe'), /a/ (the 'ah' vowel in 'arm'), and the high front rounded vowel /y/, often written as ü. This sound does not exist in English, but occurs in a number of world languages including German, French, Mandarin and Cantonese. Participants wore headphones, and followed an onscreen prompt to draw a novel glyph they feel best represents each of the vowels, on coloured paper pads.

Glyph creations were collated in a wall display at the festival, as part of a growing installation of sound-shape symbolism. They can be viewed at the *Brain Language and Intersensory Perception* (BLIP) Lab website, URL http://blogs.ntu.edu.sg/blip.

Acknowledgements

Supported by a Nanyang Assistant Professorship Grant to SJS: *The Shape of Sounds in Singapore*.

References

Daniels, P. T., & Bright, W. (Eds.). (1996). *The World's Writing Systems*. Oxford: Oxford University Press.

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. *Psychological Science*, *57*, 367-393.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright Publishing Corporation.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synesthesia - a window into perception, thought and language. *Journal of Consciousness Studies, 8*, 123-34.

Styles, S. J. (2014). 'What can ancient and unfamiliar scripts tell us about sound symbolism?' International Multisensory Research Forum, June 2014, Amsterdam.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*, 225–239.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 45-49.

Super-Normal Integration of Sound and Vision in Performance

Joel D.F. Lim, Suzy J. Styles

Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

joel0008@e.ntu.edu.sg, suzy.styles@ntu.edu.sg

Abstract

In the world of rock, blues, and popular music, guitarists who pull faces when they play are somehow more entertaining to watch – as if there is an agreement between the performer and the audience that certain facial gestures enhance the expression of the music at key moments. How does this agreement come about? What is it that viewers and performers share? Research in the field of cross-modal perception has demonstrated that humans share systematic linkages between the senses, including vision and audition. For example, people link higher pitched sounds with smaller, paler, and higher-up objects than lower sounds (e.g., Mondloch & Maurer, 2004). Yet few investigations have explored how these linkages relate to performance in the creative arts. One theory suggests that artists have stronger functional connectivity between sensory areas, and are therefore able to invoke or translate sensory experiences across modalities, using sensory metaphors, and strategic sensory exaggeration (Ramachandran, 2003). To test if viewers share a sensory experience with performers, we asked participants in an online quiz to guess which of two guitar faces 'went with' particular moments in guitar solos. One-hundred and seventeen volunteers guessed guitar faces at rates higher than would be predicted by chance, indicating that people were able to draw meaningful information about music from still images of the guitarist's face. In addition to presenting the results of the study in a poster format, this demonstration will give visitors a chance to engage with the interaction between musical performance and performative gesture, in two activities.

Keywords: musical performance, orofacial gesture, crossmodal correspondence.

1. Introduction

Some of the best guitarists worldwide contort their faces unconsciously as they perform their solos, stretching and twisting their faces, sometimes to amusing effect. Compared to a deadpan performance, these contortions seem to make for a more entertaining audience experience. We know this phenomenon as 'guitar-face'.

Music and gesture

Humans appear to be naturally inclined to move their bodies to music. Infants sway, move and dance to rhythms (Zentner & Eerola, 2010). Even 'sophisticated' adult concertgoers respond to the dynamics of musical performance by dancing, nodding their heads, or making gestures along with the music, with face or limbs. These music-related gestures demonstrate the intrinsic sensory linkages between audition and the somatosensory experience of motion.

While we often think of performers as trained specialists in their chosen field, in "air guitar championships", contestants compete to best co-ordinate sound-based gestures and physical expressions at a complex level. It is note-worthy that some of the most successful "air guitars" have not played actual guitars before (Godøy, Haga & Jensenius, 2006). As such, there appears to be a shared agreement of what a good guitarist looks like when they play certain sounds, regardless of musical background.

The spontaneous production of sound-based gestures point to certain similarities in the way musical sounds are understood by both the novice and the expert: since audiences and air-instrumentalists are able to appreciate and/or emulate sound-based performance gestures, this suggests a sensory link between the auditory and the visual components of a live musical performance.

When both the performer and the audience share a sensory mapping between the two, then their shared experience may affirm, or even enhance the dynamics of musical performance, as it is being played.

Cross-modal correspondences

In the field of cross modal research, natural mappings between sensory experiences have been well documented. For example, both adults and 5-year-old children have been shown to match the intensities of loudness and brightness across modalities (Bond & Stevens, 1969), and adults and infants match higher pitched tones to smaller, paler, objects in higher spatial position (Mondloch & Maurer, 2004). In the domain of linguistic sound symbolism, Edward Sapir observed that majority of people tend to associate nonsense words like "mal" with large objects, and "mil" with small objects; pointing out an association between the vowels /a/ and /i/ and the size of an object (Sapir, 1929).

Ramachandran and Hubbard (2001) proposed that this kind of sound symbolism arises out of cross-modal correspondences between not only sensory maps (e.g. vision, audition) but also the motor maps used for speech. Using the nonsense words "Bouba" and "Kiki", they found that majority of people pair the word "bouba" with rounded shapes, while pairing the word "kiki" with sharp, angular shapes. The authors suggested that the rounded shape of the lips in the articulation of a word-form like 'bouba' gives rise to the association with curved forms. Hence, natural mappings may exist between facial gestures and features of the audio-visual landscape – in particular, the round lips of 'oo' go with curves, and with low sounds, while the thinly spread lips of 'ee' go with spikes and high notes.

Synaesthesia and the artist

Ramachandran and Hubbard (2001) proposed an explanatory model of cross-modal correspondences in language, using synaesthesia as a lens. Synaesthesia is a condition in which the stimulation on one sensory modality results in more than one sensory experience – for example, coloured hearing or tasting shapes (Harrison & Baron-Cohen, 1997), Ramachandran and Hubbard proposed that synaesthesia arises out of exaggerated patterns of cross-talk and cross activation between sensory brain regions (Ramachandran & Hubbard, 2001), and that this extra sensory excitation may make an individual more likely to be adept at connecting seemingly unrelated concepts. In his BBC Reith Lecture: "Purple Numbers and Sharp Cheese", Ramachandran pointed out that synaesthesia is common among artists, writers and poets, whose works display skillful pairing of concepts, emotions, or sensory experiences that appear to be unrelated. It is suggested that while the neurological knack for understanding metaphors is present in all of us, the creative side is much stronger in artists and synaesthetes due to stronger cross-activation.

The creative arts utilize the vehicles of exaggeration, and distortion of stereotypes to trigger sensations of familiarity and surprise in the audience. For example in caricature drawings, unique features of a person's face are emphasized and shifted away from the facial norm, until these exaggerations seem more like the person than the actual person themselves (Ramachandran, 2003). Similarly, in animal behavior experiments, Tinburgen and Perdeck (1950) demonstrated that creating an exaggerated combination of visual features could create a "supernormal" stimulus that would elicit greater feeding responses in gull chicks, than would their own parents' beak. It may follow then that artists and performers who successfully entertain their viewers, have managed to create super-normal sensory excitation: performances that successfully combine sensory stimulation across different modalities may appeal to the hardwired nature of cross-modal human perception.

Despite the persuasive nature of these ideas, the notion of supernormal inter-sensory stimulus in artistic communication has not yet been translated into a testable hypothesis (c.f., Styles, 2016, this volume).

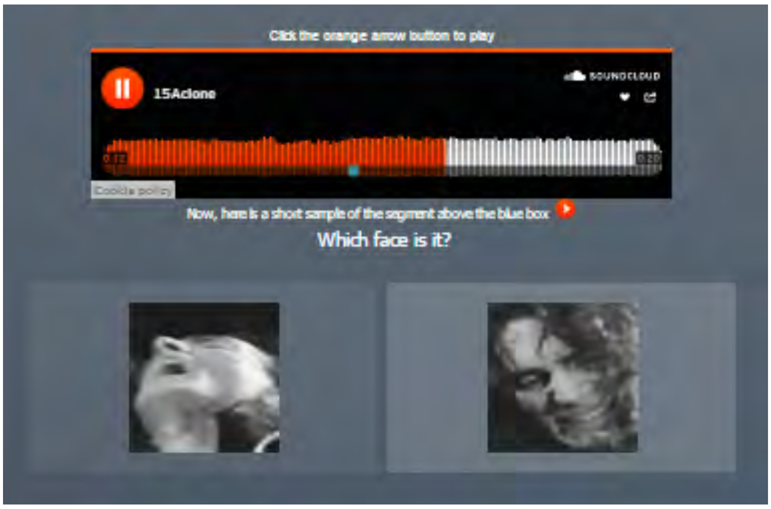
Here, we ask the very testable question: If guitar soloists incorporate guitar-faces (visual) into their instrumental performances (auditory), and if audiences implicitly integrate the combination of information from different modalities, *and if those senses combine in a satisfactory manner* (i.e., if the guitar face is good), then it follows that the audiences will be able to uncover something about the music from the face that was pulled at the time.

To test this hypothesis, we conducted an online quiz to test whether people experience cross-modal correspondences for arbitrary selected moments from live guitar solo performances, by seeing if they are able to match the correct guitar-faces to their respective moments in the guitar solo, at rates above chance.

2. Experiment

We analysed the performance of 117 volunteers who underwent an online guitar-face quiz hosted on a Brain, Language and Intersensory Perception (BLIP) Lab webpage (URL http://blogs.ntu.edu.sg/blip/). Participants ranged from 18-67 years of age and included both people with (49.6 %) and without (50.4 %) prior music experience or training in music theory or performance. Of these, 46.2 % have experienced playing a guitar. Participants resided in various countries.

Figure 1 shows a sample layout for a single quiz item. Participants were presented with a guitar solo, and were asked to pay attention to a specific moment marked by the sudden appearance of a blue box on a time progress display. A short 100ms sound slice from the moment of the guitar-face was also provided for them to play for their reference. Two faces were displayed side-by-side (one target and one distractor) and participants were instructed to guess which of the faces was produced at the specified moment of the guitar solo, for 28 quiz items.



Click the orange arrow button to play

Now, here is a short sample of the segment above the blue box

Which face is it?

Figure 1. Example quiz question with scrolling music bar at top, guitar-face 'moment' marked by a blue box in the timeline, and two faces drawn from the same solo.

3. Results

Our results show that participants are able to draw meaningful information about the musical performance by simply looking at the guitar-face and therefore be able to correctly match it to its corresponding moment on the contour of the musical solo above chance levels ($M = 15.12$, $SD = 2.95$, $t(116) = 4.108$, $p < .001$).

Figure 2. The most recognisable guitar face (left) with its distractor on the right.

Notably, the most recognisable guitar-face (see Figure 2), had a facial gesture resembling the articulation of the vowel sound 'oo' (/u/). The corresponding moment for the target guitar face was preceded by a downward undulation of the music contour, and followed by an upward inflection of the pitch contour. This "valley" of low pitch in the music contour constitutes a compelling cross-modal correspondence to lip shape, and may form a type of super-stimulus.

Discussion

Our results show that in spite of being presented with no information about other contextual and visual aspects of the performance, the ability of the participants to match the guitar-face correctly with their corresponding moments still remained above the chance level.

The fact that participants were able to draw meaningful information and make pinpoint deductions at about the musical performance simply by looking at a guitar face and tracing the contour of a musical solo, suggests a strong correspondence between the motor maps underlying facial expression and the dynamics of music in live performance, which may then provide the building blocks for a supernormal stimulus in live performance; created and conveyed by the performer to be received and appraised by the audience. This may explain a shared experience between performer and audience that is multi-modal in nature; the performer and the artist have a natural basis for communicating and understanding linkages between facial gestures and musical dynamics.

5. Si15 Demonstration

In August 2015, we gave a public demonstration of the research at Singapore's ArtScience Museum, organised by Si15 Soundislands Festival (URL http://soundislands.com/si15). A gallery space was designed to engage members of the public in two interactive explorations of co-music gestures: An ongoing interactive data array (the guitar-face board), and a live air-guitar competition.

Collective guitar-face data array

The demonstration included the opportunity to listen to a series of guitar solos, accompanied by time-progress bars, with selected 'moments' of performance for evaluation. People were asked to select from a fixed set of schematic face shapes, which face 'goes best' with the performance for evaluation. People were asked to select from a fixed set of five schematic face shapes, and to indicate their response using one of five rubber stamps, on a response slip. These response slips were compiled into a live display.

Each face-shape stamp was assigned a different ink colour, such that the growing display of public selections created a colourful data-array as the day progressed. The interactive nature of the exhibit provided a tangible insight into how much an individual's sensory experience is shared with other members of the general public.

Guitar-Face Competition

Towards the end of the demonstration, a competition was held, with performers asked to focus particularly on creating the best *guitar-face*, as highlighted by performing with their face inside a picture frame. This competition provided a light-hearted collective experience of what makes a 'good' guitar-face.

Acknowledgements

This project was supported by a Nanyang Assistant Professorship Grant (4081215) to SJS: *The Shape of Sounds in Singapore*.

6. References

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. Perception & Psychophysics, 6, 337–339.

Godøy, R.I., Haga, E., & Jensenius A.R. (2006). Playing "air instruments": Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, & J.-F. Kamp (Eds.), Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers, Volume 3881/2006, pp. 256–267. Berlin Heidelberg: Springer-Verlag.

Harrison, J.E., & Baron-Cohen, S. (1997), Synaesthesia: A review of psychological theories', in Baron-Cohen, S & Harrison, J.E., (Eds.,) *Synaesthesia: Classic and Contemporary Readings*, pp.173-182. Oxford, England: Blackwell.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. Cognitive, Affective & Behavioral Neuroscience, 4(2), 133-136.

Ramachandran, V.S. (2003). Purple Numbers and Sharp Cheese. Retrieved April 10, 2015, from http://downloads.bbc.co.uk/rmhttp/radio4/transcripts/20030423_reith.pdf

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. Journal of Consciousness Studies, 8, 3–34.

Sapir, E. (1929). A study in phonetic symbolism. Journal of Experimental Psychology, 12, 225–239.

Styles, S.J., (2016), The language of dance: Testing a model of cross-modal communication in the performing arts. IMCA Array, Volume 2016, Special Issue: Proceedings of Si15.

Tinbergen, N. & Perdeck, A. C. (1950). On the stimulus situation releasing the begging response in the newly hatched herring gull chick (Larus argentatus Pont.). Behaviour, 3, 1–39.

Zentner, M. & Eerola, T. (2010). Proceedings of the National Academy of Sciences, 107(13): 5768–5773.

array
ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 50-51.

An Implicit Association Test on Audio-Visual Cross-Modal Correspondences

Nan Shang*, Suzy J. Styles

Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

*shan0049@e.ntu.edu.sg

Abstract

Automatic connections between sounds and visual shapes have been documented for some time (c.f., Spence, 2011). We replicated audiovisual correspondences with simple linguistic sounds /i/ and /u/, this time produced in the lexical tones of Mandarin Chinese, using a modified version of the implicit association test (IAT). Although congruent blocks were significantly faster than incongruent ones (p < .001), no effect of tone congruence was observed. Since tone was an unattended stimulus dimension, we argue that attention modulates sensory congruence in implicit association tasks of this nature.

Keywords: multisensory perception, sound symbolism, implicit association test

1. Introduction

There is ample evidence of cross-modal correspondences between visual shapes and speech sounds in the literature. This kind of non-arbitrary mapping was first observed by Köhler, whose participants showed strong preference to pair a round shape with 'maluma' and an angular shape with 'takete' (Köhler, 1929 : 1947). D'Onofrio (2013) also documented a front vowel /i/-pointy, back vowel /u/-curvy preference pattern. However, their pseudowords consisted of vowels and consonants, giving rise to a confounding.

In addition to the sound of phonemes, non-linguistic pitch also gives rise to associations with visual shape: high pitch is matched to angular shapes; low pitch is matched to round shapes (Marks, 1987). Only one study has previously investigated cross-modal mapping preferences for linguistic tones. Shang and Styles (2014) report both Chinese dominant and balanced Chinese-English bilinguals share a Tone 1-curvy, Tone 4-pointy mapping tendency. See **Figure 1**. However, these results have only been seen in tasks where participants are asked to think about the relationship between the shape and the sound, meaning that their responses could reflect something 'strategic' or a kind of introspection, rather than an 'automatic' process.

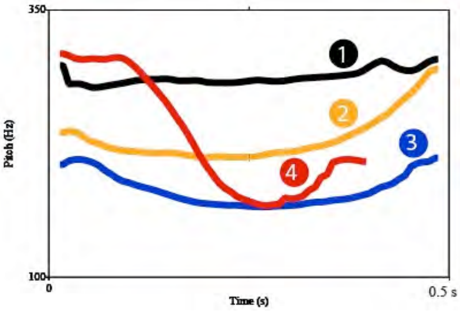


Figure 1. Time by frequency plot of 'oo' uttered in four Mandarin Chinese tones

The implicit association test (IAT) can overcome the above-mentioned ambiguity. Parise and Spence's (2012) modified version of the IAT showed that participants' performance improves when the set of stimuli assigned to a given response key share a cross-modal mapping (the congruent conditions), as compared with conditions in which a set of unrelated stimuli are assigned to the same response key (the incongruent conditions). Given the IAT's advantage in probing implicit relationships, we explored the associations between visual shapes and simple vowel sounds articulated in the lexical tones of Mandarin Chinese, using a modified version of the IAT for the first time.

2. Methods

Eighteen participants from Nanyang Technological University participated in the experiment for course credit. Sixteen were Chinese-English bilinguals. The other two were bilinguals in English and other languages.

In terms of pitch change, Tone 1 is stable and Tone 4 changes much more than Tone 1. According to the 'pitch change hypothesis', the greater the pitch change within a Mandarin Chinese tone, the pointer the shape of its associations (Shang & Styles, 2014). We therefore expect Tone 4 would facilitate congruent matches for Tone 4 with the pointy shape, while Tone 1 would facilitate congruence of Tone 1 with the round shape (**Figure 2**).

Figure 2. A congruent vowel-shape mapping

Counterbalanced across the study, were two tones of Mandarin Chinese (Tone 1 & Tone 4). However, since the task instructions asked people to press a key in response to which vowel they heard ('ee' or 'oo'), the congruence of the tone to the visual shape can be considered to be an implicit variable.

3. Results and discussion

We detected a main effect of vowel congruence ($p < .001$), suggesting the congruent blocks (/u/-curvy and /i/-pointy) were significantly faster than the incongruent blocks (/u/-pointy and /i/-curvy). No further effects or interactions with tone were observed.

The vowel congruence in our Chinese-English bilinguals is consistent with previous English-language vowel-shape matching (e.g., D'Onofrio, 2013) and an explicit vowel-shape mapping task in Chinese-English bilinguals (Shang & Styles, 2014), suggesting the Chinese-English bilinguals showed a strong effect of vowel identity. However, no tone effect was observed.

In our task, tone was embedded as an unattended dimension in the current IAT task, and the attentional demands of the task may have been too great to allow tone congruence to be observed. Ongoing investigations will inform us of whether tone congruence will emerge in IAT tests when it is the attended dimension.

Acknowledgements

Supported by a Nanyang Assistant Professorship Grant to SJS: *The Shape of Sounds in Singapore*.

References

D'Onofrio, A. (2013). Phonetic detail and dimensionality in sound-shape correspondences: Refining the Bouba-Kiki paradigm. *Language and Speech*, first published on March 24, 2013, doi: 10.1177/0023830913507694.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384–394.

Parise, C. & Spence, C. (2012) Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, 220(3-4), 319–333.

Shang & Styles. (2014) 'Cross-model correspondences between visual shape and Mandarin Chinese tone in speakers with different language backgrounds'. Poster presentation at 15th *International Multisensory Research Forum*, 2014, Pisa.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015 pp 52-53.

Pre-conscious Automaticity of Sound-Shape Mapping

Shao-Min (Sean) Hung[1†], Suzy J. Styles[2Φ], Po-Jang (Brown) Hsieh[1Ψ]

[1]Neuroscience and Behavioral Disorders Program,
Duke-NUS Medical School, Singapore

[2]Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

[†] konaes@gmail.com, [Ψ]pojang.hsieh@duke-nus.edu.sg [Φ]suzy.styles@ntu.edu.sg

Abstract

The bouba–kiki effect depicts a non-arbitrary mapping between specific shapes and non-words: an angular shape is more often named with a sharp sound like 'kiki', while a curved shape is more often matched to a blunter sound like 'bouba'. This effect shows a natural tendency of sound-shape pairing and has been shown to take place among adults who have different mother tongues (Ramachandran & Hubbard, 2001), pre-schoolers (Maurer, Pathman, & Mondloch, 2006), and even four-month-olds (Ozturk, Krehm, & Vouloumanos, 2013). These studies therefore establish that similar sound-to-shape mappings could happen among different cultures and early in development, suggesting the mappings may be innate and possibly universal. However, it remains unclear what level of mental processing gives rise to these perceptions: the mappings could rely on introspective processes about 'goodness-of-fit,' or they could rely on automatic sensory processes which are active prior to conscious awareness. Here we designed several experiments to directly examine the automaticity of the bouba-kiki effect. Specifically, we examined whether the congruency of a sound-shape pair can be processed before access to awareness?

Keywords: automaticity, sound-shape mapping, bouba–kiki effect

1. Experiments

To directly examine whether sound-shape mapping (i.e. bouba-kiki effect) happens prior to conscious awareness, we designed a first experiment (Exp. 1) in which the congruency was defined by the relationship between a non-word and a shape (Ramachandran & Hubbard, 2001b), that is, the non-word "kiki" was congruent with the angular shape, while the non-word "bouba" was congruent with the curvy shape. We blended the non-words into the shapes and presented the pairs with a series of flashing colourful squares in a dichoptic setup to suppress the visibility of the target pairs. (see **Figure 1**). The technique is referred to as 'continuous flash suppression'; for details, see Tsuchiya & Koch (2005). The time taken to see the pairs was measured, and the results exhibited a congruency effect: congruent pairs broke suppression and reached conscious awareness faster, suggesting that prior to being consciously aware of the visual stimulus, the phonology of the non-word has been extracted and matched with the shape.

Furthermore, we tested whether the congruency effect depended on the sound-shape consistency or was simply a result of the visual similarity between the non-words and shapes. In two further experiments (Exp. 2a and 2b), we implemented unfamiliar letters from the West African *Vai* Script. They were chosen from a database of 56 ancient and unfamiliar scripts that have been previously tested for their sound symbolic properties, and show neutral symbolism for linguistic contrasts (see **Figure 2**) (Styles, 2014). The congruency of the letter-shape pairs was determined by the training received prior to the main experiment. For instance, a letter paired up with the sound "kiki" was deemed to be congruent with the angular shape. Again, the "congruency effect" was found. Crucially, results showed that the congruency effect stemmed from the relationship between the shape and the sound represented by the letter, and not from shared visual characteristics.

Taken together, the three experiments suggest that sound-shape mapping can happen automatically, and that sensory congruency facilitates the access to conscious awareness.

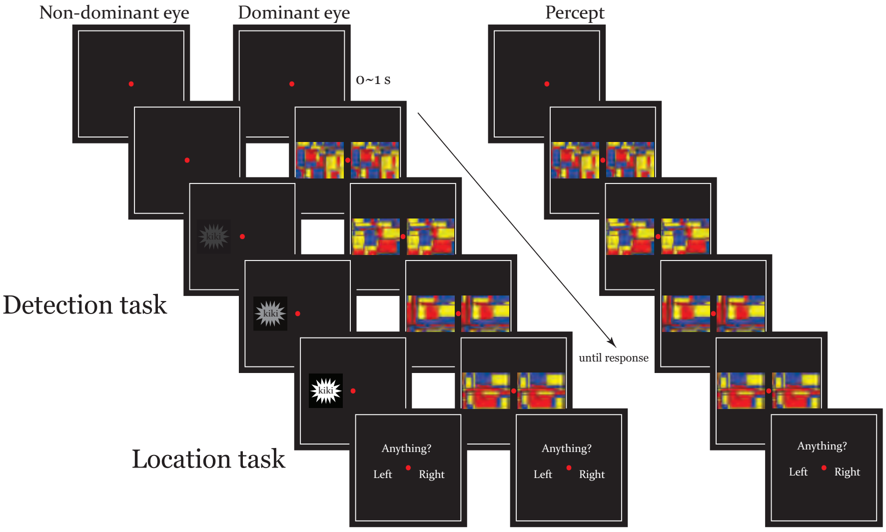


Figure 1. Illustration of the experiment procedure and stimuli in Exp. 1. After a fixation period ranging from 0 to 1 second, a series of 'colourful squares' flashing at 10 Hz was presented to the dominant eye. Meanwhile, the target stimulus was presented to the non–dominant eye either on the left or right side of the fixation point. The contrast of the target stimulus ramped up from 0 to 75% in 10 seconds. The target stimulus stayed on the screen up to 10 s or until the visibility was indicated. Participants were required to press a button immediately after any part of the target stimulus became visible (Detection task) and then report the location of it (Location task).

Figure 2. Left panels: example congruent stimuli in Exp. 1. Right panels: example stimuli in Exp. 2a (congruent) and 2b (incongruent), where the congruency of a pair depended on the training received. One word was paired up with the sound "bubu" in Exp. 2a and the sound "kiki" in Exp. 2b.

References

Köhler, W. (1929). *Gestalt Psychology*, New York: Liveright.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound–shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Journal of Experimental Child Psy chology. *Journal of Experimental Child Psychology*, *114*(2), 173–186.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Styles, S. J. (2014). 'What can ancient and unfamiliar scripts tell us about sound symbolism?' *International Multisensory Research Forum*, June 2014, Amsterdam.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*, 1096–1101

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 54-60.

Colors in Silence – The Journey of a Synesthete in Understanding the Senses

Rachel Chen Siew Yoong[†]

[*]Nanyang Technological University

chen0916@e.ntu.edu.sg

Abstract

This paper will explore how a synesthete, RC, perceives colors, shapes, and textures in silence. RC has sound-to-color synesthesia, where at least one color is associated with each pitch on the Western music scale. Listening to silence strips all sound down to the bare minimum in terms of color or texture. As discovered through weekly meditation in a "Deep Listening" class, RC places a pitch to everything at least audible when plunged into silence, trying hard to capture the colors of every available sound. The silence also forces a more pronounced tactile sense that RC tries to grasp together with the colors. Every experience with the sound however, is present but muddled, and difficult to understand. Silence becomes an uncomfortable world of uncertainty and it becomes vital to grasp the visual and tactile nuances that are a part of it. This paper reflects RC's progression through silence towards an understanding of her senses. *What is silence? And how does the synesthete grapple with it?*

Keywords: synesthesia, perception, crossmodality

1. Introduction

Synesthesia can be defined as a conscious experience of "sensory attributes that lead to involuntary experiences in a second sensory pathway" (Cytowic, 2002). An individual with synesthesia would have a specific sensory experience when encountering a different sensory experience. A common example of a synesthesic experience can include grapheme-to-color synesthesia, where distinct colors are experienced or associated with specific graphemes.

The author of this paper, RC, came to the topic through her own personal experience with synesthesia. She has auditory-to-color synesthesia, where every pitch triggers an automatic sensation of a distinct color. RC also perceives every Western musical scale or chord as a 'set' or a 'blend' of colors. Her parents and music teacher first discovered her synesthesia at the age of 6, when she started playing the violin. She had complained about the tuning of a certain string not being "red" enough, and was then tested for synesthesia. She also has perfect pitch.

The colors she perceived as a child have remained the same as an adult. In her later years however, she discovered that she the same pitches she hears across different instruments have varying textures, and sound that has no specific pitch is high on occurring textures, but continues to maintain some association to color.

She describes these associations as being the most clear for piano or synthesised sounds, where each note can be played percussively and concisely, and relatively less clear for other instruments or mediums such as the voice or trombone. This could be due to her early exposure to instrumental timbres. For instruments with such different timbre, the synesthete experiences the timbre as texture, which interferes with the color of it. In these cases, a tactile sensation is a more salient association with every note played. It is therefore the case that some sounds are more strongly associated with color while maintaining an association with texture, and other sounds are more associated with texture, while maintaining an association with color.

2. A procedural overview

The following sections begin with a description of RC's exposure to silence over a period of three months, and her personal account of the experience. These will be based on a series of journal entries over the course of 3 months. This is followed by a color chart, in which RC has selected a color to match each musical note over three octaves. Secondly, a table of texture mappings will be presented. Finally, color and texture will be compared in terms of salience.

3. Sounds in silence

RC underwent one semester of a 'Deep Listening' class given by Pauline Oliveros. (Oliveros, 2005, p. 2) where she practiced weekly listening meditations, which involved listening to 10 minutes of silence. After each session, she wrote a journal entry about her experience and thoughts. Oliveros (2005), in a description of the class, states that the journal serves as a "record of progress for each individual" (p. 3).

RC's deep listening environments can be divided into three main categories; firstly, the music room at West Hall in Rensselaer Polytechnic Institute, secondly outdoors, and last, rooms in the *Experimental Media and Performing Arts Center* (EMPAC). These environments gave RC a variety of 'silent environments' to be immersed within.

The music room at West Hall can be described to be a low-noise room, and contained a consistent, quiet hum that was low in pitch. Outdoors, a wide variety of quiet sounds could be heard – slight chatter, birds singing, breeze blowing etc. In the soundproof rooms within EMPAC, there was usually almost no sound at all from the environment, so quieter sounds were easily discernible. Small sounds of distant humming could be distinctly heard on occasion.

These low-sound environments made RC question the definition of 'silence'. She writes in her journal entry (29/2/12):

"*I can't hear silence. I try hard to listen to it and find it but hear everything else instead. Should I try blocking a sound out, another sound replaces it. Is silence even possible*?"

In a discussion of the absence of sound, Hodkinson (2007) draws upon John Cage's composition *4'33"* to illustrate silence as an object of perception. It states that in having the performers remain silent, the performance is then transposed onto audience members in "shifting perception toward other sound" (Hodkinson p. 51), a phenomenon that also occurred with RC when faced with deliberate silence.

RC found that even in extremely quiet environments, sounds were distinct and obvious, describing all sounds as 'sharp' – these quiet environments brought out the subtlest sounds. This is interesting for a synesthete because it strips sound down to its barest, bringing to attention the smallest sounds with no other distraction. These become clear or distinct enough for the synesthete to visualise.

'Silence' for RC therefore presents unusually prominent sounds that are either color-salient or texture-salient. Sometimes however, both ring out, and these associations, although very present and real, become hard to grapple and understand: to 'catch'. Silence becomes uncertain to the synesthete since it is difficult to grasp the visual and tactile nuances of such sounds.

4. Color-mapping

To aid with the analysis of her silence, RC made a pitch-color chart for three octaves on synthesised piano pitches.

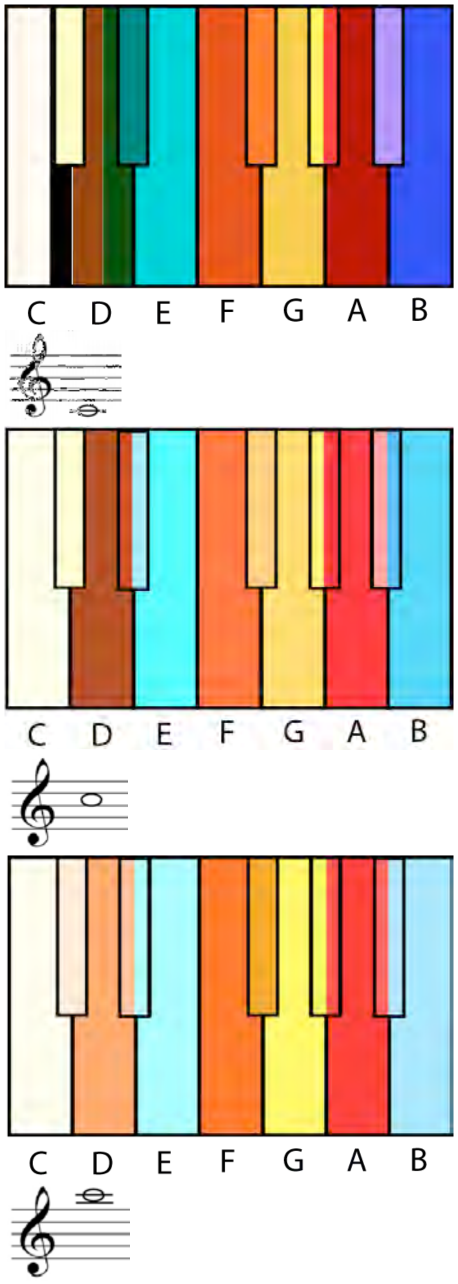


Figure 1. Pitch to color mapping for three octaves (from top to bottom in increasing pitch). Colour available in online pdf.

Some notes were found to have ambiguous colors, or to reflect two colors based on the musical context of its presentation. Pitches across octaves seem to share the same color with different shades and hues – lighter going up the pitches, and darker going down the pitches. The association of higher pitches with lighter colors, and lower pitches with darker colors is noted to be a common tendency in synesthesia literature (Cutsforth, 1925; Riggs and Karwoski, 1934; Whitchurch, 1922; Zigler, 1930; in Ward et al., 2006). Both synesthetes and non-synthetes have been found to exhibit this trend, and **Figure 1** above is consistent with these findings (Ward et al., 2006, p. 268).

It is however, worth noting that RC had problems matching the appropriate colors to the relevant pitches. In her reflections she writes (11/4/12):

"I tried my best with color-matching, but the RGB colors are missing a certain dimension to them. Especially as the pitches increase, it is difficult to express the lightness in color. Making red more "light" makes it look pink even though it isn't, and "B" and "E" seem to blend together although "E" really is more turquoise."

Perhaps a next step in matching color to pitch could involve the use of paint, which would allow for the manual mixing of colors in a way that RGB colors cannot. In the meantime, these colors must be seen as portrayals, and some amount of inconsistency is inevitable.

5. Textures in silence

To aid with understanding her perception of texture within sounds, RC kept a record of all of the distinct sounds she heard during her weekly deep meditation, and drew textures to represent each of the sounds. A summary of the textures are presented in **Figure 2** below.

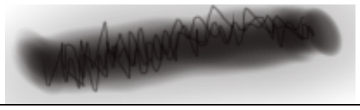
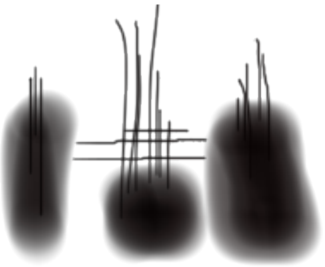
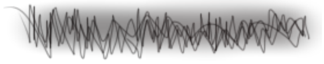
Projector humming 1		*This was a consistent stream of thickness with small and even 'static-like' waves in the middle.*
Projector humming 2		*This was soft and had a fuzzy quality to it. I felt that I could even knead it with my hands – it felt flexible in that way and open to potential change.*
Heater humming		*This produced a consistent sound that was light and thin. There were sharp spokes within the texture.*
Construction work		*This had huge chunks in them and had these lines above them that crossed one another every now and then, but were mostly vertical.*
Stomach grumbling		*This came in a variety of shapes and forms, but all with this sort of flow. It was a scribbly, springy texture.*
Cough		*This was rough and round and small.*
Phone Vibration		*This was consistent and dense and messy at the same time*
Rustling of Leaves		*There was a scrunchiness to this texture and it felt malleable.*
Birds Singing		*This was smooth and delicate, despite the interruptions between the chirpping. The sound quality was textured this way.*

Figure 2. Sounds and depictions of texture.

6. Pitch and texture: saliency and interaction

Of all the distinct sounds that RC experienced, the ones that can be described to have a clear pitch were:

1. projector humming;
2. heater humming;
3. birds singing;
4. phone vibration.

These presented clear colors or a stream of colors that either interfered with, interacted with, or overwhelmed the tactile nuances of the sound.

On one occasion, the humming sound of a projector was described to have 'a mix of pitches' in it, "namely A♮-B♮ and E♮-E♭, blended together in a pulse three or four times faster than my heartbeat" (**Figure 3**). This instance was not affected by texture, and reflects a pitch-salient sound. Although there was a distinct texture to it, the pitch was much more salient, and both texture and pitch could coexist, not affecting the other.

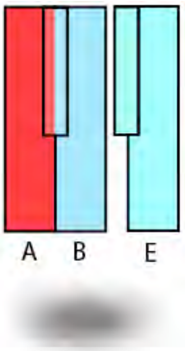


Figure 3. Projector humming (first occasion): color and texture.

However, another projector humming sound produced a different effect. As quoted from RC's accounts, "usually B♮-B♭ is a deep blue, but this time because of its deep and intense humming, I identified a very dark blue, almost black." This instance is different from the previous; the texture changed the color of the sound; even though RC still heard it as the same pitch. It is thus a case of interference (**Figure 4**).

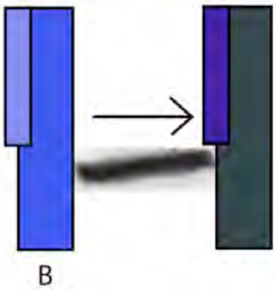


Figure 4. Projector humming (second occasion): color transformation by texture.

The heater humming had a high pitch; between B♮ and C♮ (**Figure 5**). According to previous-ly collected color mappings, this was supposed to be blue, or a shade of cream white. The texture of the heater changed the color of the sound to a shade of pinkish-red.

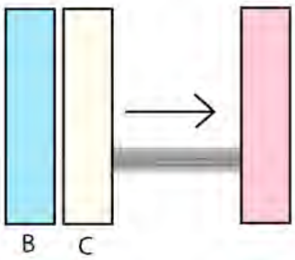


Figure 5. Heater humming: color transformation by texture.

Two other sounds presented both distinct pitches and texture: the birds' singing, and phone vibrations. The bird's singing had very dinstinct pitches in place, and they were easily dicernible because they contained light and delicate texture. Phone vibrations tended to be made up of several pitches at once, and these pitches were usually individually detectable without any effect from their texture (**Figure 6**).

Figure 6. An example of the colors of phone vibration

Although the textures of phone vibrations were dense, they contained some amount of 'clarity' – they were made up of discernible lines that were packed together. This stands in contrast to a dense texture, such as the first projector humming (**Figure 3**), where the texture was much denser and there were fewer lines since all were "melded together in a blurry mess" (journal entry: 11/4/12).

A summary of textures showcased in this section is displayed in **Figure 7**.

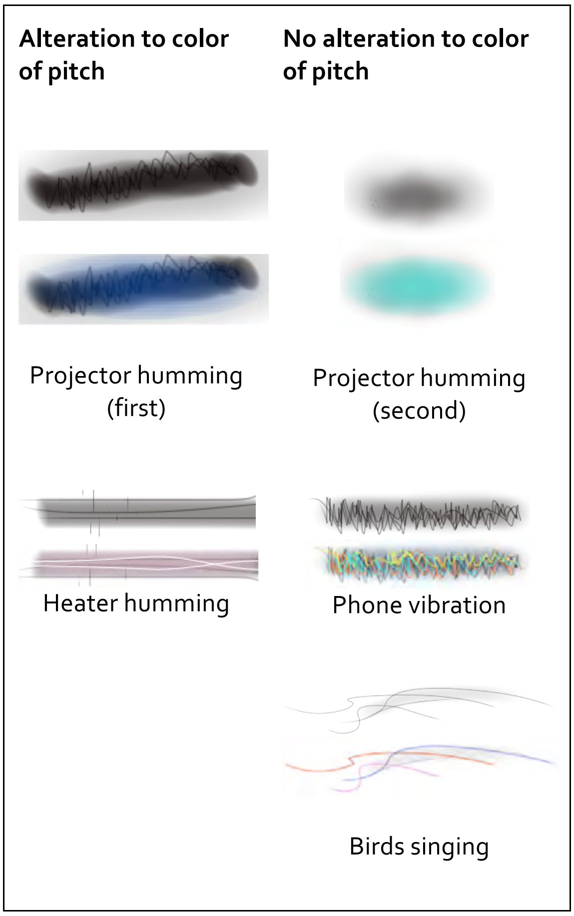


Figure 7. Effect of texture on pitch: textures from RC's journal, and artist impression for illustration purposes

In this section we took a look at the textures of sounds with clear pitch, and their interference with color associations. It is quite clear that a dense texture can cause the darkening of perceived colors, as in the case of projector humming 1. Although phone vibrations were dense as well, they had discernible, more delicate lines within the texture which allowed RC to identify matching colors to their pitches. The heater humming sound seemed to also have a dense enough texture, with fewer 'lines' that changed the pitch-associated color of the sound. The second projector humming (**Figure 4**) and birds singing however, did not receive any change – in the first case, the perceived colors tended to encompass the entire texture. In the latter, like the phone vibration, the individual colors were easily discernible through clear lines within the texture.

For textures that did change the perceived color of the sound, the denser the texture, the darker the color became. The lighter and sharper the texture, the lighter the color of the sound was perceived to be. This is similar to the tendency to perceive lightness with higher pitches, as discussed in **Section 4**.

7. Discussion

In this paper we have described how silence provided RC the opportunity to better understand her senses, and the ways in which she perceives sound. For a moment though, we step back for a closer look at the fundamental concept which allowed for the insight that forms the basis of the paper:

What is silence? To this question there is no straightforward answer, apart from it being a medium for a heightened awareness of sound, and in this case, RC's senses. But we do know that it is most definitely – the absence of sound, and deliberately so, in RC's case.

Returning once more to Hodkinson (2007), it is stated that "perception is influenced by intellectual understanding", and that "the sounds available for an audience to hear at a performance of (John Cage's) *4'33"* will actually vary depending on their prejudices and expectations" (Hodkinson p. 51). The expectation of *hearing sounds* made listening to silence an impossible concept for RC, thus allowing for the vivid perception of such small sounds by RC.

Perhaps silence could take on a different meaning in an anechoic chamber, where even low-level sounds are eliminated. But for now, it is clear that there was no other way that RC could grapple with silence, but by being especially attentive to the few sounds left behind.

References

Cytowic, R. E. (2002). *Synesthesia: A union of the senses*. MIT press.

Cytowic, R. E. (2003). *The man who tasted shapes*. Imprint Academic.

Eagleman, D. M., & Cytowic, R. E. (2009). *Wednesday Is Indigo Blue: Discovering the Brain of Synesthesia*. MIT Press.

Hodkinson, J. (2007). *Presenting Absence: Constitutive Silence in Music and Sound Art Since 1950s* (Doctoral dissertation, University of Copenhagen, Faculty of Humanities, Department of Musicology).

Oliveros, P. (2005). *Deep listening: a composer's sound practice.* IUniverse.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, *42*(2), 264-280.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 61-68.

Same Time, Same Place, Keep it Simple, Repeat: Four Rules for Establishing Causality in Interactive Audio-Visual Performances

Yago de Quay

Department of Radio-Television-Film, University of Texas at Austin, USA

yagodequay@utexas.edu

Abstract

Recent consumer-grade technologies that extract physiological biosignals from users are being introduced into interactive live performances and innovating its practice. However, the relationship between these signals and the responsive audiovisual content is often not understood by the audience. Recent discoveries in neuroscience can address this issue by proposing perceptual cues that help us connect the things we see and hear in our environment. Drawing from the field of neuroscience, and more specifically the theory of crossmodal binding, this paper proposes four rules that govern the mechanism which attributes causality between audiovisual elements: *same time, same place, keep it simple, repeat.* Intended as a set of guidelines for artists, they will help the audience unify, and understand the underlying cause to, what they see and hear in a performance. The last section describes a brainwave-based performance called *Ad Mortuos* that applies these four rules. A video of the performance is available at http://www.tiny.cc/admortuos.

Keywords: crossmodal binding, EEG based brain computer interface, audiovisual performance

1. Introduction

There is an increasing interest in developing interfaces that convert physiological, neurological and motor processes like heartbeat, muscle contraction, brain activity and body motion into digital media. Powered by the growing availability of consumer-grade medical devices, artistic experiments with biosignals are enriching the performative landscape.

However, these new interfaces tend to conceal the cause and consequence relationship between the performer's actions and the audiovisual output (Barbosa, 2006; d'Escriván, 2006; Hook et al., 2012; Jensenius, 2007). Although interactive transparency is up to the discretion of the artist, some argue that there is a need for approaches that make it clear and meaningful to the audience what and how audiovisual materials are being manipulated (Armstrong, 2006; Schloss, 2003; Wechsler, 2006).

Two popular approaches have been the application of Embodied Cognition schemas and Human Computer Interaction principles (Hook et al., 2012; Leman, 2008). Together they provide suitable analogies that link the somatic qualities of a gesture or sound to the affordances of an interface (Larson & Johnson, 2011; Tanaka, 2010). These analogies are useful for user-centered design but they box our creativity within intuitive, generic models (Downie, 2005).

An alternative approach proposed by (Callear, 2012; Cooke, 2010; Coulter, 2010; Whitelaw, 2008) and pursued here, draws from neuroscience a means to leverage the neurological process that governs the unification of audio and visual information into coherent, meaningful chunks. This process called crossmodal binding picks up spatiotemporal cues that indicate to the brain that an object and sound in our environment are related and have a causal relationship. When crossmodal binding is created synthetically with a live performer and interactive digital media, we are tricking our brain into creating these physical-virtual objects that share characteristics with the natural environment responsible for the evolution of crossmodal binding.

This paper argues that unrelated audio and visual elements in interactive live performance can be bound into a single percept by following four rules based on crossmodal binding: *same time*, *same place*, *keep it simple*, *repeat*. Designed with the audience in mind, these four rules promote audiovisual unity and clarify the cause and effect relationships between the performer and interactive audiovisual content. If followed, these rules will free the artists to work with abstract content and obscure interactive technologies and still have them perceptually and cognitively bound into a unified concept in the eyes of the spectator.

Section 2 of this paper investigates literature on crossmodal binding supporting the four rules. **Section 3** prescribes eight simple design strategies that help establish causality and coherence between the live performers, visuals and sound. **Section 4** describes a performance called *Ad Mortuos* that was presented in the MOVE! concert series at UT Austin. The goal of the piece was to create a coherent expressive medium that combines thought, movement, sound and image using the four binding rules.

2. Crossmodal binding

A cat accidentally pushes a vase off a table and it smashes on the floor. We are able to rapidly orient our eyes to the scene, catch a glimpse of the culprit, and deduce what happened partially because we treated what we heard and saw as referring to a single spatiotemporal event (Holmes & Spence, 2005). The process of grouping vision and sound into a single percept, referred to here as crossmodal binding, is a crucial, perceptual task that allows us to recognize objects in the real world that represent the common cause to the things we sense (Whitelaw, 2008). The way we combine sound and image depends on the causal relationship one infers through top-down decisions (the cat pushed the vase) (Körding et al., 2007), as well as automatic bottom-up combinations (the loud smash came from the vase) (Whitelaw, 2008). The ecological goal of crossmodal binding is to find a common cause of what we see and hear to an object or event in the environment.

Sounds and images that share similar qualities such as timing, localization, and semantic compatibility, are more likely to be treated as referring to a single multi-sensory event or object rather than separate unisensory events (Spence, 2007). These qualities are referred to as amodal properties since they are not specific to any particular sensory modality (Welch, 1999). Furthermore, belief in a causal relationship between auditory and visual information, also known as unity assumption, can also serve as a cue for binding (Schutz & Kubovy, 2009). Experiments with simple auditory and visual stimuli suggest that arbitrary sounds and images can be bound together simply through spatiotemporal coincidence (Effenberg, 2007; Körding et al., 2007; Schutz & Kubovy, 2009), and that the recognition of spatiotemporal coincidence is enhanced when the observer believes that the auditory and visual stimuli should go together (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Molholm, Ritter, Javitt, & Foxe, 2004).

2.1 Same time

When an auditory phenomenon and a visual phenomenon are presented at the same time an inevitable and irresistible weld is produced between them (Chion, Gorbman, & Murch, 1994). This weld allows seemingly unrelated pairs of audiovisual material to come together. For example, using temporal disparity as the independent variable, (Lewald & Guski, 2003) conducted three experiments that asked participants to judge the impression of the likelihood of a common cause, spatial alignment, and synchrony, of simple audio and visual stimulus. In all three experiments temporal disparity had a significant negative effect on participants' judgment. Optimal results were found when the audio stimulus was presented 50-100 ms after the visual stimulus.

Consecutive stimuli tend to "pull" the perceived time of occurrence towards each other in an effect called "temporal ventrilo-quism" (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In their studies, Morein-Zamir, Soto-Faraco and Kingstone found that two consecutive visual events were perceived as happening in a shorter time frame when sound markers intervene between them compared to when the sounds were not present. A longer time frame was experienced when the sound markers happen before and after both visual events. The authors claim that the audio markers were attracting the perceived time of occurrence of each visuals event towards them.

Stimuli with slow attacks make asynchronies harder to detect (Jaskowski, 1993). (Vatakis & Spence, 2006) found that subjects watching a reverse video clip of a guitarist playing classical music had difficulty in identifying the starting order of the asynchronous string sound and string pluck. The authors claim that the reverse video made it hard to detect when the string were being plucked. Asynchronies in the normal and reverse video clip were detected at around 30 ms and 150 ms, respectively, using just noticeable difference (JND) and a test for the percieved point of subjective simultaneity (PSS).

2.2 Same place

If sound and image are presented close to each other in space, observers tend to perceive a single underlying cause, if they are far apart from each other observers tend to infer two independent causes (Körding et al., 2007). Using as an example a ventriloquist and a puppet, if the voice and puppet appear together in the same place we ascribe one source to the auditory and visual stimuli. Increasing the distance between the voice and puppet makes us infer that each originates separately from an auditory source and visual source.

Spatial distance between the interaction of two visual object is negatively correlated with causality (Yela, 1952). Using the classic Michotte launching display experiment, Yela's study demonstrates that increasing the spatial gap between the stopping point of the first moving object and the starting point of the second moving object decreases the chances of causal impression, that is, we are less likely to believe that the first object caused the second one to move.

Spatially moving events with the same trajectory may produce stronger audiovisual binding than static events (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002) Furthermore, visual motion tends to influence the auditory motion but not vice versa (Körding et al., 2007). This sensory dominance asymmetry can be explained by modality precision hypothesis which predicts that the sensory channel that provides the most reliable information will dominate crossmodal integration, which in the case of localization is vision (Soto-Faraco, Spence, & Kingstone, 2004).

2.3 Keep it simple

"The austerity of the experimental situation may in-and-of-itself lead the participants into an 'assumption of unity' concerning the likely relation between the two signals" (Spence, 2007, p. 4). The presence of events that have been associated with each other strengthen audiovisual binding (Jack & Thurlow, 1973) while the presence of distractor events that do not share amodal properties weaken it (Welch, 1999). Even consciously, distractors are hard to ignore because the brain automatically detects discordance between stimuli (Spence, Ranson, & Driver, 2000) .

Complex unimodal (either visual or auditory) stimuli can reduce crossmodal binding in favor of grouping elements within one sensor modality, a phenomena referred to as intramodal binding (Spence, 2007). Experiments by (Sanabria, Soto-Faraco, Chan, & Spence, 2005) with intramodal binding investigate whether the number of equally moving visual distractors has an influence on the perceived movement of an auditory stimulus (left-to-right or vice versa). Results are counterintuitive. They report that increasing the number of visual distractors from two to six made it easier for participants to judge the trajectory of the auditory stimulus. Since previous research by (Morein-Zamir et al., 2003; Schutz & Lipscomb, 2007; Slutsky & Recanzone, 2001; Watanabe & Shimojo, 2001) found contamination of amodal properties between two sensory input as evidence for crossmodal binding, a reduction of contamination with six distractors suggests a reduced magnitude of crossmodal binding. Taken together, these studies show that simple audiovisual conditions enhance crossmodal grouping while complex audiovisual conditions segregate both senses and can encourage intramodal grouping.

Bimodal stimuli are registered faster and more accurately than unimodal stimuli (Teder-Sälejärvi, Russo, McDonald, & Hillyard, 2005). In their study, Teder-Sälejärvi et. al. presented subjects with randomized sequences of unimodal and bimodal light/sound stimuli and tested them for speeded responses to infrequency targets of higher intensity. When presented with only light or sound, subjects in Teder-Sälejärvi's study took longer and made less correct detections than when both light and sound were present.

2.4 Repeat

Repeated, simultaneous exposure to artificial audio and visual pairs increases their association as a congruent crossmodal object (van der Linden, van Turennout, & Fernandez, 2011) and produces integration-related cortical plasticity in the frontal lobe (Naumer et al., 2009).

Instructional and/or situational context favoring the assumption that two audio and visual stimuli should go together (unity assumption) can lead to changes in our perceptual mechanism that facilitate crossmodal binding (Welch, 1999). (Vroomen, Keetels, de Gelder, & Bertelson, 2004) noticed that when presented repeatedly with slightly asynchronous tone burst and flashing light (~+/- 200ms difference), observers tended to recalibrate their tolerance window towards the direction of the lag (up to 32 ms; ~10% of the lag). Aside from demonstrating our brains' dynamic approach to spatiotemporal congruency, these studies suggest that the unity assumption gives strong impetus to the emergence of a crossmodal representation of an object.

Unimodal familiarization of an audio or visual stimulus increases identification speed and accuracy within and across modalities in congruent audiovisual pairs (Ballesteros, Gonzalez, Mayas, Garcia-Rodriguez, & Reales, 2009). In Ballestero's study a group of participants was exposed to 30 ecological sounds or pictures and then told to identify a similar set of sounds or pictures in a speeded object naming implicit task. Results show that familiarization reduces identification time (1860 ms) and errors (2.51%) compared to no-study event (2339 ms and 3.26%). Furthermore, unisensory familiarization decreases response time in within-modal as well as crossmodal conditions, supporting the argument that crossmodal priming is not modality specific.

3. Design guidelines

This section presents a preliminary effort to explain how the principles *same time, same place, keep it simple, repeat* can be applied in the composition and design of an interactive audiovisual piece. They increase feature congruency (same time, same place), reduce anti-binding distractors (keep it simple), and suggest relatedness (repeat). These guidelines provide an approach to enhancing the binding of the various creative elements such as music, sound, scenery, performers, light and video. The goals are twofold: enable artists to bind abstract material, and help the audience understand the cause-and-effect between interactive elements in cases where the technology hides the mechanism of causality. The limitations are also twofold: artistic consideration supersede the need to establish causality, and the four principles govern only those elements that are implicated in a cause-and-effect relationship. Each rule breaks down into two guidelines that strengthen crossmodal binding.

Same time: Synchronous events and asynchronous events that have slow attacks;

Same place: Superposition of elements in the physical space and correlated spatial displacement;

Keep it simple: Absence of elements or events that are not related and audiovisual effects as opposed to uniquely auditory or visual;

Repeat: Repeated exposure to each cause-and-effect relationship and explicit information about the interactive relationships.

4. Application of the rules in a performance

Ad Mortuos is a collaborative work that features live drawing, voice, brainwaves and dance. The theme was inspired by a poem that tells the story of the immortal soldier Athanatoi (Greek –"without death") who sings on "bright love," weaving a voice of hope through the first tableau. Then Athanatoi falls into deep séance, conjuring thoughts of the afterlife and immortality. A video of the performance can be watched at the webpage http://www.tiny.cc/admortuos. Athanatoi is played by performer and vocalist Yago de Quay using a brain-computer interface (BCI). He wears a headset device (*Emotiv EPOC*, detecting EEG) that enables deliberate mental tasks and facial expressions to manipulate visual and musical parameters. The device connects wirelessly to the control software in a Windows 8.1 tablet that he wears on the chest, attached to a hoodie. A gyroscope sensor on the performer's upper chest, attached to an under shirt, reports body orientation to the tablet as well. The control software is trained to recognize Quay's facial expressions and mental commands. Together with the orientation data, these are sent by a datagram protocol (UDP) through a wireless router above the stage to the sound and visuals computers. Sound is produced using a Max/MSP patch (*Cycling '74*) that layers voice and sound effects over a base track and then plays them through a stereo system. The visuals are manipulated in a VDMX software (*Vidvox*) and projected on the floor from overhead.

Table 1 illustrates the mapping between the BCI performer and the digital audiovisual output. The interactive elements consist of the BCI performer, six dancers, projections and sound effects. The BCI performer serves as the input while the six dancers, projections and sound effects as outputs. The three input types—body orientation, facial expressions and mental commands—demonstrate different degrees of perceived cause-and-effect, from obvious to invisible, respectively. Since mental commands are imperceptible to the audience, the performer uses his right hand as a visual cue to help connect the gesture to the size of the circles.

Table 1. Input and output mapping

Input	**Digital Output**
Body orientation	Slant of concentric circles Play position looped vocals
Facial Expressions	Distortion of circles Bleeps and noises
Mental Commands	Size of circles Right hand height

To help the audience understand the cause-and-effect mechanism in the BCI section of *Ad Mortuos*, the beginning strictly adhered to the four rules *same time, same place, keep it simple, repeat* before allowing independent, non-interactive events to occur such as solos by each one of the dancers. **Photos 1-4** on the next page illustrate how these principles impacted the overall visual aesthetic of the piece. The next paragraphs describe how the rules were implemented.

Same Time: Changes initiated by the BCI performer are immediately reflected on all interactive audiovisual elements. There are no long-term, delayed interactions. The performer's facial expressions and body orientation are imitated by the dancers and produce an audiovisual effect. Mental commands on the other hand, take time to formulate. In order to keep them synchronous with the outputs we settled for slow, continuos audiovisual changes.

From top to bottom. Photo 1. *Same Time.* Body orientation rotating the visuals. Photo by Rodrigo Guedes. Photo 2. *Same Place.* Superimposed elements moving in the same direction. Photo by Rodrigo Guedes. Photo 3. *Keep it Simple.* Absence of non-interactive elements. Photo by Chian-ann Lu. Photo 4. *Repeat.* Repeated facial expressions deform the visuals. Photo by João Beira.

Same Place: All the interactive, visible elements are co-located. Body orientation, facial expressions, and hand gestures produce changes on the adjacent dancers and projections. Displacement of the different body parts and slant of the projections follow the same trajectory.

Keep it Simple: Aside from the lights and base track, which hardly change, all elements on stage are interactive. The three input types are never performed simultaneously and always produce a predictable effect on all the outputs (dancers, projection and visuals).

Repeat: It was at the discretion of the BCI performer when and how to perform the inputs. Nevertheless, they were repeated sequentially multiple times to help establish causality. The concert program informed the audience, without getting into much detail, that the BCI performer was controlling the projections and sound effects.

5. Conclusions

Introducing new interactive technologies in performance can foster innovative artistic practice. However, when technological mediation does not exhibit the physical causality properties present in the natural world that we are used to experiencing it runs the risk of obscuring the relationship between the performer's input and the media output. Furthermore, performances based on physiological biosignals leave the audience clueless to the nature of the input. This paper looks at empirical studies in crossmodal binding that are concerned with how our brains combine sound and image to suggest four rules *same time, same place, keep it simple, repeat* that can help re-establish the causal link between the performer, image and sound. The magnitude of synchronicity and proximity (same time, same place) of audio and visual events has a positive effect on their grouping. Audiovisual distractors and incongruity has a negative effect in grouping (keep it simple). Explanation about, and exposure to, audiovisual pairs enhances the observers' assumption that they should go together (repeat). Applicable to meditated performances, these four rules give hints to the audience's brain that the performer, images and sounds are related and referring to an underlying cause. This paper ends by suggesting eight design guidelines and applies them to an interactive audiovisual piece entitled *Ad Mortuos* that uses brainwaves to control music and visuals.

Credits for *Ad Mortuos*
Poet: Stephanie Pope
Spoken word: LaQuerra Carpenter
Choreographer: Yacov Sharir
Composer/Sound Design: Bruce Pennycook
Visuals: João Beira, Rodrigo Carvalho
Costume Design: Kelsey Vidic
BCI performer/Vocalist: Yago de Quay
Dancers: Emily Snouffer, Becca Bagley, Gianina Casale, Summer Fiaschetti, Katie McCarn and Allyson Morales
Camera: Pedro Miguel Resende, Filipa Rodrigues

References

Armstrong, N. (2006). An Enactive Approach to Digital Musical Instrument Design. (Ph.D.), Princeton University.

Ballesteros, S., Gonzalez, M., Mayas, J., Garcia-Rodriguez, B., & Reales, J. M. (2009). Cross-modal repetition priming in young and old adults. *European Journal of Cognitive Psychology, 21*(2-3), 366-387.

Barbosa, Á. M. (2006). *Displaced Soundscapes: computer supported cooperative work for music applications.* (Doctor per la Universitat Pompeu Fabra with Mention of European Doctor), Pompeu Fabra University, Barcelona. Retrieved from http://www.abarbosa.org/docs/t_abarbosa.pdf

Callear, S. (2012). Audiovisual particles: parameter mapping as a framework for audiovisual composition. (Ph.D.), Bath Spa University.

Chion, M., Gorbman, C., & Murch, W. (1994). *Audio-vision: sound on screen*. New York: Columbia University Press.

Cooke, G. (2010). Start making sense: Live audio-visual media performance. *International Journal of Performance Arts and Digital Media, 6*(2), 193-208.

Coulter, J. (2010). Electroacoustic Music with Moving Images: the art of media pairing. *Organised Sound, 15*(1), 26-34.

d'Escriván, J. (2006). To sing the body electric: Instruments and effort in the performance of electronic music. *Contemporary Music Review, 25*(1-2), 183-191.

Downie, M. (2005). *(Context) Choreographing the Extended Agent: performance graphics for dance theater.* (PhD), MIT, Boston. Retrieved from http://openendedgroup.com/index.php/publications/thesis-downie/

Effenberg, A. O. (2007). *Movement Sonification: Motion perception, behavioral effects and functional data.* Paper presented at the 2nd International Workshop on Interactive Sonification, York, UK.

Holmes, N. P., & Spence, C. (2005). Multisensory Integration: Space, Time and Superadditivity. *Current Biology, 15*(18), R762-R764.

Hook, J., Schofield, G., Taylor, R., Bartindale, T., McCarthy, J., & Wright, P. (2012). *Exploring HCI's relationship with liveness*. Paper presented at the CHI '12 Extended Abstracts on Human Factors in Computing Systems, Austin, Texas, USA.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept Mot Skills, 37*(3), 967-979.

Jaskowski, P. (1993). Temporal-order judgment and reaction time to stimuli of different rise times. *Perception, 22*(8), 963-970.

Jensenius, A. R. (2007). Action-sound: developing methods and tools to study music-related body movement. (Ph.D.), University of Oslo.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE, 2*(9), e943.

Larson, S., & Johnson, M. (2011). Something in the Way She Moves: The Metaphor of Musical Motion *Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington, IN: Indiana University Press.

Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158*(4), 405-414.

Leman, M. (2008). Embodied Music Cognition and Mediation Technology: The MIT Press.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468-478.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual–Auditory Object Recognition in Humans: a High-density Electrical Mapping Study. *Cerebral Cortex, 14*(4), 452-465.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res, 17*(1), 154-163.

Naumer, M. J., Doehrmann, O., Muller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audio-visual object representations. *Cereb Cortex, 19*(7), 1641-1653.

Sanabria, D., Soto-Faraco, S., Chan, J., & Spence, C. (2005). Intramodal perceptual grouping modulates multisensory integration: evidence from the crossmodal dynamic capture task. *Neuroscience Letters, 377*(1), 59-64.

Schloss, W. A. (2003). Using Contemporary Technology in Live Performance: The Dilemma of the Performer. *Journal of New Music Research, 32*(3), 239-242.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Experimental Psychology: Human Perception and Performance, 35*(6), 1791-1810.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception, 36*(6), 888-897.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7-10.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res Cogn Brain Res, 14*(1), 139-146.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-Modal Dynamic Capture: Congruency Effects in the Perception of Motion Across Sensory Modalities. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 330-345.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology, 28*(2), 61-70.

Spence, C., Ranson, J., & Driver, J. (2000). Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Perception & Psychophysics, 62*(2), 410-424.

Tanaka, A. (2010). *Mapping Out Instruments, Affordances, and Mobiles.* Paper presented at the New Interfaces for Musical Expression, Sydney, Australia.

Teder-Sälejärvi, W. A., Russo, F. D., McDonald, J. J., & Hillyard, S. A. (2005). Effects of Spatial Congruity on Audio-Visual Multimodal Integration. *Journal of Cognitive Neuroscience, 17*(9), 1396-1409.

van der Linden, M., van Turennout, M., & Fernandez, G. (2011). Category Training Induces Cross-modal Object Representations in the Adult Human Brain. *Journal of Cognitive Neuroscience, 23*(6), 1315-1331.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research, 1111*(1), 134-142.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32-35.

Watanabe, K., & Shimojo, S. (2001). When Sound Affects Vision: Effects of Auditory Grouping on Visual Motion Perception. *Psychological Science, 12*(2), 109-116.

Wechsler, R. (2006). Artistic Considerations in the Use of Motion Tracking with Live Performers: a Practical Guide. In S. Broadhurst & J. Machon (Eds.), *Performance and Technology: Practices of Virtual Embodiment and Interactivity*: Palgrove Macmillian.

Welch, R. B. (1999). Meaning, Attention, and the "Unity Assumption" in the Intersensory Bias of Spatial and Temporal Perceptions. In G. Aschersleben, T. Bakhman, & J. Müsseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events*. Amsterdam: Elsevier.

Whitelaw, M. (2008). Synesthesia and Cross-Modality in Contemporary Audiovisuals. *Senses & Society, 3*(3).

Yela, M. (1952). Phenomenal causation at a distance. *Quarterly Journal of Experimental Psychology, 4*(4), 139-154.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 69-78.

Interactive Computation of Timbre Spaces for Sound Synthesis Control

Stefano Fasciani

Digital Signal Processing Lab, School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore

stefanofasciani@stefanofasciani.com

Abstract

Expressive sonic interaction with sound synthesizers requires the control of a continuous and high dimensional space. Further, the relationship between synthesis variables and timbre of the generated sound is typically complex or unknown to users. In previous works, we presented an unsupervised mapping method based on machine listening and machine learning techniques, which addresses these challenges by providing a low-dimensional and perceptually related timbre control space. The mapping maximizes the breadth of the explorable sonic space covered by the sound synthesizer, and minimizes possible timbre losses due to the low-dimensional control. The mapping is generated automatically by a system requiring little input from users. In this paper we present an improved method and an optimized implementation that drastically reduce the time for timbre analysis and mapping computation. Here we introduce the use of the extreme learning machines for the regression from control to timbre spaces, and an interactive approach for the analysis of the synthesizer sonic response, performed as users explore the parameters of the instrument. This work is implemented in a generic and open-source tool that enables the computation of ad hoc synthesis mappings through timbre spaces, facilitating and speeding up the workflow to get a customized sonic control system.

Keywords: interactive timbre space, generative mapping, perceptual sound synthesis.

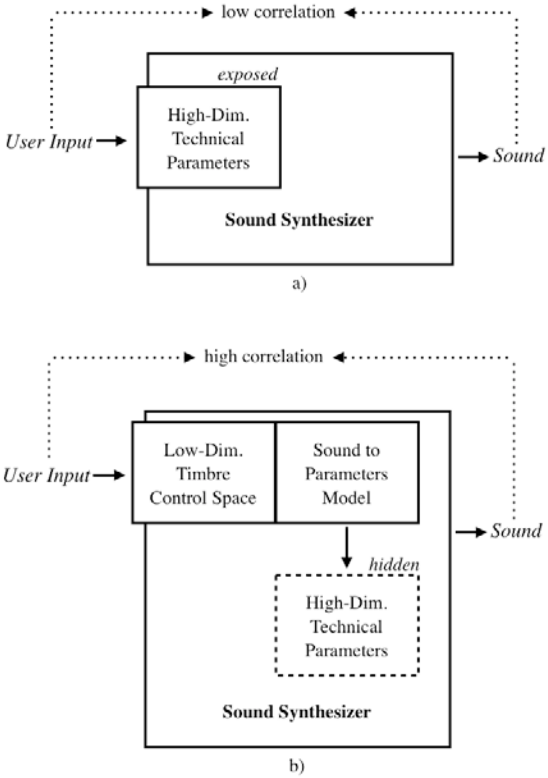
1. Introduction

The sonic potential of synthesisers has been constantly progressing in the past decades, supported by novel techniques and rapid advances in related technologies. These devices provide almost boundless resources for the creative task of composers, designers, and musicians. Users directly interact with variables of the synthesis algorithm, as in the early days of sound synthesiser, when these were mostly operated by engineers and scientists. Generating a sound with intended timbre characteristics requires sufficient control intimacy (Fels, 2000; Moore, 1988) that implies knowledge of the relationship between parameters and sonic output, which may be complex. Most synthesis methods present a weak or missing relationship to sound perception models (Wishart, 1996). This determines a sound interaction that is object oriented rather than model oriented. Existing control strategies are typically process-based rather than result-based, which is unusual in Human Computer Interaction (HCI). This can drastically limit the efficiency of users' creative workflow. The time spent in implementing a sonic intent may exceed the time spent in conceptualizing the ideas, contradicting the creative nature of the task. Mapping strategies have contributed in easing the complexity and boosting the expressivity of the interaction (Miranda & Wanderley, 2006). The use of machine learning techniques for mapping the user input to instrument parameters has been proposed too. However, the control abstraction has not changed and the full exploitation of the synthesis potential is still challenging for users.

To address this problem, control strategies that map the user input onto synthesis variables through timbre spaces or perceptually related layers, reviewed in Section 2, have recently proliferated. Along this line, in our previous works (Fasciani, 2014; Fasciani & Wyse, 2012, 2015) we introduced a control method, implemented in an open-source software, which concurrently address the high dimensionality of synthesis control spaces and the lack of relationship between variation of control parameters and the timbre of the generated sound. In addition, our work introduced an unsupervised and automated generative mapping, independent of the synthesis algorithm and implementation, which does not require users to provide training data. Machine listening and machine learning techniques are employed to compute mappings that, given a set of synthesis variable parameters, maximize the breadth of the explorable perceptual sonic space covered by the synthesizer, and minimize the possible timbre losses due to the lower dimensionality of the control. The resulting interactive timbre space can drive simultaneously any number of parameters, and it is adapted to generic controllers, with components are directly mapped onto the most varying timbral characteristics of the sound.

A usability limitation of our system is the time required for the mapping generation. This includes the time for executing the automated parameters-to-sound analysis of the specific synthesizer (it generates the training data), and the time to compute the mapping with machine learning algorithms. The time increases exponentially with the number of synthesis parameters users aim to control through the timbre space. Here we address this issue introducing improvements in method and implementation. The adoption of the Extreme Learning Machines (ELM) algorithm (Huang, Zhu, & Siew, 2006) to train the Artificial Neural Network (ANN) performing the regression between control and timbre space, significantly reduces computation time and improves accuracy. Further reduction is achieved performing the parameters-to-sound analysis interactively as users operate the synthesizer.

The system described in this paper enables users to compute ad hoc timbre-oriented interaction, reducing the dimensionality of synthesis control space without losing timbre expressivity, which is central in modern music, despite its high dimensionality and blurry scientific definition (Risset & Wessel, 1999). The sense of hearing provides a secondary feedback on performer-instrument interaction, but for digital musical instrument this relationship can be loose. This work provides a tighter relationship between control and sound perception, independently of the specific control modality. It contributes to a reduction in the overhead for the realization of sonic intent arising by the complexity of sound generation algorithms. Further, it supports users' customization of instrument mapping while minimizing setup time and effort. **Figure 1** illustrates the traditional synthesis control method against the proposed one, which hides the technical parameters and provides higher correlation between sound and user input.



a)

b)

Figure 1: Diagrams of a) traditional and b) proposed synthesis control method. The introduction of the timbre space hides the technical parameter and reduces the dimensionality of the control.

The rest of the paper is organized as follows. In Section 2 we present a survey on methods for perceptually related sound synthesis control. Section 3 includes a brief summary of our generative mapping strategy and sonic interaction. In Section 4 we introduce the interactive and optimized the mapping computation. Section 5 describes system integration and implementation. Conclusion and future work are discussed in Section 6.

2. Perceptually related synthesis control strategies

A synthesis control strategy is perceptually related when it explicitly manipulates timbral attributes of the generated sound. Timbre is considered "the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled pitch or loudness" (McAdams & Bergman, 1979), or the attribute that allows to distinguish two sounds with equal pitch and loudness. Timbre as a whole can be measured only at nominal level. It can be represented with a high dimensional space, whose individual components are acoustic ordinal descriptors (e.g. brightness, noisiness, color, attack, decay, vibrato, tremolo). For an accurate representation, the computation of the descriptors must include psychoacoustic principles and nonlinearities of the human auditory model. In specific applications, the timbre space can be reduced to few dimensions with multidimensional scaling techniques (Grey, 1977). However, it is necessary to preserve the element-by-element proximity between original and low-dimensional space by using nonlinear reduction techniques and non-Euclidean distances (McAdams & Cunible, 1992).

The control parameters of some synthesis techniques have a tighter relation with the output timbre due to the intrinsic characteristic of the algorithm generating sound. Additive synthesis provides control of amplitude and frequency of individual spectral components, which explicitly characterize the timbre of both harmonic and inharmonic sounds. When sound synthesis is based on a database of heterogeneous samples, it is possible to analyze the timbre characteristics of the sources and use this information in the control layer. Concatenative synthesis (Schwarz, Beller, Verbrugghe, & Britton, 2006), an extension of granular synthesis, explicitly provides control over specific sound descriptors. The synthesis sequences grains from a corpus of segmented and analyzed samples, according to the coordinate of a low dimensional interactive descriptor space, which represent an interactive timbre control structure. Through experience, humans can identify mechanical aspects (material, shape, cause, location) of the event that generates sound by auditory cues (Emmerson, 1998). In the same way we can predict the sound of a mechanical event, including performers playing musical instruments. Physical modelling synthesis simulates the mechanical phenomenon governing acoustic sound generation. The correlation between control and timbre is still complex and nonlinear, but known to users. Therefore, explicit timbre manipulation is almost immediate.

Timbre-oriented control strategies has been proposed from the pioneer work of Wessel (1979), in which subjective timbre dissimilarities between orchestral instruments synthesis patches were measured, reduced to a 2D interactive space, and then used to control the additive synthesizer. The parameters were generated by interpolating between those of the original patches used for the listening tests. This allowed generating a wider spectrum of timbres than that presented in the listening tests. Moreover this provided a drastic reduction of the control space and perceptually meaningful control dimensions. Subjective listening was replaced with computerized perceptual analysis of sounds by Jehan & Schoner (2001) in a synthesis engine that models and predicts the timbre of acoustic instrument. The analysis included in pitch, loudness, and timbre's descriptors such as brightness, noisiness and energy of the bark critical bands. Cluster-weighted modeling approximates synthesis parameters from timbre descriptions, predicts a timbre given a new set of parameters. Arfib, Couturier, Kessous, & Verfaille (2002) generalized the different approaches for a timbre-oriented control introducing perceptual layer in the mapping from gesture to musical instrument. Including the intermediate perceptual space in the modular mapping of musical interfaces improves sensitivity and efficiency of the interaction. Seago (2013) proposes a method for searches in timbre spaces based on weighted centroid localization, claiming the quasi-linearity of these spaces and explicit association of individual dimensions with acoustical attributes. Therefore, searches should return a single and optimal solution. This technique addresses usability problems of sound synthesis and it affords engagement with the sound itself, rather than with irrelevant timbre visual representations.

When the parameters-to-sound relationship of a specific synthesis algorithm is unknown, it is still possible to implement a perceptually related control by measuring psychoacoustic sound descriptors and generating a model of the synthesizer response, required for the mapping. This approach presents challenges in computing an accurate model and in retrieving coherent parameters, but it guarantees independency between control strategy and synthesis method. In this work we adopt and extend a similar strategy, which is also found in other systems. A timbre space composed of the principal components of the spectral output of an FM synthesizer is used to generate interactively auditory icons and 'earcons' (Nicol, Brewster, & Gray, 2004). A generic and modular framework, including feature evaluators, parametric synthesizers, distance metrics, and parameter optimizers controls the synthesis with a set of well-defined and arbitrary acoustic features (Hoffman & Cook, 2006), extending the work of Puckette (2004). This approach minimizes the distance between target and synthesized sonic characteristic, which are user-defined or computed from a reference sound. Klügel, Becker, & Groh (2014) propose a collaborative and perceptually related sound synthesizer controlled by multi-touch tabletop, where the interactive timbre space is mapped and visually represented. Generative topographic techniques are used to map perceptual audio features to synthesis parameters. Computational challenges in performing a deep time-frequency analysis with high dimensional parameters space are identified but not addressed. An approach based on fuzzy models, manually defined by experts, expose to users a large set of intuitive and natural timbre descriptors, which allows novice users to effectively control the synthesis algorithm by visual programming (Pošćić & Kreković, 2013).

Strategies for perceptually related control of audio mosaicing systems has been introduced as well. In this case the sound synthesis is replaced with the sequential playback of samples retrieved from a database organized according to audio descriptors. This generates evolving timbre and textures. Proposed implementations (Lazier & Cook, 2003; Schnell, Cifuentes, & Lambert, 2010; Grill, 2012) map the analyzed samples collection onto a two- or three-dimensional sonic map defined after long- and short-term sound analyses.

3. Timbre space for synthesis control

This section briefly summarizes our approach to perceptually related synthesis control through timbre spaces. We assume having no prior knowledge of the synthesis algorithm and of the relationship between parameters and output sound, neither this provided by users. The parameters-to-sound model is derived from input-output observations. In particular we use machine listening techniques to 'hear' the timbre almost like humans. Further we assume univocal relation between parameters and sound, excluding random variables in the generation algorithm, while we tolerate not univocal sound to parameters relationship. Therefore, this control strategy is compatible with any deterministic sound synthesizer.

Performing a full parameters-to-sound characterization of a synthesizer is a demanding task. The set of possible parameter permutations we have to analyze grows exponentially with the number of parameters and their possible values, which is equal to at least 127 for the real-valued ones (e.g. 33G permutations with 5 real-valued parameters). The full characterization requires excessive computational resources and time. In this context this is generally not required. Firstly, when performing, users interact only with few synthesis parameters, the remaining are fixed given a specific synthesis patch. Secondly, the control resolution of real-valued parameters is high to avoid sound glitches in the real-time synthesis control, though for the analysis, a lower resolution can still provide accurate modeling while drastically reducing the number of permutations to measure. Further, the sound-to-parameters analysis aims to identify only the timbre components that change varying the non-fixed synthesis parameters. We do not intend to carry out a comprehensive timbre characterization. Therefore, users can contribute in reducing analysis and mapping complexities also by discarding descriptors that are steady or not relevant, and by limiting the analysis only to a specific sub-region of the ADSR envelope. Further, the method presented here is independent of the specific descriptors selected for the analysis, detailed in Section 5.

Given the set of variable parameters, with their analysis range and resolution, we compute the matrix $\mathbf{I}$ which includes all permutation vectors $\mathbf{i}_j$ of variable parameters. One at a time and automatically, the synthesizer is driven with the $\mathbf{i}_j$, and the sound is analyzed computing vectors of perceptually related descriptors $\mathbf{d}_j$, accumulated in $\mathbf{D}$. The matrices $\mathbf{I}$ and $\mathbf{D}$ model the parameters-to-sound relationship of a specific synthesizer, according to user-defined settings. The matrices have different dimensionality but identical number of elements ($\mathbf{i}_j$ and $\mathbf{d}_j$, associated pairwise).

Our control strategy maps the user control space $\mathbf{C}$ onto the parameter space $\mathbf{I}$ through the timbre space $\mathbf{D}$. We assume that $\mathbf{C}$ has up to three components, which are statistically independent and uniformly distributed in a fixed range. Most general-purpose musical controllers are compliant with this model. For instance the control data generated interacting with a touchpad can be represented with a uniformly distributed square. The proposed generative mapping strategy address issues such as high dimensionality of the timbre space $\mathbf{D}$, the arbitrary shape and distribution of $\mathbf{D}$, and possible not one-to-one relationship between vectors $\mathbf{d}_j$ and $\mathbf{i}_j$ (similar timbres generated with different synthesis parameters). To compute the mapping, the dimensionality of $\mathbf{D}$ is reduced to the one of $\mathbf{C}$ using ISOMAP (Tenenbaum, Silva, & Langford, 2000). Then, the entries in the low-dimensional timbre space $\mathbf{D}^*$ are rearranged into a uniformly distributed space $\mathbf{D}_U^*$ (line, square or cube) using a homotopic iterative transformation that preserve the neighbourhood relationships (Nguyen, Burkardt, Gunzburger, Ju, & Saka, 2009). The obtained space $\mathbf{D}_U^*$ shares identical geometrical and statistical characteristics with the space $\mathbf{C}$. We model the inverse of this transformation with the function *m( )*, which is implemented with a feed-forward ANN performing the regression from the low-dimensional and uniformly distributed timbre space $\mathbf{D}_U^*$ to the low-dimensional timbre space $\mathbf{D}^*$. The ANN is trained with a back propagation iterative algorithm. Reduction and redistribution do not alter the number of entries in the timbre space. Pairwise association between $\mathbf{d}^*_j$ or $\mathbf{d}_U{}^*_j$ and the relative $\mathbf{i}_j$ are still valid. The mapping from the user control vector $\mathbf{c}$ to the synthesis parameter vector $\mathbf{i}$ is therefore defined in the equation below.

$$\mathbf{i} = \mathbf{i}_j : \underset{j}{\operatorname{argmin}} \left| \mathbf{d}^*_j - m(\mathbf{c}) \right|$$

The mapping returns the parameter vector $\mathbf{i}$ equal to the $\mathbf{i}_j$ of $\mathbf{I}$ associated with the timbre vector $\mathbf{d}^*_j$ in $\mathbf{D}^*$ closer to the projection of the user control vector $\mathbf{c}$ onto the reduced timbre space $m(\mathbf{c})$. To cope with the limited parameter resolution of the analysis stage, the real-time mapping includes temporal and spatial interpolations (inverse distance weighting) of the synthesis parameters, as detailed in (Fasciani, 2014). For addressing the possible parameter discontinuities in the $\mathbf{i}_j$ stream due to the not one-to-one relationship between sound and parameters, we also presented a search strategy for the closer $\mathbf{d}^*_j$ restricted to the subset of $\mathbf{D}^*$ that guarantee continuity and space explorability.

In **Figure 2** we illustrate a basic example of our control strategy, which provides a perceptually related and dimensionality reduced synthesis control space. The figure shows a two-dimensional parameter space. Each parameter can assume 8 different values, giving a total 64 permutations, represented with the black circles. The blue, green, and red lines illustrate different divergent mappings that reduce the control space from two to one dimension. The linear mapping, in blue, can retrieve only few entries in the parameter space, losing most permutations and likely timbres. The nonlinear mapping, in green, arranges the 64 possible parameter permutations on a single line, but it does not guarantee any timbre continuity or meaningful organization of the generated sound. Finally the red dotted line represents our mapping strategy, where the 64 possible synthesizer states are sorted on a single dimension according to their previously analysed timbre. The arrangement can follow the most varying perceptually related descriptor within the parameter space, or a user-defined descriptor. This mapping appears random in the parameter space but it is systematic in the timbre space. This control strategy supports any dimensionality of the parameter space **I**, timbre space **D**, and user control space **C**.

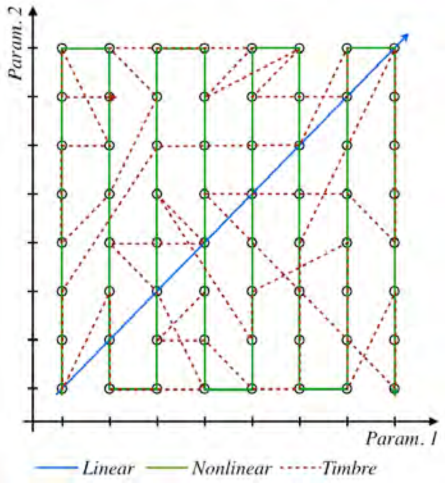


Figure 2: Illustration of synthesis control space two-to-one reduction, using linear mapping (blue), nonlinear mapping (green), and timbre space mapping (red dotted). For the latter, the synthesis states are sorted on a single control dimension according to timbre characteristics.

4. Interactive and optimized timbre space computation

Implementing the method described above, the time required for performing the parameters-to-sound analysis and computing the mapping was often exceeding half an hour. The time strongly depends on the number of entries in **I** and on the dimensionality of **C**. We address this issue introducing an interactive analysis and mapping strategy improvements.

To speed up the analysis, the execution can be run in real-time while users interactively explore the parameters space of the synthesizer. Parameters and descriptor vectors are progressively accumulated in **I** and **D**. The analysis can be executed on the current parameter configuration generating a single pair **i**–**d**, or it can run continuously producing a stream of **i** and **d**. After setting the synthesis parameters to **i** we typically wait 50-100 ms before starting to analyze the sound. This allows sufficient time to the synthesizer for responding to the new configuration. This is not possible when vectors **i** are presented as a stream and the analysis runs continuously. To compute a reliable parameters-to-sound model it is therefore essential to measure and consider the response time of the synthesizer. The two streams of vectors have to be re-aligned associating each **i** with the corresponding **d**. In our implementation, presented in Section 5, Virtual Studio Technology (VST) synthesizers are hosted in Max/MSP. In this context we found that the **i** to **d** misalignment T_{resp} (measured in number of vectors) depends on analysis window size, analysis window step, and Max/MSP signal vector size, as in the equation below.

$$T_{resp} = \mathrm{round}\Big((win / step) + (vect_{MaxMSP} / step)\Big)$$

When performing the analysis we usually compute multiple **d** for each **i**, which are post-processed and merged into a single descriptor vector. This reduces measurement noise taking the average of more observations, and allows including longer-term features of instantaneous descriptors, such as their variation range and periodicity. In the continuous analysis this can be performed in a later stage, by grouping descriptor vectors **d** related to the same parameter vector **i**. However, it is unlikely to find identical vectors **i** when users interact with high-resolution controllers. This is detrimental for the system because **I** and **D** would determine poor modeling (a single observation **d**) and high computational load (large number of **d-i** pairs). Hence, before computing the mapping we reduce the resolution of the parameters in **I** to user-defined values. Then we merge **i-d** pairs with identical synthesis parameters, computing average and range of the descriptors. The periodicity cannot be estimated because the observations may not be consecutive. We also provide an alternative method to reduce the automatic analysis time, running VST and Max/MSP in non-real-time mode so that audio samples are generated and analyzed at a rate higher than the audio sampling period, fully using the CPU power.

In the mapping computation the most computationally demanding part was the back propagation iterative training algorithm for the multilayer ANN, which performs the regression $\mathbf{D}_U^*$ to $\mathbf{D}^*$. We replaced this with the ELM (Huang et al., 2006), which is an efficient and not iterative training procedure. Conventional learning methods require seeing the training data before generating parameters of the hidden neurons, while ELM can randomly generate these in advance. With ELM we train a single-hidden layer feed-forward ANN randomly choosing the input weights and bias coefficients of the hidden nodes. Thereafter the output weights are computed analytically. The introduction of ELM allows the training in shorter time of a larger ANN, which determines faster mapping computation and higher accuracy. Over a large set of mapping test cases, the introduction of ELM training algorithm determined an average 98% reduction of the training time, a 87% reduction of the real-time mapping time, and a sensible improvement of the regression accuracy, measured with the Mean Squared Error (MSE). We use a growing algorithm to design the ANN providing a satisfactory regression. Starting with 10 neurons in the hidden layer, we grow the network by 10 units at a time, allowing a maximum of 200 neurons, until the MSE is below 0.01.

The homotopic uniform redistribution and of the ISOMAP dimensionality reduction are other computationally demanding components. In the functional prototype, described in Section 5, their implementation has been optimized to provide a tradeoff between accuracy and fast execution. Moreover we integrated synthesizer, analysis, mapping computation, real-time mapping, visualizations, and user interface in a single environment. This had improved the computation efficiency, reduced the data exchange overhead, and simplified the user workflow, executed in a single Graphical User Interface (GUI). Overall we reduced by 82% the mapping computation time, and by 94% the analysis automated computation time. The alternative interactive analysis determines no time overhead for computing the interactive timbre space. In **Figure 3** we illustrate the previous and current system implementations. The first was also part of a system for the vocal control of sound synthesis through timbre spaces. The seamless integration of the different components of the system further simplifies the procedure to compute personalized mappings through the timbre space, according to users' preferences and customized for their specific synthesizers.

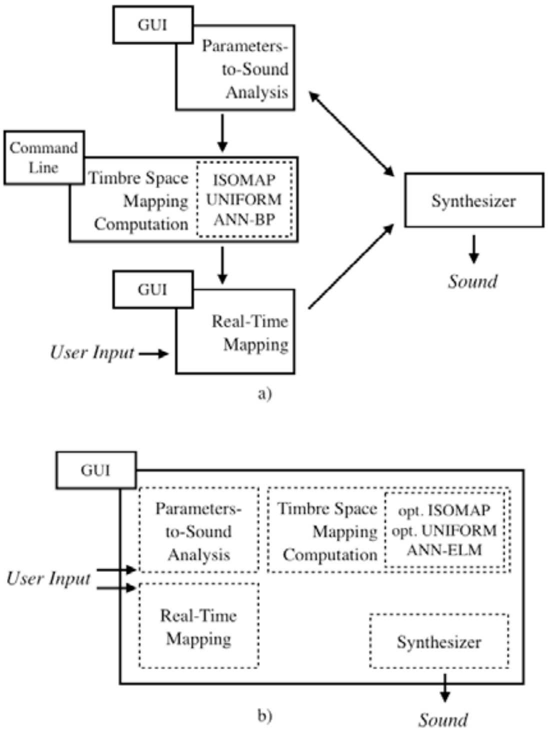


Figure 3: Illustration of a) previous and b) current system implementations. The latter, including improved mapping algorithms, presents a single GUI and seamless integration of the different components.

The proposed timbre space mapping method provided satisfactory performances (evaluated in Fasciani, 2014; Fasciani & Wyse, 2015). We measured quantitative metrics such as the percentage of retrievable synthesis parameter permutations, parameter continuity, and timbre space losses over an extensive set of synthesizers. The computational optimizations and usability improvements described here did not determine any performance drop.

5. Fully-functional integrated prototype

The sound-to-parameter analysis, the mapping computation, and the real-time mapping have been implemented in a standalone and fully functional prototype, presenting an integrated GUI, which exposes system settings and mapping options. Prototype, source code, user guide, screenshots, and demo videos are available at http://stefanofasciani.com/tsm4vsts.html. The prototype is implemented in Max/MPS and it hosts VST software synthesizer using the native vst~ object. The FTM library (Schnell, Borghesi, Schwarz, Bevilacqua, & Muller, 2005) and ad-hoc externals are extensively used for the analysis and real-time mapping. The ISOMAP and homotopic uniform redistribution algorithms are particularly complex and have not yet been ported to Max/MSP. Therefore a part of the mapping computation is still implemented in a MATLAB script. However this has been compiled in a standalone executable that runs in background and communicates with Max/MSP via the Open Sound Control (OSC) protocol. This process is completely transparent to users. This system can be seen as a VST synthesizers wrapper that exposes an alternative control strategy, as illustrated in the diagram of **Figure 3** b). The VST and synthesis patch can be stored to and recalled from presets, which also save mapping, analysis and system settings. The GUI of the prototype, in **Figure 4**, includes advanced options to further customize the timbre space mapping. The system implementation provides sufficient flexibility to support different workflows and interaction strategies. The parameters-to-sound analysis includes five different modalities. Available analysis descriptors include loudness of Bark critical bands, MFCC, PLP, and spectral moments, with possible addition of variation range and periodicity. The mapping can be recomputed changing most timbre options without executing a new analysis. Users are provided with options to tune the mapping in real-time. An interactive color-coded visualization of the timbre space, as in **Figure 5**, supports the development of control intimacy with mappings generated unsupervisedly.

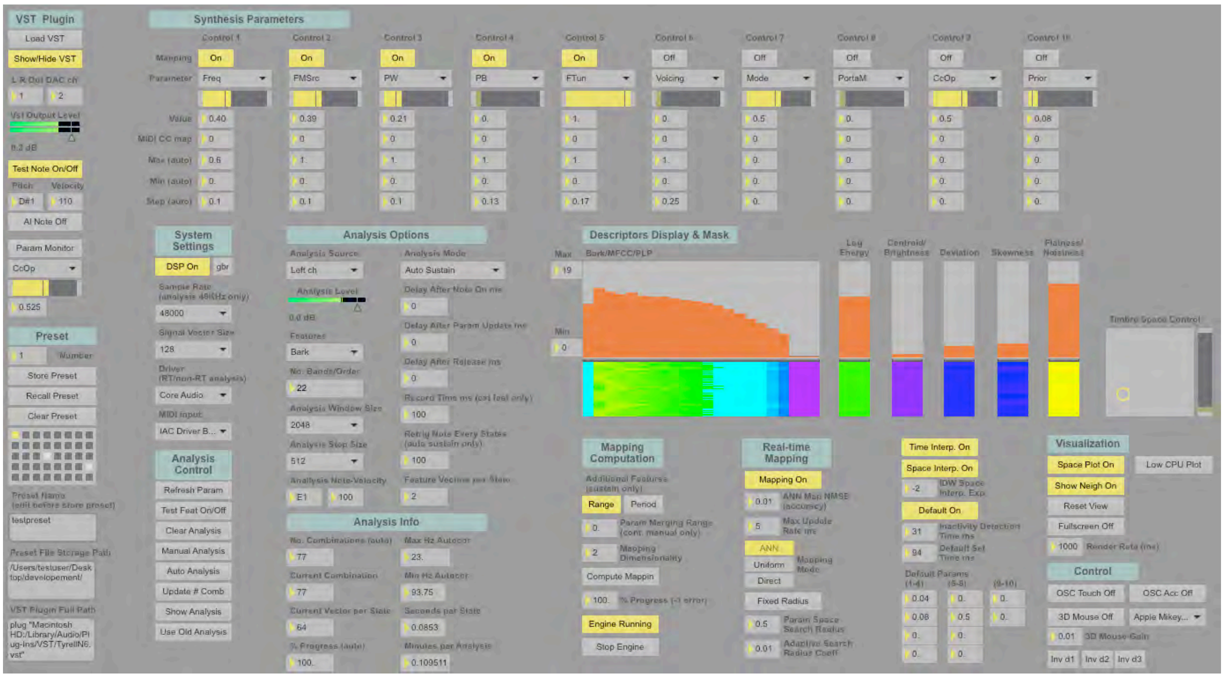


Figure 4: GUI of the fully functional prototype implementing the timbre space mapping for VST synthesizers.

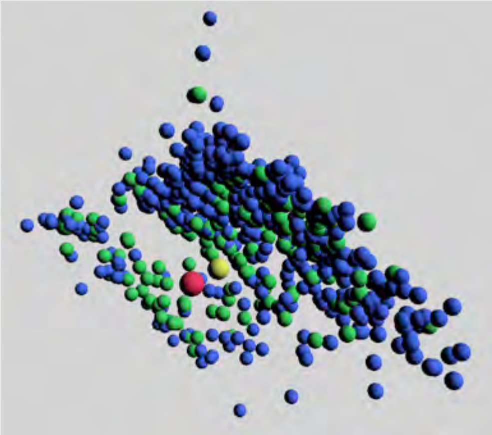


Figure 5: 3D interactive timbre space visualization. The blue spheres represent the entries in the reduced timbre space $\mathbf{D}^*$, the red one is the projection $m(\mathbf{c})$ of the user control with the yellow one being the closer $\mathbf{d}^*_j$. The green spheres represent the instantaneous restricted search space. Camera position and angle are user-defined.

6. Conclusion and future work

We have presented a method to interactively compute timbre space of specific sound synthesizers, and to use this as control structure, which provides a perceptually related interaction. Our previous studies on a system presenting a similar interaction approach, demonstrated that the control through the timbre space successfully hides the synthesis technical parameters, reduces the control space, and provides perceptual related interaction. However, further comprehensive user-studies are necessary to evaluate the interactive and unsupervised timbre mapping generation. The implementation of an open-source and fully functional prototype, compatible with any software synthesiser, can be a resource for performers and musicians. At the same time, it can be used as an exploration and tool by researchers and developers to study alternatives and improvement to the proposed method, including the use of different sound descriptors. The algorithm and the implementation can be further improved finding suitable alternatives to the ISOMAP and homotopic uniform redistribution algorithms, or introducing drastic optimizations. These can be ported in Max/MSP for the system integration in a single programming environment.

References

Arfib, D., Couturier, J. M., Kessous, L., & Verfaille, V. (2002). Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. Organized Sound, 7(2), 127–144.

Emmerson, S. (1998). Aural landscape: musical space. Organised Sound, 3(2), 135–140.

Fasciani, S. (2014). Voice-controlled interface for digital musical instruments (Ph.D. Thesis). National University of Singapore, Singapore.

Fasciani, S., & Wyse, L. (2012). Adapting general purpose interfaces to synthesis engines using unsupervised dimensionality reduction techniques and inverse mapping from features to parameters. In Proceedings of the 2012 International Computer Music Conference. Ljubljana, Slovenia.

Fasciani, S., & Wyse, L. (2015). Vocal control of digital musical instruments personalized by unsupervised machine listening and learning. Submitted, ACM Transactions on Intelligent Systems and Technology.

Fels, S. (2000). Intimacy and embodiment: implications for art and technology. In Proceedings of the 2000 ACM workshops on Multimedia (pp. 13–16).

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 61(5), 1270–1277.

Grill, T. (2012). Constructing high-level perceptual audio descriptors for textural sounds. In Proceedings of the 9th Sound and Music Computing international conference. Copenhagen, Denmark.

Hoffman, M., & Cook, P. R. (2006). Feature-based synthesis: Mapping acoustic and perceptual features onto synthesis parameters. In Proceedings of the 2006 International Computer Music Conference. New Orleans , US.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1–3), 489–501.

Jehan, T., & Schoner, B. (2001). An audio-driven perceptually meaningful timbre synthesizer. In Proceedings of the 2001 International Computer Music Conference. Havana, Cuba.

Klügel, N., Becker, T., & Groh, G. (2014). Designing Sound Collaboratively Perceptually Motivated Audio Synthesis. In Proceedings of the 14th International Conference on New Interfaces for Musical Expression. London, United Kingdom.

Lazier, A., & Cook, P. R. (2003). Mosievius: feature driven interactive audio mosaicing. In Proceedings of the 7th international conference on Digital Audio Effects. Napoli, Italy.

McAdams, S., & Bergman, A. (1979). Hearing musical streams. Computer Music Journal, 3(4), 26–43, 60.

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. Royal Society of London Philosophical Transactions, 336, 383–389.

Miranda, E. R., & Wanderley, M. M. (2006). New digital musical instruments: control and interaction beyond the keyboard. A-R Editions, Inc.

Moore, F. R. (1988). The dysfunctions of MIDI. Computer Music Journal, 12(1), 19–28.

Nguyen, H., Burkardt, J., Gunzburger, M., Ju, L., & Saka, Y. (2009). Constrained CVT meshes and a comparison of triangular mesh generators. Computational Geometry, 42(1), 1–19.

Nicol, C., Brewster, S. A., & Gray, P. D. (2004). Designing Sound: Towards a System for Designing Audio Interfaces using Timbre Spaces. In Proceedings of the 10th International Conference on Auditory Display. Sydney, Australia.

Pošćić, A., & Kreković, G. (2013). Controlling a sound synthesizer using timbral attributes. In Proceedings of the 10th Sound and Music Computing international conference. Stockholm, Sweden.

Puckette, M. (2004). Low-dimensional parameter mapping using spectral envelopes. In Proceedings of the 2004 International Computer Music Conference. Miami, US.

Risset, J. C., & Wessel, D. (1999). Exploration of timbre by analysis and synthesis. The Psychology of Music, 113–169.

Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., & Muller, R. (2005). FTM - Complex Data Structure for Max. In Proceedings of the 2005 International Computer Music Conference. Barcelona, Spain.

Schnell, N., Cifuentes, M. A. S., & Lambert, J. P. (2010). First steps in relaxed real-time typo-morphological audio analysis/synthesis. In Proceeding of the 7th Sound and Music Computing international conference. Barcelona, Spain.

Schwarz, D., Beller, G., Verbrugghe, B., & Britton, S. (2006). Real-time corpus-based concatenative synthesis with CATART. In Proceedings of the 9th international conference on Digital Audio Effects (pp. 279–282). Montreal, Canada.

Seago, A. (2013). A New Interaction Strategy for Musical Timbre Design. In S. Holland, K. Wilkie, P. Mulholland, & A. Seago (Eds.), Music and Human-Computer Interaction (pp. 153–169). Springer.

Tenenbaum, J. B., Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500), 2319.

Wessel, D. (1979). Timbre space as a musical control structure. Computer Music Journal, 3(2), 45–52.

Wishart, T. (1996). On Sonic Art. Harwood Academic Publishers.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 79-85.

Beyond the Keyboard and Sequencer: Strategies for Interaction between Parametrically-Dense Motion Sensing Devices and Robotic Musical Instruments

Jingyin He[*†], Jim Murphy[*†], Ajay Kapur[†], Dale A. Carnegie[†]

[*]New Zealand School of Music, Victoria University of Wellington, New Zealand

[†]School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

jon.he@vuw.ac.nz

Abstract

The proliferation and ubiquity of sensor, actuator and microcontroller technology in recent years have propelled contemporary robotic musical instruments (RMIs) and digital music controllers to become more parametrically dense than their predecessors. Prior projects have focused on creating interaction strategies for relatively low degrees-of-freedom input and output schemes. Drawing upon prior research, this paper explores schemes for interaction between parametrically-dense motion-based control devices and contemporary parametrically-dense robotic musical instruments. The details of two interaction schemes are presented: those consisting of one-to-one control (allowing the actions of a performer to directly affect an instrument) and those consisting of a recognition system wherein user-created gestures result in output patterns from the robotic musical instrument. The implementation of the interaction schemes is described, and a performance utilizing these schemes is presented.

Keywords: human-robot interaction, wearable sensor interfaces, musical robotics

1. Introduction

Contemporary robotic musical instruments (RMIs) have become more parametrically dense than their predecessors, with more actuators controlling greater numbers of musical parameters. As these new instruments present users with increasing numbers of actuators, they potentially afford composers and musicians greater expressive control than was possible with earlier systems. An increased number of parameters is accompanied by an increased level of difficulty in interfacing with them in real time: a system with many high-resolution output actuators will likely prove more time-consuming and difficult to interface with (in an intuitive and expedient manner) than a simpler RMI with few low resolution parameters. To address this problem, new mapping strategies between a human performer and a parametrically-dense RMI must be explored.

In attempting to create mapping strategies between human and robot, a challenge emerges: how to control what is essentially a complicated assembly of actuators in a manner as musically-sensible and relevant to a performer's skillset as possible. The goal of this paper is to explore this challenge, developing means of allowing the use of such complicated robotic instruments in performances through the use of parametrically-dense input interfaces. In doing so, a wearable input interface (designed by the first author) with many sensors is used to control a robotic output device with many actuators (designed by the second author). In the act of interfacing these two new I/O devices, a need was identified for a streamlined method of integration between them. To this end, new interfacing schemes were developed and a performance case study was undertaken. This work presents one of many possible mapping strategies that complicated I/O devices may employ.

Following a brief history of interaction schemes between human performers and RMIs, the hardware used in the performance case study is described. After introducing the hardware, the two human-robot performance paradigms explored in this paper are presented with implementation details, followed by a description of a performance featuring the hardware and software. Finally, this paper closes with a discussion of the work's outcomes and the potential for future work arising from that presented in this paper.

2. Background and Related Works

A challenge facing builders of robotic musical instruments is to create interaction schemes that allow users to create music with the instruments in a manner that affords creative exploration. The instruments themselves do not necessarily afford such interaction without significant intervention on behalf of the instrument builder: most of the actuators are electronically driven and numerically controlled. To enable musicians to use the instruments in a manner more similar to the instruments with which they are familiar, interaction schemes must be imposed upon the instrument. The work presented in this paper builds upon a history of such schemes.

Perhaps the fundamental way of defining a musical interaction paradigm for a musical robotic instrument lies in the mapping of the robot's outputs to a scheme based upon the MIDI keyboard control paradigm. In such systems, pitch-shifting elements are often assigned MIDI commands pertaining to pitch information. MIDI NoteOn and NoteOff commands may control actuation and damping mechanisms. In essence, the musical robot is mapped to respond to typical MIDI keyboard commands. While this system is widespread and applied to many systems both simple and relatively complicated (exemplified in and (Focke, 2011) and (Singer et al., 2004)), it fails to account for the current trend in increasing parametric density on contemporary musical robots (Murphy et al., 2015). As musical robots gain more parameters, the number of relevant MIDI commands falls short. Further, attempting to map a robot's functionality to keyboard-oriented MIDI schemes neglects to account for those systems that contain expressive parameters quite different from keyboard-like instruments.

To address the shortcomings of the manual MIDI mappings presented above, a number of instrument builders have created mapping schemes for their systems that attempt to more closely match a particular robot's performance capabilities. A drum robot, for example, might have a percussion-based means of performer input. Notably, third author Ajay Kapur has created a mapping scheme between his eSitar and his *MahaDeviBot* (Kapur, 2008). This scheme, exemplified in the live performance *Digital Sankirna*, consists of a direct mapping between *eSitar*-mounted and wearable sensors and the robot's actuators (Kapur et al., 2011). During performance, the human user's motions explicitly trigger a musical robot output in a "one-to-one" mapping scheme: pressure on one of the *eSitar*'s frets, for example, results in the direct triggering of a robotic drum mechanism. While this scheme is used to control the relatively parametrically-simple *MahaDeviBot*, it allows a performer to escape the MIDI keyboard or sequencer paradigm typical of numerous other musical robots. Musical robot researcher Gil Weinberg, who has created a number of instruments that allow for unusual performer/robot interactions, exemplifies a second mapping paradigm. In (Weinberg and Driscoll, 2006), for example, a human-played drum pattern is input into a system to allow a musical robot to "improvise" in a manner based upon the human-input pattern. While sophisticated, this system (like Kapur's) consists of a parametrically-simple low-degree-of-freedom musical robot as the output device coupled to a low-number-of-inputs input device. As the complexity of the robotic instrument increases, the challenge of developing musically-useful mapping schemes grows accordingly.

These three mapping paradigms encompass the majority of existent human/robot mapping schemes. While the MIDI keyboard-inspired scheme is used on numerous devices, it was deemed insufficiently applicable to contemporary parametrically-dense robot systems. The one-to-one performer/robot mapping and the higher-order mapping schemes mentioned in the preceding paragraph, however, are appropriate for new instruments with large numbers of controllable output parameters. These schemes are further explored below.

3. Interaction System Overview

With higher degrees of freedom afforded by new sensors, actuators and computing power, it is now possible to interact more intuitively with robotic musical instruments through musical gestures in both sonic and control paradigms. Two interaction schemes, based on musical gestures to control parametrically-rich RMI, were investigated in a performance context. The first interaction scheme utilizes the method of mapping sensor data of the input device into musical gestures. These gestures are then parameterized as instructions for the actuators of the RMI. The second interaction scheme involves the RMI playing a musical phrase (either stochastically generated or precomposed) based on previously played phrases and user input. Following a description of the hardware and software used in this project, the subsequent subsections detail these interaction schemes.

3.1 *Kontrol*, *Tangle* and *Swivel 2*

The input interface used in this work is the *Kontrol* hardware and software system, while the output interface used is the *Swivel 2* mechatronic chordophone connected to the *Tangle* software suite (which allows for for human-RMI interaction via MIDI). **Table 1** details the parameters of *Kontrol* and *Swivel 2*.

The *Kontrol* physical gesture acquisition system is a novel motion-based wearable (seen in **Figure 1**) that measures finger flexion and physical dynamics of the hand (He, et al., 2015). It samples data at 100 Hz and sends it wirelessly to the laptop. The software processes and inputs the incoming sensor data to classifiers. The classifiers then categorize the hand postures and gestures of a Guqin (plucked 7-string Chinese instrument) performer in real-time.

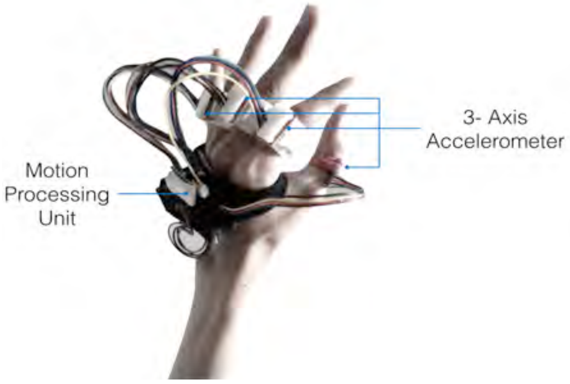


Figure 1. Kontrol wearable interface

Table 1. Summary of hardware parameters of Kontrol and Swivel 2

Kontrol Physical Gesture Acquisition System
Finger: 3-axis Accelerometer. Measures the flexion/extension of the fingers' metacarpophalangeal joint with respect to gravity.
Back of left palm: 9-axis Motion Processing Unit. Measures the pronation/supination, flexion/extension and ulnar/radial deviation of the left wrist, and acceleration with respect to gravity.
Xbee series 1 module. Transmits serial data to the host computer wirelessly.
Swivel 2 module (note: the complete instrument consists of 6 of these modules)
Fretter Arm Control: Miniature servo actuator. Positions fretter arm along string length for pitch shifting.
String picker: Miniature servo actuator. Rotates a guitar pick against the module's guitar string.
Clamper for fretter: Miniature servo actuator. Presses the fretter arm against the string, changing the string's pitch.
Damper for string: Miniature servo actuator. Rotates a foam pad against the vibrating string, dampening it.

Swivel 2 is a RM chordophone designed with high-resolution actuators that allow for the precise continuous control of the instrument's parameters. Illustrated in **Figure 2**, these parameters include as damping, fretting position, fretting depth, and picking (Murphy et al., 2013). *Swivel 2* is controllable through MIDI using the Tangle musical robot abstraction framework.

Tangle, described in more detail in (Mathews, et al., 2014), allows for user-defined MIDI, Open Source Control (OSC), or serial input to be mapped to RMI-specific output scheme. The *Tangle* framework is designed to be flexible regarding both input and output, and may be configured to listen for any input pattern and output in a scheme suitable for varying RMI communications methods. In the system described in this paper, *Tangle*'s role is to provide an easy-to-use interface between the classification system and the high number of actuators present on *Swivel 2*, allowing for simplified communications between *Kontrol*-related software and *Swivel 2*-related software.

A typical data flow between *Kontrol* and *Swivel 2* is outlined in **Figure 3**. During a live performance, the performer wears *Kontrol*, which serially transmits the measured data to the host computer using wireless Xbee Series 1 modules. The gesture recognition system outputs hand posture classifications and motion data such as linear body acceleration and jerk without gravity. The output values are then sent via OSC to *Tangle*, in which the data is interpreted and converted to the MIDI-formatted data used to communicate with the microcontrollers responsible for the control of the actuators of *Swivel 2. Swivel 2* utilizes single-coil pickups (like an electric guitar) to capture the sounds produced and sends the audio signal back to the host computer through a Firewire audio interface for further processing. A chain of audio effects to effect the original audio is set up in either a DAW (such as Ableton Live) or a music programming language (Max/MSP or ChucK). Additionally, audio effects are controllable by the performer by mapping either the raw sensor data or the output of the gesture recognition system to the parameters of the audio effects.

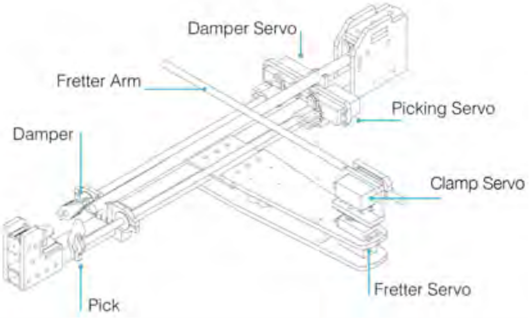


Figure 2. Module of Swivel 2 (the complete instrument consist of 6 modules)

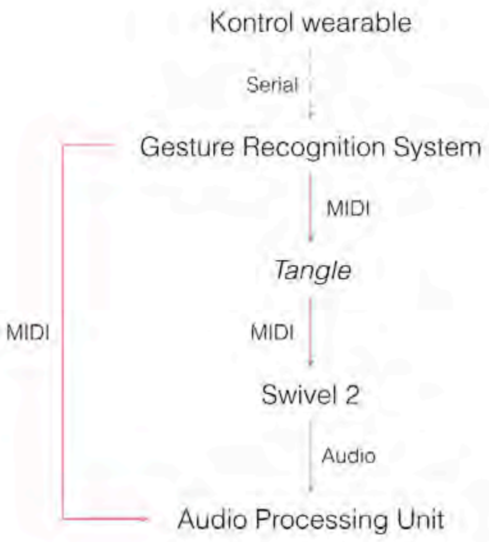


Figure 3. Data flow of interactive system

3.2 Connecting the dots between *Tangle* and *Kontrol Gesture Recognition System*

As mentioned in earlier sections, the interaction schemes deployed aim to explore the interaction paradigms beyond the traditional one-to-one mapping schemes between performer and robot.

In the first interaction scheme (IS-1), abstractions of sensor data as hand postures and gestures from the *Kontrol* system are mapped to the abstractions of actuating mechanisms of *Swivel 2* through *Tangle*. This mapping scheme addresses the parameterization between abstractions at the *sound object* time scale, whereby sonic events with duration from a fraction of a second to several seconds occur (Roads, 2001). A general mapping scheme between abstractions is as follows: If

Posture A is active, then *n* numbers of strings (up to six) are fretted at *f1, f2… fn* position(s) on string(s). An example would be that a clenched first gesture could lead to all strings being fretted with depth determined by wrist flexion, and with each of those being fretted at different points based on the shape of instantaneous finger flexion data. While the wrist radial/ ulnar deviation offsets each of the fretter positions, high jerk in wrist motion would result in the excitation of the strings by the RMI's string picker. **Table 2** shows a summary of the mapping used in IS-1.

Table 2. General scheme of parameterizing N number of postures to actuating Swivel 2

Kontrol	Tangle + Swivel 2
Hand posture [0]	No strings fretted
Hand posture [1]	1 string fretted, 4 strings touching
Shape of instantaneous finger flexion data	Fretter position
Wrist flexion	Fretting depth
Hand posture [2]	2 string fretted, 3 strings touching
Shape of instantaneous finger flexion data	Fretter position
Wrist flexion	Fretting depth
…	…
Hand posture [*N*]	*N* strings fretted, (5 - *N*) strings touching
Shape of instantaneous finger flexion data	Fretter position
Wrist flexion	Fretting depth
Wrist ulnar/ radial deviation	Offset of fretter position

In interaction scheme 2 (IS-2), the user's hand postures and gestures are mapped to control *Swivel 2* in the *meso* timescale, whereby phrases of musical/ sonic structures of various lengths measured in minutes or seconds occur (also presented in (Roads, 2001)). Each posture and gesture is mapped to trigger a sequence, while their motion data are used to alter the parameters of the sequence such as playback speed. Additionally, the postures and gestures also progress a score (similar to that of Max Mathew's score following with the Radio-Baton (Mathews, 1989)). The score following system used here is capable of branching, allowing for nonlinear compositional progression at the performer/composer's discretion as seen in **Figure 4.** The pre-composed variable-length sequences are grouped according to their sonic characteristics, and the sections' labels do not follow an order of progression. Each variation in each of the sections can be accessible from any section based on the requirements set for each sequence. The requirements are based on how many performed postures and/or gestures have been detected. Along with the requirement threshold for triggering each sequence, the histogram of performed postures and gestures is reset each time a new sequence is triggered (regardless of whether it is within or beyond the section).

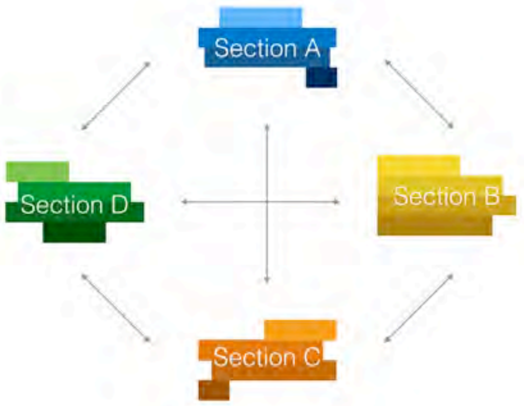


Figure 4. A network of sequences for non-linear compositional progression.

4. Applications

The interaction system presented above premiered at the solo recital performance of acclaimed Guqin performer Wu Na at the Adam Concert Room, New Zealand School of Music. The main objectives of the performance were to: 1) demonstrate the system outside of the laboratory in a live performance situation, 2) investigate the immediacy of the integration of the interaction system, and 3) explore the sonic possibilities of the presented human-robot interaction schemes in improvisatory music performance. The instrumentalist did not have any prior rehearsals with the system, although there had been discussion about the general form of the piece.

Figure 5. Wu Na, during IS-2, waiting for Swivel 2's response (https://vimeo.com/129071867).

The piece had two distinct movements characterized by its musical outcome influenced by the interaction system employed. The first movement was presented as introduction and exposition. During this movement, *Swivel 2* was controlled using the traditional interaction methods of on-the-fly triggering and editing pre-composed sequences. The implication of this model of interaction in musical outcome was that the human performer was mostly leading the improvisation, while *Swivel 2* assumed a more supportive role. In the second movement, the Guqin performer put on the *Kontrol* interface as *Swivel 2* finished the remainder of its current ongoing sequence. The new interaction schemes were then deployed, and the roles of the performers (human and robot) were observed to evolve during the second movement. With IS-1 in place, the performer was observed to intuitively dictate *Swivel 2*'s output through the use of both traditional and non-repertoire playing techniques. Conversely, with IS-2, *Swivel 2* is observed to be more autonomous, leading to a more dynamic conversation with the Guqin performer - influencing, complementing and contrasting the musical intentions of the human performer. The interaction in this case is bi-directional, resulting in tension and release, as well as complimentary and contrasting human-robot interaction. This is observed to a degree not present in IS-1.

Through this performance, it is observed that the three aforementioned objectives were satisfactorily met. In meeting the first objective, the system performed in a stable, glitch-free manner during a real-world performance scenario. In providing this system to a performer with little prior knowledge of the apparatus, a compelling and sonically-interesting live improvisation was performed in a manner that complimented the repertoire of both traditional and extended techniques of Guqin performance. This indicates that both the second and third goals of the *Kontrol-Swivel 2* interaction schemes were also met.

5. Conclusion and Future Works

When attempting to control a densely-parameterized musical robot with a parametrically-rich output interface, a difficulty emerges in mapping the output device's values to the input parameters of the robot. The work presented in this paper has sought to address this challenge through the use of a number of interaction strategies, presenting new abstraction and mapping schemes for live performance use.

With the interaction schemes presented in this paper, a performer may gain a level of real-time access and control over the RMI that they may not have had they been required to manually input each actuator control value. Indeed, where many prior musical robots are essentially treated as output devices to be connected to traditional sequencers, the interaction strategies presented herein allow musicians to gain direct and musically-relevant control over and otherwise-complicated instrument. In addition to providing musicians with this musically-relevant control, gesture-to-sound schemes provide audiences with an overview of performative cause and effect between musician and digital instrument that is difficult to achieve within the aforementioned sequencer-like paradigm.

Though the work presented in this paper has focused on a particular performance paradigm, it is hoped that it is merely the first in a series of works featuring novel output devices controlled by novel input devices. Such would serve to bridge the authors' research group's creative avenues. To go toward allowing any new musical output interface to control any new RMI, the Tangle musical robotic framework must be further extended to be applicable to any new devices. After extending Tangle, the creation of new musically-relevant mapping strategies may be explored. Additionally, *Kontrol* hand gesture recognition framework should also expand its recognition framework to other instruments such that performance techniques of other instrumentalists or even dancers may be explored.

References

Focke, A. (2011). Trimpin: Contraptions for Art and Sound. Marquand Books, Seattle, Washington, 1st edition.

He, J., Kapur, A., and Carnegie, D. (2015). Developing a physical gesture acquisition system for Guqin performance. Proceedings of 2015 International Conference on New Interfaces for Musical Expression. Baton Rouge, Louisiana.

Kapur, A. (2008). Digitizing North Indian Music: Preservation and Extension using Multi- modal Sensor Systems, Machine Learning and Robotics. VDM Verlag.

Kapur, A., Darling, M., Diakopoulos, D., Murphy, J., Hochenbaum, J., Vallis, O., and Bahn, C. (2011). The machine orchestra: An ensemble of human laptop performers and robotic musical instruments. Computer Music Journal, 35(4):1–15.

Mathews, M. V. (1989). The conductor program and the mechanical baton. In M. Mathews & J. Pierce (Eds.). Current Directions in Computer Music Research (pp. 263-282). Cambridge, MA: The MIT Press.

Mathews, P., Murphy, J., Carnegie, D., Kapur, A. (2014). Tangle: a flexible framework for performance with advanced robotic musical instruments. In Proceedings of the 2014 Conference on New Interfaces for Musical Expression. London.

Murphy, J., Kapur, A., and Carnegie, D. A. (2013). Swivel 2: A systems overview and evaluation of a new mechatronic slide guitar. In Proceedings of the 2013 Engineering New Zealand Conference (ENZCON), Auckland, New Zealand.

Murphy, J., McVay, J., Mathews, P., Carnegie, D., Kapur, A. (2015). Expressive Robotic Guitars: Developments in Musical Robotics for Chordophones. *Computer Music Journal 39(1)*, 59-73.

Roads, C. (2001). Microsounds. The MIT Press.

Singer, E., Fedderson, J., Redmon, C., and Bowen, B. (2004). Lemur's musical robots. In Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME), Hamamatsu, Japan.

Weinberg, G. and Driscoll, S. (2006). Toward robotic musicianship. Computer Music Journal, 30(4):28–45.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 86-88.

A Brain-Computer Music Interface for Music Expression

Rafael Ramirez, Sergio Giraldo, Zacharias Vamvakousis

Music Technology Group , Universitat Pompeu Fabra, Barcelona, Spain

rafael.ramirez@upf.edu

Abstract

Active music listening is a way of listening to music through active interactions. In this paper we present an expressive brain-computer interactive music system for active music listening, which allows listeners to manipulate expressive parameters in music performances using their emotional state, as detected by a brain-computer interface. The proposed system is divided in two parts: a real-time system able to detect listeners' emotional state from their EEG data, and a real-time expressive music performance system capable of adapting the expressive parameters of music based on the detected listeners' emotion. We comment on an application of our system as a music neurofeedback system to alleviate depression in elderly people.

Keywords: Brain-computer interfaces, emotion, expressive music performance

1. Introduction

In recent years, active music listening has emerged as a study field that aims to enable listeners to interactively control different aspects of music in real-time. Most of the work in this area has focused on controlling music aspects such as playback, equalization, browsing and retrieval. However, there have been very few attempts to control the expressive aspects of music performance, such as pitch, timing, amplitude and timbre. The manipulation of these sound properties is clearly distinguishable by the listeners who often experience strong emotions as a consequence of this. Expressive music performance research (for an overview see Gabrielsson, 1999, 2003) investigates the manipulation of these sound properties in an attempt to understand how it conveys emotion and to recreate expression in performances.

The study of emotions in human-computer interaction has increased in recent years. This is due to the growing need for computer applications capable of detecting the emotional state of users (Picard, 2002). Motivated by every day interaction among humans, a great part of the research in this area has explored detecting emotions from facial and voice information. Under controlled conditions, current emotion-detection computer systems based on such information are able to classify emotions with considerable accuracy (Takahashi, 2004). However, emotions are not always manifested by means of facial expressions and voice information. Facial and voice information is related only to behavioral expression, which can be consciously controlled and modified, and which interpretation is often subjective. A still relatively new field of research in affective brain-computer interaction attempts to detect a person's emotional state using electroencephalograms (EEGs) (Chanel, 2006; Choppin, 2000).

In this paper we go one step beyond brain activity data sonification by mapping brain activity data into a high-level emotional space description, and use this description to control expressive aspects of music performances. The result is an expressive brain-computer interactive music system for active music listening, which allows listeners to manipulate expressive parameters in music performances using their emotional state, as detected by a brain-computer interface. We describe the application of our system as a music neurofeedback system to alleviate depression in elderly people.

2. A Brain-Computer Music Interface for Music Expression

Our approach to emotion-based expressive brain-computer music interaction is depicted in **Figure 1**. The system consists of a real-time feedback loop in which the brain activity of a person is captured as an EEG signal, the signal is processed in order to filter some frequency bands and to compute the estimated emotional state of a person as a coordinate in the arousal-valence 2D space. This coordinate, together with a music score or audio file, is the input to a previously trained expressive music performance model, which outputs an expressive rendition of the score/audio. The expressive performance output is presented to the user as feedback of his/her emotional state.

For acquiring the brain activity of the user, we use the Emotiv EPOC headset, recently released by the Emotiv Company (Emotiv, 2014). This headset consists of 14 data-collecting electrodes and 2 reference electrodes, located and labeled according to the international 10-20 system (Niedermeyer, 2004).

From the EEG signal of a person, we determine the level of arousal, i.e. how relaxed or excited the person is, by computing the ratio of the beta and alpha brainwaves as recorded by the EEG. In order to determine the valence level, i.e. negative or positive state of mind, we compare the activation levels of the two cortical hemispheres. Details about arousal and valence calculation can be found in Ramirez et al. 2012.

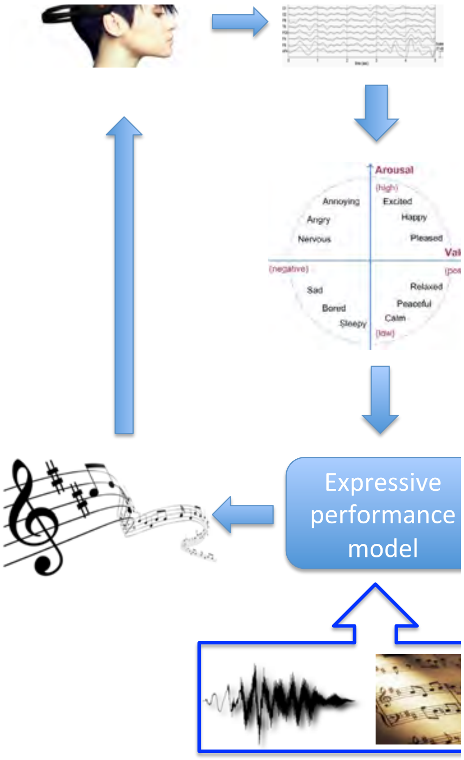


Figure 1. System overview

We trained four expressive performance models (one for each emotion, happy, sad, angry and tender) using machine learning techniques. In order to train the model we have collected a training set of music scores of pieces and recorded performances of the pieces in the four different emotions by a professional musician. The modelled expressive performance actions are duration ratio (ratio between the score duration and the performed duration), the energy ratio (ratio between loudness of a note and average loudness), and articulation ratio (level of staccato - legato).

In order to apply the expressive model with an arousal-valence coordinate as input, we interpolate (using a linear regression model) the four trained (happy, sad, angry and tender) models. This allows us to apply intermediate models, not just the four trained models. The performance actions, i.e. duration, energy and articulation, are calculated using the model and the interpolated coefficients based on a control position input on the arousal valence plane.

3. A music neurofeedback application

The proposed expressive brain-computer music interface has been applied as a music neurofeedback system for treating depression in elderly people. Ten adults (nine female and one male, mean=84, SD=5.8) with normal hearing participated in the neurofeedback study consisting of ten sessions (two sessions per week) of fifteen minutes each. Participants listened to music pieces preselected according to their music preferences, and were encouraged to increase the loudness and tempo of the pieces, based on their arousal and valence levels, respectively: Pre and post evaluation of six participants was performed using the BDI depression test, showing an average improvement of 17.2% in their BDI scores at the end of the study.

4. Conclusions

We have presented an expressive brain-computer interactive music system for active music listening, which allows listeners to manipulate expressive parameters in music performances using their emotional state, as detected by their brain activity as an EEG signal. The proposed system is divided in two parts: a real-time system able to detect listeners' emotional state from their EEG data, and a real-time expressive music performance system capable of adapting the expressive parameters of music based on the detected listeners' emotion. For acquiring the brain activity of the listener, we use the low-cost Emotiv EPOC headset, and we determine his/her emotional estate by computing arousal and valence values from the alpha and beta waves in the prefrontal cortex. The expressive performance model is trained using recordings of musical pieces in four emotions (happy, relaxed, sad, and angry). The input of the system is the coordinate consisting of the computed instantaneous arousal and valence values. The coefficients of the four expressive models (one for each emotion) are interpolated to obtain the prediction of the performance actions, which serve as input for synthesis. Finally, we have briefly described the application of the proposed system as a music neurofeedback system to alleviate depression in elderly people.

Acknowledgements

This work has been partly sponsored by the Spanish TIN project TIMUL (TIN2013-48152-C2-2-R)

References

Chanel, G., Kronegg, J., Grandjean, D., Pun, T.: Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. In: Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) MRCS 2006. LNCS, vol. 4105, pp. 530–537. Springer, Heidelberg (2006)

Choppin, A.: Eeg-based human interface for disabled individuals: Emotion expression with neural networks. Masters thesis, Tokyo Institute of Technology, Yokohama, Japan (2000)

Emotiv Systems Inc. Researchers, (2014). URL http://www.emotiv.com/researchers/.

Gabrielsson, A. The performance of music. In: Deutsch, D. (Ed.), The Psychology of Music, second ed. Academic Press (1999).

Gabrielsson, A. Music performance research at the millennium. Psychol. Music 31 (3), 221–272 (2003)

Niedermeyer, E., da Silva, F.L.: Electroencephalography, Basic Principles, Clinical Applications, and Related Fields, p. 140. Lippincott Williams & Wilkins (2004)

Picard, R.W., Klein, J.: Toward computers that recognize and respond to user emotion: Theoretical and practical implications. Interacting with Computers 14(2), 141–169 (2002)

Ramirez, R., Vamvakousis, Z. (2012). Detecting Emotion from EEG Signals Using the Emotive Epoc Device, Brain Informatics, LNCS 7670, pp. 175–184.

Takahashi, K.: Remarks on emotion recognition from bio-potential signals. In: 2nd International Conference on Autonomous Robots and Agents, pp. 186–191 (2004)

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 89-94.

Assembling Music

Yoichi Nagashima

Department of Design, Shizuoka University of Art and Culture, Japan

nagasm@suac.ac.jp

Abstract

In this paper, I report on my three projects - sound installation, new musical instruments and new platform of computer music - the common key concept is "assembling music". Musical performance is a kind of expansion of human-body activity. I have been developing many kinds of musical instruments as a part of my composition, and I focused essentially on human "assembling" action in music performance in these projects. Of course, musical composition is an assembling process of musical parts. However, I want to expand the concept of "assembly" - not only for composition but also for performance. I hope this report expands the possibilities in interaction design for media art, and hope to discuss the technological detail and artistic approach.

Keywords: musical building block, expanded LittleBitsSynth, assembling performance

1. Introduction

Musical performance is a kind of expansion of human-body activity, for example - beating, rubbing, bending, pushing, pulling, repelling, etc. I have developed many kinds of musical instruments as a part of my composition, sometimes with "contactless” interface like a Theremin. But now, as CV (computer vision) using a Webcam or Kinect for music interface is very popular, I do not follow the trend.

In this paper, I will show some experimental works of "Assembling Music". Of course, the musical composition is essentially an assembling process of musical parts. I want to expand the concept of "assembly" – not only for the composition but also for the performance. The human action of assembling something is familiar in many industries, but is beautiful in productions by experts. Children enjoy assembling and disassembling building block toys, because this simple behavior is essential for human culture. In this paper, I report on three experimental works of musical interface - an installation work, new instruments and a new platform for improvisational composition/performance.

2. Installation work ‘Color Orchestra’

This sound installation (or musical instrument) was created by Ryu Junhee who was an undergraduate student in my seminar. He targeted the "musical building blocks" as a toy, and the system had a total 44 colored wooden blocks and a wooden base machine (playing table).

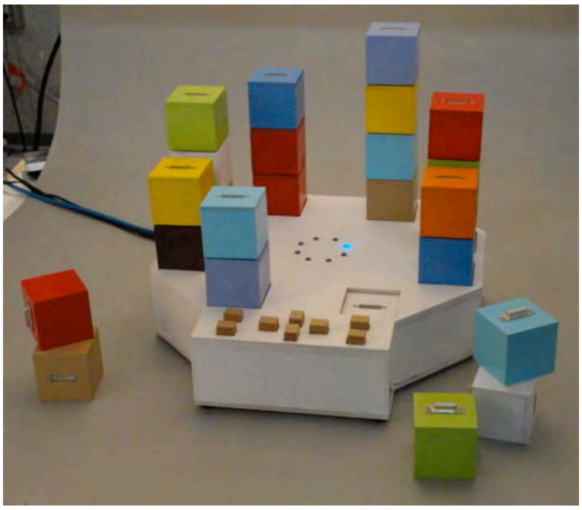


Figure 1. The installation "Color Orchestra"

Figure 1 shows the system which had a circle of eight bright blue LEDs, flashing in sequence in the center of the table. On the table, there was a shallow recessed square depression outside each of the eight LEDs. The shallow recessed square depression acts as a "connector" for blocks which people can pile up on the table freely.

At first, there are no blocks on the stage, and the blue LEDs flash sequentially as musical tempo alone. If people stack one block on one location, the assigned sound related to the color is generated at the moment the assigned blue LED flashes. Because there are eight points of sound generation, this instrument is "8-beat" sequencer.

People can stack any number of blocks on any location, thus the generated sound is the combination of stacked blocks - four blocks for each of eleven colors which means four percussion (BD, SD, HH, clap) and a seven-note scale (C, D, E, F, G, A, B). If same-color blocks are stacked in the same location, the volume of the assigned sound doubles. People can push other control buttons on the stage of the system – to change the tempo, to change the timbre of scale notes, to change total loudness and to change the depth of echo effect.

Technically, this system is realized by Max/MSP patch connected by microprocessor AKI-H8 inside of the system via MIDI. Each block has a 15-pin DSUB male connector at the bottom of the block and has a 15-pin DSUB female connector at the top of the block. These blocks can be piled up (connected) higher on top of each other and can easily be disconnected. Each block receives the common "address" signal from the system via DSUB connector, and the decoder IC (inside the block) generates an output signal "I am here!" to the common signal line with an open-drain buffer (wired-OR connection).

The superior specification of this system is realtime assembling / disassembling performance of musical blocks – improvisation on the stage. After finishing the creation of this installation work, I composed a new work of computer music using this system as a musical instrument. **Figure 2** shows the performance of the new work called "Joyful Boxes" on stage. Behind the stage, the realtime 3D graphics (created by Open-GL of Max/jitter) was projected on the screen - which is synchronized with the beat of eight blue LEDs of the system.

Figure 2. The performance of "Joyful Boxes"

At the beginning part of "Joyful Boxes", I performed like a child playing a block-building toy with the beat of blue LED, making sounds, and moving CG on the screen. As the music scene proceeds, I hold one block and sing/shout something with the microphone – then the system works as a sampler of the voice, and the assigned sound of the block changes into my realtime sampled voice. When I pile up the "sampled voice" color blocks, my voice is mixed with other sounds of blocks. This musical performance means "assembling sound itself" and "assembling musical elements" of musical beats with improvisation. You can view this performance on YouTube (Nagashima 2012).

3. New Instruments "GHI2014"

This section is a report of a new instrument from GHI project, called "GHI2014". In 2007, I proposed a new concept/project called "GHI" - meaning "It might be good that musical instrument shines, wouldn't it?" in Japanese.

I selected three policies while looking straight at the essence of musical instruments as a new approach. (1) Musical instruments are a tool of music, and are good partners for musical expression in the performance. (2) For natural musical instruments, the sound is radiated, and not only the audience but also the player receives the sound from the whole space. (3) The player's expression mesmerizes an audience from not only the aural route but also the visual route. The idea of a new project arose here. I produced a new instrument called "Cyber Kendang", and new piece (using it) was selected and performed by the author at NIME07 concert in NYU (**Figure 3**: Nagashima 2007).

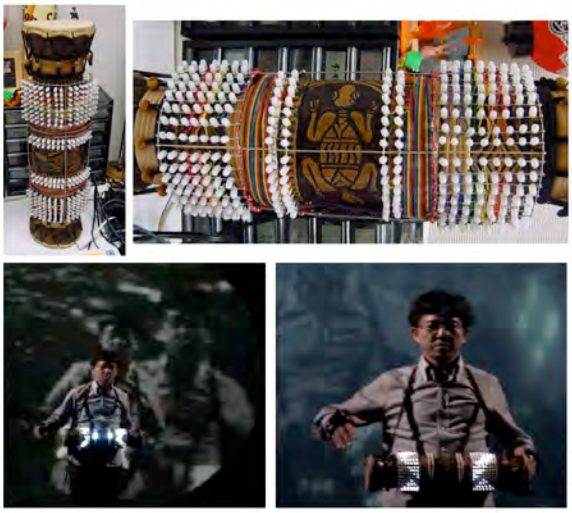


Figure 3. The performance of "Cyber Kendang"

I had reported that the "Propeller" processor is a good platform to realize multimedia interactive systems (Nagashima 2009). In 2013, I supported my student's project – "Shining one-piece dress". Her concept was very simple and cute - to produce a black semi see-through one-piece dress which shines with many internal white LEDs. I developed a system with Propeller processor as 24-channel LED driver for individual PWM control (**Figure 4**). I used the reel-type LED ribbon, which contains 3 powerful white LEDs per 5cm/block. The 24 output ports were driven by 2SC1815 for 12V60mA (3 LED) or 12V120mA (6 LED) each, and controlled by fast PWM driver module (originally developed).

Inspired by that project, bright LED ribbons and Propeller PWM driver, I have decided to start "GHI" again. I aimed for a new instrument with the same concept, self-many-shining instrument called "GHI2014". The dress had a total of 304 LEDs, but "GHI2014" has a total of 864 strong LEDs. The impact of this light is very expressive, and sometimes performers use sunglass on stage. In addition to this, I aimed to add one idea. I have developed many other instruments and brought them on world tours, and my big theme was the portability of instruments. Small instruments/interfaces like iPhone are easy to carry, but bigger-scale instruments have more effect/impact on stage, so this is the essential opposition.

Figure 4. The "Shining one-piece dress"

A good idea came to me – an instrument which can be carried while decomposed and can be assembled/used. This disassembly/assembly process was a good concept for my composition using it later. Eventually the overall shape of this instrument became two 12 ridges outlining the shape of an octahedron side-by-side. Each ridge line is a square pillar which has 15 LEDs on each of 4 surfaces. All edges of these ridge lines are separable, with small strong Neodymium magnets. The LED blocks of the total 24 ridge lines of the double octahedron are realtime controlled with PWM via MIDI. Around this sculpture, 6 ultrasonic range finder sensors can be setup, and the output is also MIDI. **Figure 5** shows the shape of "GHI2014".

Eventually, the GHI2014 contained two Propeller drivers. The interesting specification of the GHI2014 are the three modes of the interface – standalone mode, sensor mode and instrument mode. In the standalone mode, the 1st Propeller receives the distributed output from the 2nd Propeller (sensors), and reacts

Figure 5. New Instruments "GHI2014"

Figure 6. The performance of "GHI2014_Ogaki"

without MIDI input. So, GHI2014 works without a computer, works as an installation - detects human action and shines by itself. In the sensor mode, with all LED cables disconnected, GHI2014 works as six-channel ultrasonic range finder sensors (removable/re-settable) for other applications as interactive installations. In the instrument mode, the 1st Propeller can cut the distributed output from the 2nd Propeller by special MIDI command, GHI2014 works with the concept of "GHI", and shining pattern can be programmed with realtime generated sound/music/graphics performances.

I have composed a new work featuring this new instrument in 2014 (Nagashima 2014). At the beginning of the work "GHI2014_Ogaki", I performed with the "expanded LittleBitsSynth" (explained later) and started assembling this instrument. With the realtime-generated BGM (Rhythm section), each part of the instrument was flashing synchronized with the beat, and I connected them to construct the shape on stage. After the whole shape was built, I attached the six ultrasonic distance sensors, then the total 24 ridge lines of the double octahedron started shining with my performance, as controlling the Theremin. This assembling performance was an important part of the work, so the live camera image (captured by staff) was projected on the screen behind the stage. From this projection, the audience were able to understand the relation between the assembling and the sound. **Figure 6** shows the performance on stage.

4. The "expanded LittleBitsSynth"

In 2013, the release of "LittleBitsSynth" (http://littlebits.cc/kits/synth-kit) was an exciting development in both the Computer Music community and the Sketching (Physical Computing) community. I arranged the "LittleBitsSynth" for expanded expression in computer music as a concept of "Assembling Music", and I will introduce the details here. Figure 7 shows the performance.

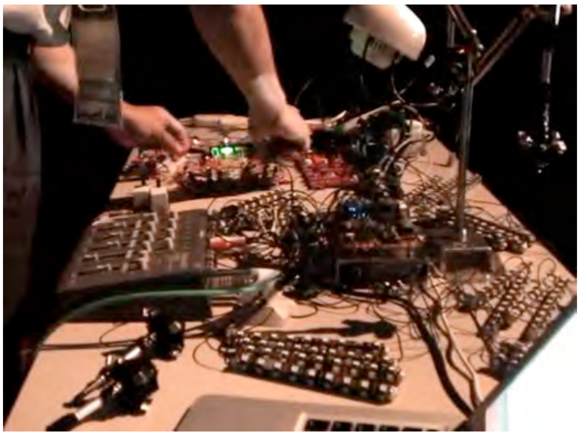


Figure 7. Performance with "LittleBitsSynth"

At first, I cut the connecting cables and inserted male-female a "pin header" connector between them. Then, I connected cables from three types of sensors (infra distance sensor, 2D acceleration sensors, 3D acceleration sensors) to the "pin header" connectors. Thus, I could insert special external sensors, or could pass through (directly connecting). **Figure 8** shows the close-up.

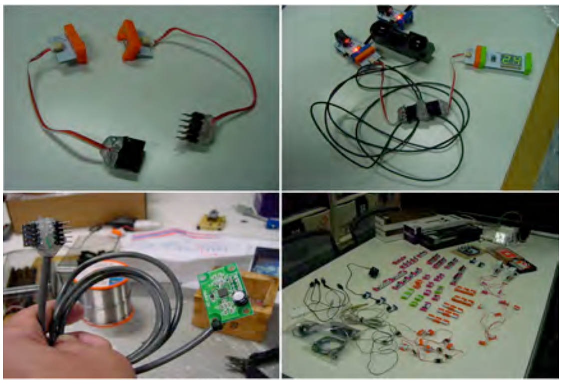


Figure 8. Inserting sensors/connectors

Next, I arranged the "LittleBits Arduino" (**Figure 9**) for usefulness/controllablility in the system. The "LittleBits Arduino" has [3 input and 3 output] voltage ports using the "bitSnap" magnetic connector, and pin holes of other ports on the board – compatible with standard Arduino. We can program it with the Arduino IDE – like Arduino UNO, Arduino Pro, etc. However, the programmed Arduino works only as a standalone system, not as a peripheral system with Max environment. Therefore, I replaced the firmware of "LittleBits Arduino" with Firmata (http://firmata.org) which is a protocol for communicating with microcontrollers from software on a computer. The protocol can be implemented in firmware on any microcontroller architecture as well as software on any computer software package.

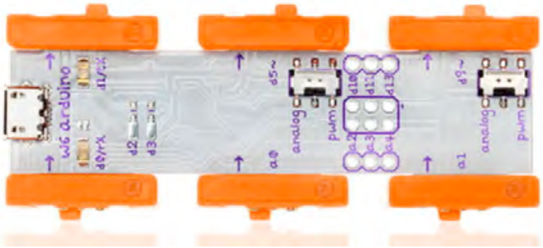


Figure 9. LittleBits Arduino

Then, I arranged the Max object/patch "maxuino" (http://maxuino.org) as a simple/useful tool. **Figure 10. Upper)** shows the default maxuino patch, but I analyzed and modified the internal "maxuino.js" object, and finally produced the original "3 Arduino controller" maxuino-like patch (**Figure 10. Lower**). In this way, we can control the "LittleBits-Arduino running Firmata" by using "arranged maxuino" just like as Gainer.

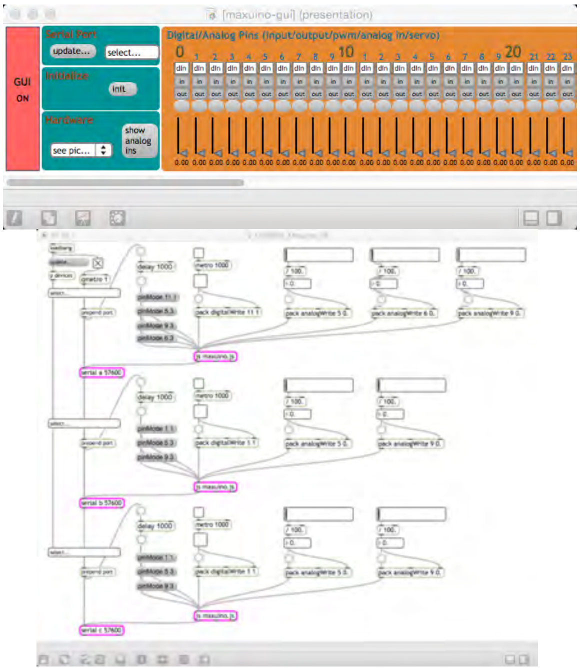


Figure 10. The "maxuino" (default vs. arranged)

The "LittleBits-Arduino(Firmata)" and "arranged maxuino" was a very smart combination with Max, but there was a serious problem – the magnetic connector of LittleBits. The "bitSnap" magnetic connector is a very smart and funny idea for our "Joy of Assembly/Disassembly", and the "slight" momentary contact failure is almost OK for voltage connection as sensor (control), sound (pitch, amplitude, waveform). Humans cannot recognize the momentary skip of the voltage. On the other hand, the "LittleBits-Arduino (Firmata)" and the "arranged maxuino" system are undermined by a serious obstacle, because the momentary contact failure causes the RESET of Arduino, and the Max patch (maxuino) loses sight of the "serial" object and requires a restart of the patch. Of course, in the music scene, this is not acceptable.

Based on my experience and knowledge of computer music over 20 years, the most reliable connection is MIDI. Thus I decided to develop a new system to control LittleBitsSynth system from Max without “LittleBits-Arduino (Firmata) + arranged maxuino". This system was constructed with new a microcontroller "mbed" - NucleoF401RE - ARM processor developed by open source tool on the Web. **Figure 11** shows the system - the upper left is six channel voltage output ports, upper right is 10 channel voltage input ports, lower left is MIDI IN/OUT and three channel trigger output ports and lower right is my wiring. With this system, the realtime control is widely expanded with LittleBitsSynth and the reliability is very good. This system is robust and tough for improvisational assembling and disassembling performance on stage. I have plans for composing and performing in the future using this system.

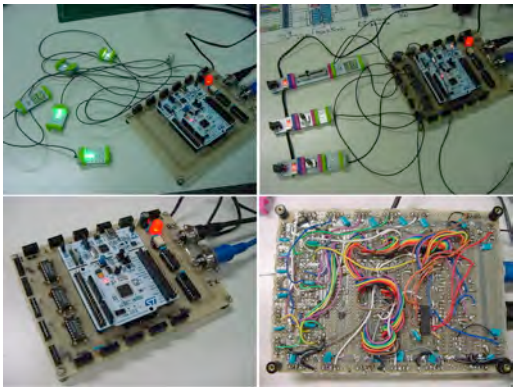


Figure 11. LittleBits Controller (mbed)

5. Demonstration

After this paper was accepted by the organizer of Soundislands Festival, I was requested to give an additional demonstration presentating my instruments at the ArtScience Museum on Sunday. I am planning three demonstrations there.

The first demonstration will be a live performance of "GHI2014". I will construct a musical instrument during the music performance time, and perform live sound /light generation.

The second demonstration will be a live lecture of "LittleBits-Arduino(Firmata)" and "LittleBitsSynth controller by mbed". I will show some techniques of the Open Source Culture in computer music.

The third demonstration – somewhat related to my theme – will be the recent two novel instruments. One is the "Double Myo" sensors, and the other is a "New tactile instrument which has not yet been named". Both will be accompanied by a simple performance. In particular, the latter instrument has 10 unique tactile sensors, and it is a useful case study of "mbed" as a controller.

6. Conclusion

I have reported on my three projects – sound installation, new musical instruments and a new platform for computer music. The common key concept is "assembling music". I hope to expand the concept of "assembly" – not only for composition but also for performance. I hope this report expands the possibilities in interaction design for media art.

References

Yoichi, Nagashima. (2012). Piece "Joyful Boxes". URL http://www.youtube.com/watch?v=GiOohVnoAoc

Yoichi, Nagashima. (2007). "GHI project and [Cyber Kendang]". Proc. NIME 2007.

Yoichi, Nagashima. (2007). "GHI Project : New Approach for Musical Instrument". Proc. ICMC 2007.

Yoichi, Nagashima. (2009). "Parallel Processing System Design with [Propeller] Processor". Proc. NIME 2009.

Yoichi, Nagashima. (2014). Piece "GHI2014". URL
http://www.youtube.com/watch?v=sRhxdGaSQfE,
http://www.youtube.com/watch?v=c1e1U4tTkcU,
http://www.youtube.com/watch?v=wh_iztSy-B8

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 95-97.

Spandex Shoji Synthesizer Transforming Elastic Interaction into Images and Sounds

Shumpei Tamura, Yasuo Kuhara*

Department of Interactive Media, Tokyo Polytechnic University, Japan

*kuha@int.t-kougei.ac.jp

Abstract

We developed an interactive artwork generating images and sounds using the traditional Japanese interior fixture known as Shoji. We used spandex fiber, which is a polyurethane fiber having remarkable elasticity, as a Shoji screen. When we push and expand the elastic fiber using our hand, it generates images and sounds according to the degree of its expansion and contraction. The images are projected onto the area of the Shoji as a screen touched by a performer. We can enjoy the interaction of the elastic feeling synchronized with the images and sounds by touching the Shoji screen.

Keywords: interactive art, sound Installation, shoji, spandex, elasticity, Kinect

1. Introduction

Shoji is a traditional Japanese interior fixture. It is a sliding door or window made of wooden lattice pasted with Japanese paper to make a screen. In contrast to a normal door, the translucency of its screen enables more light to be let in while it is closed. Additionally, different from a window, because of its lack of transparency we cannot clearly see objects on the other side of the screen but only their silhouette images depending on the light condition. Because of these characteristics, Shoji is a typical interior fixture directing the Japanese beauty of shade and shadow in a dim light.

The Japanese have a proverb: “Walls have ears; Shoji have eyes.” This saying represents a peeping person, who rips the paper to make a small hole in the Shoji by poking it with their finger in order to look at the other side through it. Adequate translucency of Shoji stimulates a feeling of curiosity about the other side.

Some projection mapping works have used Shoji as a screen, for example, “Shouji ni Mary.” However, these are movies that are mainly for watching without touching the Shoji. On the other hand, “Fire wall” is an interactive artwork whereby we can touch the screen made of soft fabric causing projection of images onto the screen and playing of music. In this work, the action of pushing the fabric is used to trigger the playing of movies and change the speed and volume of music based on the depth, therefore the degree of pushing strength is not reflected in musical expressions of performers. Additionally, “Humanaquarium” is a movable performance space with solid touch screens combined with projection responding to audience interaction. In this work, players can not feel a soft and elastic touch because of a glass window.

In the case of Shoji, when you push the paper of the screen with your finger, you can feel the elasticity of the Shoji against your finger because of its stretching properties. The harder you push it, the greater the feeling of tension against your finger whether the paper is ripped or not. Our work was inspired by Shoji filled with a Japanese emotional atmosphere, and we pursue the graphical and musical expression of the sense of tension arising from the elasticity when we touch Shoji.

2. Concept

This work is an interactive artwork that reacts to poking the Shoji screen with one's finger. We emphasize this action of "poking Shoji with one's finger." If you have ever seen Shoji in childhood, you have probably felt the temptation to rip the Shoji paper by poking it with your finger. This feeling may arise from certain causes, which are curiosity about the other side of the screen and the material of soft paper in which a hole can be easily made. Shoji as an interior fixture has many factors stimulating various emotions. This work aims to express these deep feelings for Shoji.

This work uses spandex fiber, which is made of strongly elastic fabric, for the Shoji screen. In the case of paper, we have to repair it each time someone rips the paper of the screen. Using white spandex fiber, we can enjoy the action of poking Shoji with our finger repeatedly while keeping the appearance of the Shoji. Additionally, we can enjoy the elastic force of this fabric, and we use it as a method of interactive art transforming the elasticity into images and sounds.

3. Configuration

This work comprises projection mapping onto a spandex fiber screen attached to the wooden lattice of Shoji, which is configured by a grid square with nine cells of 3 x 3 as shown in **Figure 1**. Spandex is made from polyurethane fiber known for its extraordinary elasticity. Because of this feature, it is often used for sportswear. This work is composed of a Shoji main unit, Microsoft Kinect for a depth sensor measuring the displacement of the screen, a projector, speakers, and a PC controlling the entire system as shown in **Figure 2**. We use Max/MSP/Jitter software and jit.freenect.grab of its external object for grabbing the Kinect image.

When you push the Shoji screen with your finger, images and sounds synchronized with the elasticity that you feel are made. The depth sensor of Kinect, which is located on the other side, measures the deviation from the normal position of the screen in proportion to how hard you push it with your finger. The deviation is transformed to images and sounds. As images, contours of the color gradient corresponding to the deviation are drawn on the screen as shown in **Figure 3**. For example, if you push the screen hard, the color of the image is displaced toward yellow, and if you push the screen gently, the color of the image is displaced toward red.

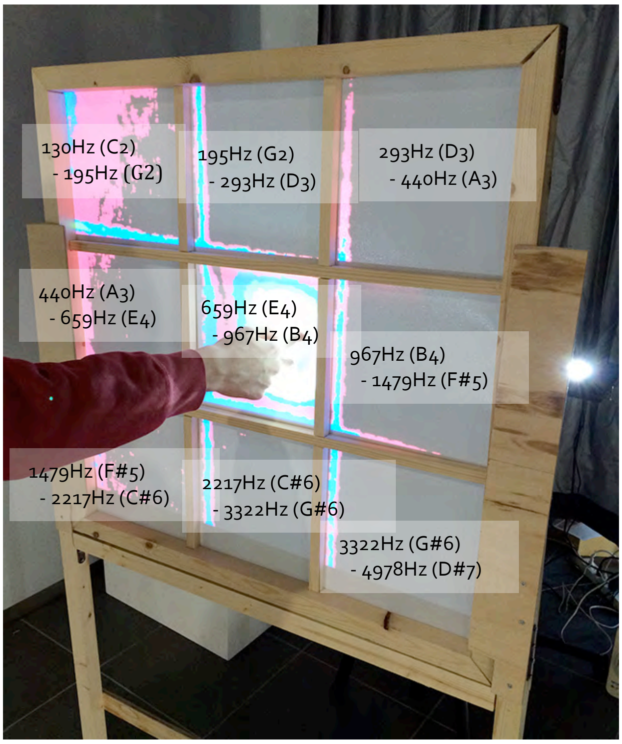


Figure 1. Overview and example of frequency bands.

At the same time, the sound of sine waves is generated with images as described above. Nine different frequency bands are allocated to the nine square grid cells of the Shoji screen as shown in **Figure 1**. For example, we assigned various perfect fifth intervals, which are 130 to 195Hz, 195 to 293Hz, 293 to 440Hz, 440 to 659Hz, 659 to 967Hz, etc. If you push a square cell of the Shoji, the sound of the sine wave with the allocated frequency are generated. If you push a cell deeply, the pitch of the sound is shifted higher in proportion to the deviation of the cell made by your hand. If you push multiple cells, a sound synthesized with multiple sine waves of different frequencies is made. You can feel the synchronization of the

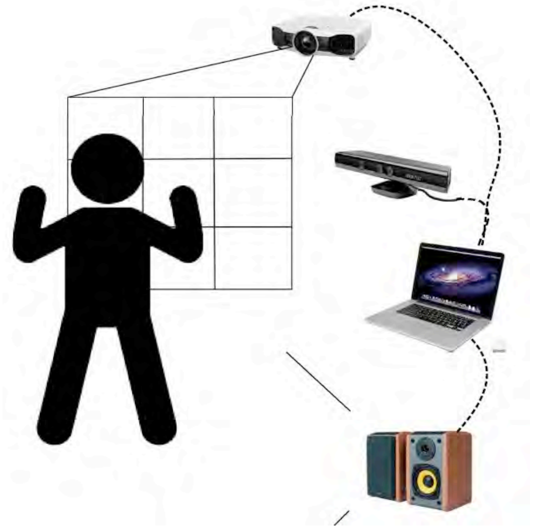


Figure 2. System configuration.

deviation of the screen that you push and the sound made, because the pitch of the sound corresponds to the force with which you push.

In this way, we can enjoy various images and sounds synchronized with the elasticity by touching the Shoji.

4. Performance

First, a performer stands in front of the Shoji and pushes the Shoji screen made of spandex. Then, the Kinect sensor detects the deviation of the screen, which is transformed into certain values for generating images and sounds. Next, the sound of the sine wave with the frequency according to the deviation is made. At the same time, contour images of the color gradient corresponding to the deviation are projected onto the cell of the screen touched by the performer.

Because of the extreme elasticity of spandex fiber, the screen expands and contracts in sensitive response to the performer's action of pushing or releasing. The surface of the screen wrinkles or fluctuates as you push hard or more gently. This fluctuation creates graphical effects such as ripples sensitively reflecting the degree of force by pushing the screen.

Moreover, you can poke the screen in the same way as ripping it to make a small hole in the paper of the Shoji with your finger. Paper is damaged if you push it too hard. However, in the case of spandex fiber, the screen is not damaged and can expand it to a large extent even if you push pretty hard. The harder you push, the more you feel the action-reaction force and the higher the pitch of the sound that is gradually shifted. As a result, this work generates a unique sense of tension similar to ripping Shoji paper with your finger. You can watch some movies of our works on Internet, at URL http://www.kuhalabo.net/shojisynth/.

5. Conclusion

We developed an interactive artwork expressing the features of Shoji, which is a traditional Japanese interior fixture. We enjoy the feeling of tension in elasticity caused by pushing the screen as well as the images and sounds synchronized with our action. In future, we will improve the color gradient to be finer by making the resolution higher. Additionally, we will install a low-frequency oscillator, or LFO, to some square cells of the screen, and generate more expressive sounds by synthesizing various waves.

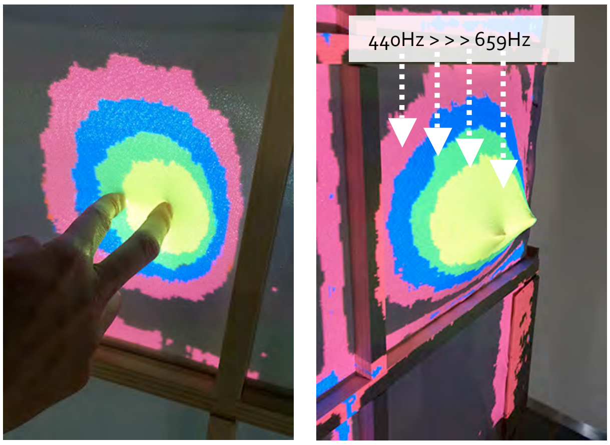


Figure 3. Image mapping and example of pitch shift.

References

Noya Kudo (2012). "Shouji ni Mary". URL http://yeah70.tumblr.com/

Aaron Schwarz (2013). "Firewall". Future Interfaces Conference, New York 2013.

Robyn Taylor et al. (2011). "Designing from within: humanaquarium", The ACM Conference on Human Factors in Computing Systems 2011.

array

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 98-100.

Si15

Having a Ball With the *Sphere*

Dirk Johan Stromberg[*], Robert Casteels[†]

[*]Department of Music, LASALLE College of the Art, Singapore

[†]Centre for the Arts, National University of Singapore

dirk.stromberg@lasalle.edu.sg

Abstract

Robert Casteels has composed a growing corpus of more than 100 musical works cross cultures, genres and disciplines. These range from miniature to large-scale works in the European tradition, as well as multidisciplinary works which combine European, Chinese and Indian instruments, as well as the gamelan, together with computer-generated sound and images. Instances of space for creative interaction between the performers in these compositions have been far and few between. The work, *time:space:*, has unveiled new possibilities for interaction and a blurring the line between performer and composer creating an artistic interconnectivity and interdependency.

Keywords: new instruments, composition, live performance

1. Introduction

A research grant from the National Arts Council (Singapore) enabled Casteels, together with electro-acoustician Dirk Stromberg and visual art director Andrew Thomas, to research and create two works, *2014:time:space:* and *2015:time:space:* (collectively referred to as *xxxx:timespace:*). Each was a one-hour work with "e-instruments" such as the *Sphere* which is presented in this article. Both works also involved dance, human voice, and video in three superimposed layers: a pre-recorded soundtrack of electro-acoustic music, the sounds produced by the electronic instruments and the human voice, and thirdly the transformation of these sounds (Casteels, 2014).

2. The *Sphere*

The *Sphere* is an electronic instrument built by Stromberg in 2014. It was developed from the physical modeling work of Nicky Hind and Eric Lindemann, whose research looks at synthesis approaches and sound parameters. Nicky Hind describes various models for physical modeling which involve:

> *Articulation* – the manner in which a sound is initiated, and:
>
> *Resonant body* – the manner in which a sound propagates within the physical structure of the instrument.

This has then been described in synthesis terms as:

> *Excitation source* – the initial source of a sound (i.e. a model of a blown or bowed sound);
>
> *Spectral characteristics* – The resonance of sound that models the behavior of different instrumental shapes and bodies (Hind, 2015).

As suggested by Lindemann, the sonic parameters for synthesis and physical models can be simplified as:

> *Spectral skew* - the tuning of partials;
>
> *Spectral parity* - the balance between even and odd partials;

Spectral tilt - the emphasis on either high or low partials and the degree of difference between the high and low partials;

Spectral angle - a subcategory of spectral tilt, the rate at which partials decay (Lindemann, 2015).

The *Sphere* was created with these concepts of physical modeling in mind. The goal was to create an instrument that allows for a performer to sculpt the sound through physical means. The choice of pressure sensors articulated with the fingers requires for the performer to use force in the creation of the sound. This is akin to the need for a wind player to blow and a string player to bow at different pressures to change register and color. The accelerometers require the player to use the wrists to orient the two halves of the sphere to further sculpt the sound, not unlike the changing of the angle of a bow or a plectrum. Keeping in line with the instrumental modeling, each parameter is interconnected. An example of this is the interdependent mapping of various aspects of amplitude, timbre and pitch. **Figure 1** shows the layout and types of sensors, as well as the parameters they control, and **Figure 2** shows the instrument in action.

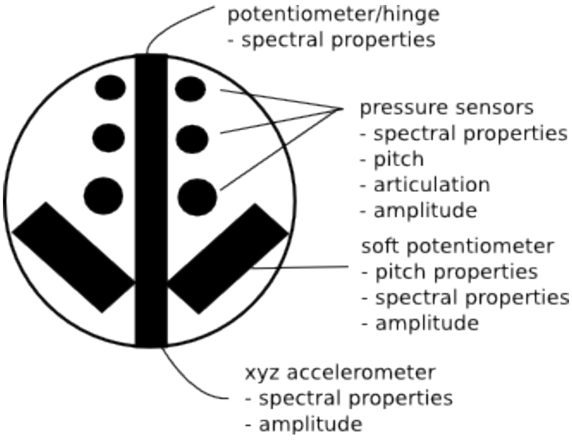


Figure 1. Diagram of the *Sphere*.

3. The *Sphere* in *xxxx:time:space:* and its interconnectivity

The musicians (on performing on the *Sphere* and singers) produced sounds. These sounds were processed, transformed and redistributed in the performance space in real time via a 22-speaker diffusion system with a custom software built on psychoacoustic principles. The musicians reacted to these transformed and spatially distributed sounds.

The musicians depended on the sound engineers to control the dynamics, the timbral transformation and the spatial distribution of the sounds from the mixing desks. The sound engineers depended on the sounds produced by the musicians. In summary: in order to function as a composer and performer in the new sonic and visual environment of the *xxxx:time:space:* project, Casteels had to adapt his tactile pianistic sense to the sensitivity of the ten sensors on *The Sphere*, had to relinquish levels of control to sound engineers and had to change the type and level of interactivity with fellow performers.

The traditional labor division between composers, performers and engineers is blurred. Already in 1936, French composer Edgar Varèse noted the solidarity between scientific development and the progress of music when he declared: “I dream of instruments obedient to my thought and which with their contribution of a whole new world of unsuspected sounds, will lend themselves to the exigencies of my inner rhythm” (Schwartz, 1998).


© Sheryl Stromberg

Figure 2. Manipulating the *Sphere*.

4. The score

The *Sphere* player and singers not only reacted to sounds but also anticipated sonic events of the pre-recorded track thanks to a score visible on a computer screen. This score consists of a waveform. Performers and engineers also had to relate to the images that were projected on four screens (Casteels, 2014). **Figure 3** shows the cover of the score.

Figure 3. Cover of the score of *2014:time:space*

5. Further development

The *Sphere* was a stepping-stone in the approach of instrument development through the exploration of tactile feedback and its musical implementation in various synthesizes. The use of pressure sensors was successful because an acceptable level of expressiveness and interconnection between performer and instrument was reached. The physical structure of the instrument was frail and the reapplication of the instrument is limited. The outcome is to develop a more sturdy and universal instrument design system that allows for instrumentalists to create instruments, rather than impose a design on a performer.

References

Casteels, R. (2002, February 12). URL Robercasteels.com. Retrieved July 2, 2015.

Casteels, R. (2014, October 15). Xxxx:time:space:. Retrieved July 2, 2015, from URL http://timespace.sg

Casteels, R. (2014, December 10). 2014:time:space:. Singapore: Casteels.

Hind, N. (n.d.) Physical Modeling Synthesis. Retrieved July 31, 2015, from URL https://ccrma.stanford.edu/software/clm/compmus/clm-tutorials/pm.html.

Lindemann, E. (n.d.) Synful Orchestra Users Guide. Retrieved July 31, 2015, from URL http://www.synful.com/Synful_Orchestra_Users_Guide.pdf.

The Liberation of Sound. (1998). In E. Schwartz & B. Childs (Eds.), *Contemporary Composers on Contemporary Music* (Expanded Edition ed., P. 196). New York, New York: Da Cap Press.

Thomas, A. [creativehybrid]. (7 March 2015). 2014:time:space: Premiere Live Performance [video file]. Retrieved from URL https://youtu.be/9IZkvs3XZVA.

array
ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 101-103.
Si15

Resonating Spaces

Paul Fletcher, Mark Pollard

School of Film and Television (Animation), School of Contemporary Music , VCA and MCM
University of Melbourne, Australia

fp@unimelb.edu.au

Abstract

Resonating Spaces is a collaborative research project titled "reconstructing the familiar". This research addresses the sonic, kinetic and visual resonances of specific locations. Through the creation of new perspectives we make the familiar unfamiliar and in so doing bring a sharper focus on our familiar surroundings of place, time and atmosphere.

Keywords: animation, music, resonating-spaces

Making the familiar unfamiliar is done with the intention of invigorating the observations of new and old locations of time place, time and phenomenon. The artists involved research a physical and experiential phenomenon as a location. The reconstruction process then undertaken by the artists includes collaborative interdisciplinary and trans-disciplinary research. Processes and ideas are exchanged for example between the specialist fields of audio, kinetic, haptic and visual artistic disciplines. Rather than isolated disciplines of specialization, these individual areas of study and production are enriched when treated as overlapping sensory fields. These trans-sensory responses and new understandings become embodied in the form of new artworks for instance film, performance, and installations incorporating animation and music composition. Today we bring to you two works captured as short films; *Gridlife* (2013) and *With Every Beat of its Wing* (2014; see **Figure 1**). *Gridlife* (see **Figure 3)** is a colorful and energetic reinterpretation of the grid patterns found across many aspects of city life. *Gridlife* responded specifically to the cultural and architectural space of the Ian Potter Musuem of Art in Melbourne. It was first screened as an installation at the Ian Potter Museum of Art. In its short film form it screened at Animex London (2015) and at Punto Y Raya Visual Music Festival in Iceland (2014). *With Every Beat of its Wing* (2014) was created for an installation at the Royal Melbourne Zoo. Butterfly flight patterns were filmed and observed in the Royal Melbourne Zoo's Butterfly House. The sound and animation was then constructed in response to these observational studies of oscillating patterns of color and movement. *We notice raindrops as they fall* (2015; see **Figure 2**) is a new work for 2015. This project examines and responds to observed, re-imagined and re-mapped patterns of raindrops falling. The project studies and re-imagines the trans-sensory characteristics of falling raindrops. The concert presentation of this work mixes improvisation with predetermined composition. There is a dynamic observation and reflection process in the live performance. Many aspects of the sound and video can be adjusted and recomposed in a live performance. In this performance, artists and audience will notice falling raindrops of sound and image.

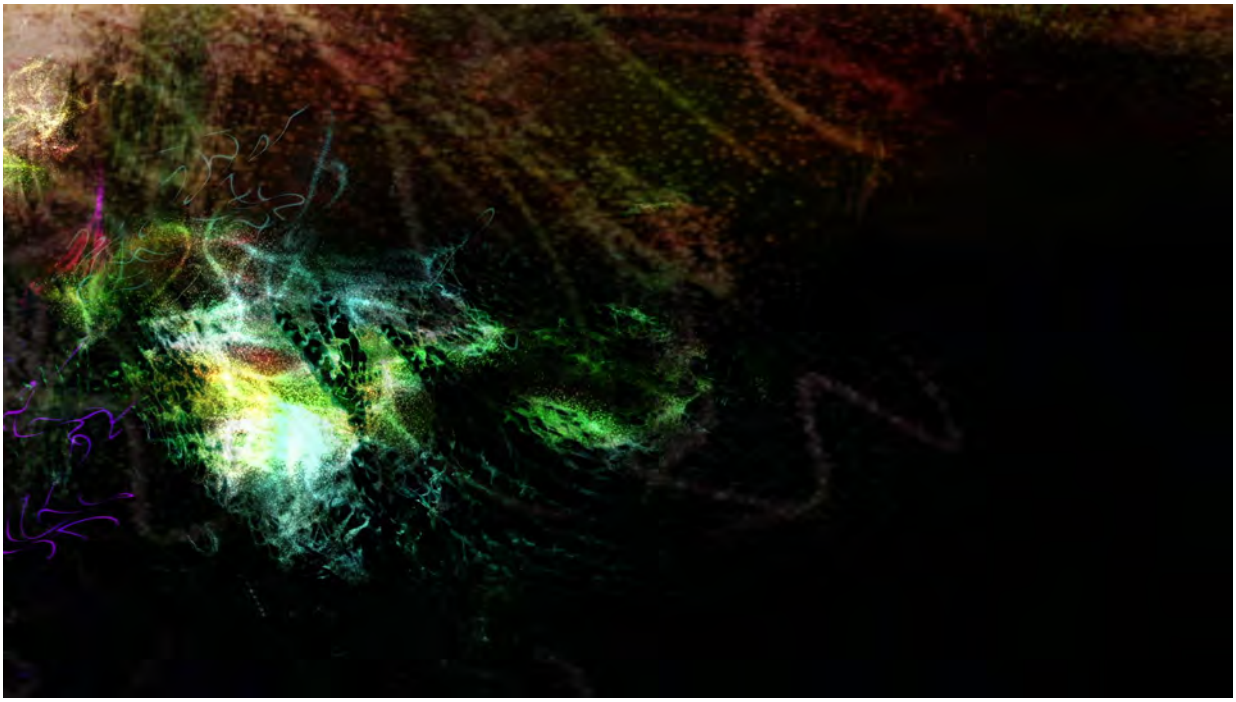


Figure 1. Still from *With every beat of its wing.* Mark Pollard, Paul Fletcher (2014)

Figure 2. Still from *We Notice Raindrops As They Fall*. Paul Fletcher, Mark Pollard (2014)

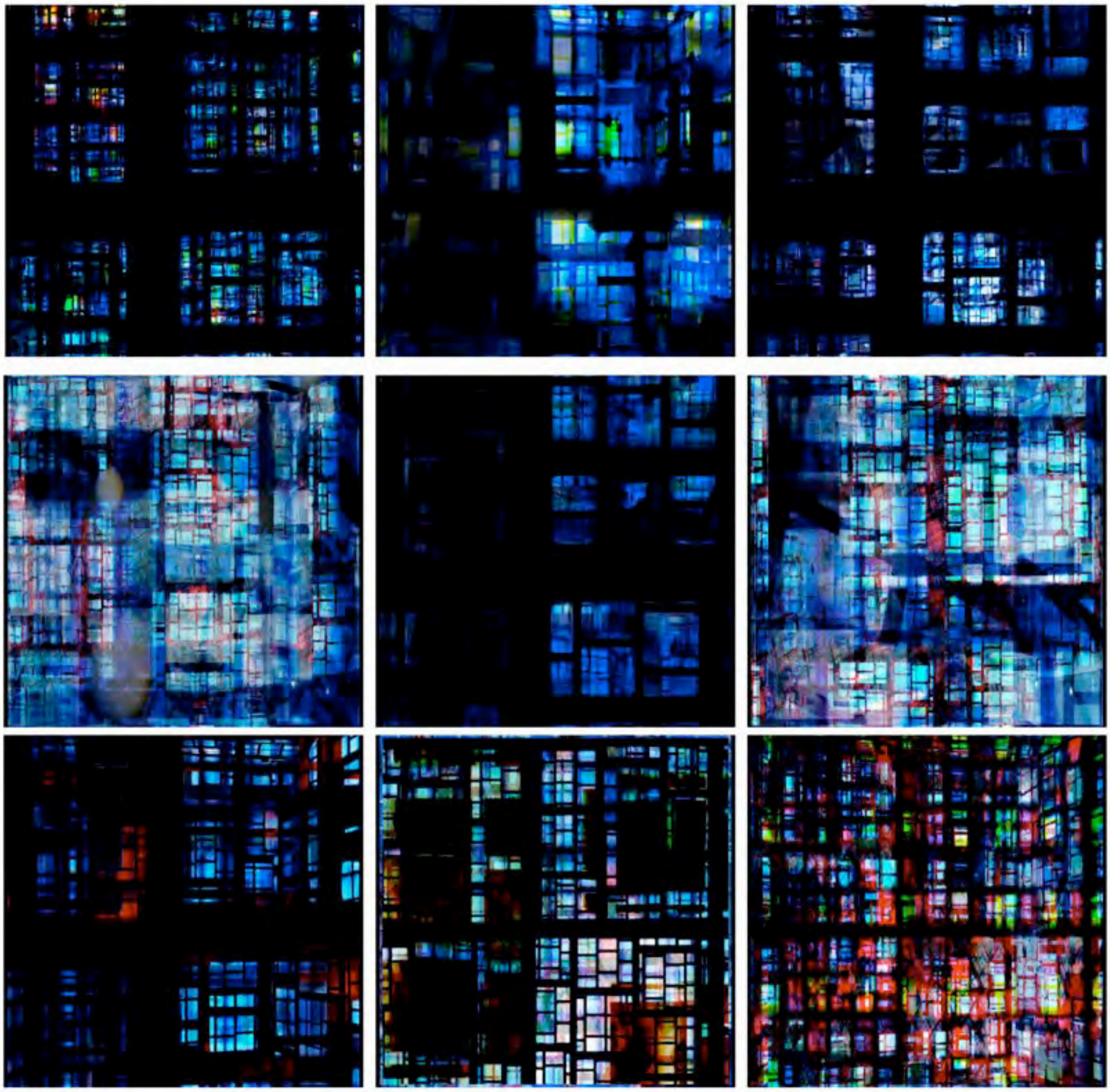


Figure 3. Still from *Gridlife*. Mark Pollard, Paul Fletcher (2014)

References

Fletcher, Paul. "Resonating Spaces We Notice Raindrops as They Fall - Process Notes." URL http://digitalcompost.net/weNoticeRaindropsAsTheyFall/weNoticeRaindrops.html.

Fletcher, Paul. "With Every Beat of Its Wing- Press Kit." URL http://www.digitalcompost.net/EveryWingBeat.

ICMA Array, vol 2016, Special Issue:
Proceedings of Si15, Singapore, August 2015, pp 104-107.

About *When We Collide*: A Generative and Collaborative Sound Installation

PerMagnus Lindborg [*], Joyce Beetuan Koh [†]

[*] School of Art, Design, and Media, Nanyang Technological University

[†] Nanyang Academy of the Fine Arts, Singapore

permagnus@ntu.edu.sg

Abstract

The idea for *When We Collide* sprang from Douglas Hofstader's metaphor of creativity as the meeting between records and record players, appearing in his 1979 book "Gödel, Escher, Bach: An Eternal Golden Braid". In our case, the records are soundfiles, whilst the record player is a generative system. The player analyses, selects, mixes, transforms, and spatialises the material created by the composers (monophonic and quadraphonic soundfiles). The system negotiates between algorithms that tend towards monotony (in terms of loudness, spatialisation, and frequency spectrum) and algorithms that tend towards variability (in terms of soundfiles, transformations, and scenes). In a nutshell, the installation is a space where sonic ideas collide and co-exist.

Keywords: generative, installation, composition

1. Introduction

When We Collide is a generative sound installation based on audio material contributed by six composers. It is a development from two previous productions: *On the String*, inspired by String Theory, commissioned by the Singapore Arts Festival 2010; and *The Canopy*, presented at International Computer Music Conference 2011 and World Stage Design Festival 2013. These works are about our fascination with the building blocks of the universe, in particular, the distinctive decay signatures of elementary particles. Imagine attributing their distinctive features to sonic characteristics. *What musical possibilities arise from the idea of particle collisions, interactions, and decay?*

Originally, *When We Collide* was conceived as an interactive installation of four independent large physical structures of strings, configured in the Chinese character 'ru' /入 which means 'to enter' (**Figure 1**). The visitor, in the role of a dancer-musician, would modify the sonic characteristics of a particle by movements. The image evoked is a *pas de deux* with an invisible, dark super partner - dancing elementary particles - yielding a multimodal experience that mediates between the eyes, ears and body. The construction of this scale became too resource intensive. Many discussions ensued and various guises of the installation were deliberated. One of the guises was inspired by Iannis Xenakis' drawings of the Montreal Exposition 1967 (**Figure 2**).

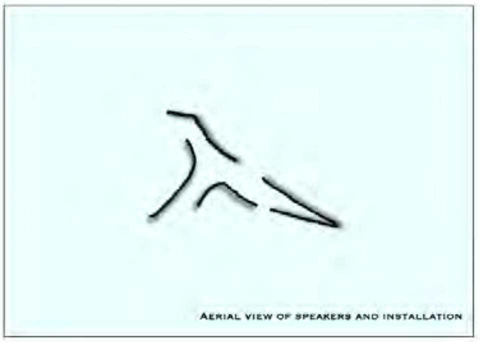


Figure 1. Installation draft (top-down view) of string structures resembling a Chinese character.

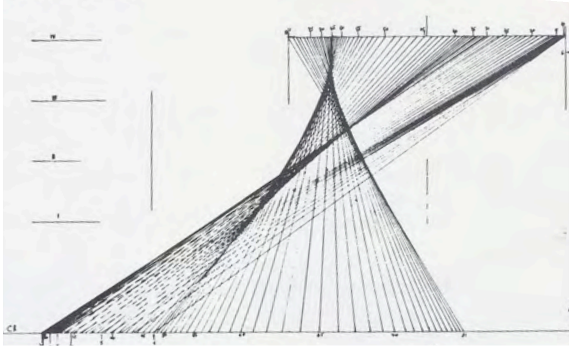


Figure 2. A visualisation of the pitch and architectural structures by Iannis Xenakis.

Our structure was to have strings fixed with stretch sensors. The intention was to 'wrap the space' in a specific pattern reminiscent of the particles decay signatures (**Figure 3**).

Figure 3. Mockup of the staging for *On The String*. Note the strings across the space.

2. A collaborative and generative installation

In *When We Collide*, we have revisited the concept of particle-collisions. Instead of simulating the concept with physical material, we could lean on our aural and spatial senses. What could be more apt than to have six composers, each has a distinct language, alluding to the decay signatures of particles, compose musical material for a generative system that is designed to allow collision, morphing and becoming. The invited composers were Andrián Pertout, Seongah Shin, Stefano Fasciani, Dirk Stromberg. The authors also contributed material. There were thus six artists in all, each of us providing five quadraphonic soundfiles and twenty monophonic soundfiles: quadraphonic files are 55 seconds long, and mono files 5 seconds.

3. Player design

The ideas behind the 'player' come from more than one source of inspiration; important to mention are John Cage's *Williams Mix*, the *Freq-out* installations curated by CM von Hauswolff (where one of us, Lindborg, participated), and our own previous work, in particular Lindborg's *Khreia* for orchestra (2002) and Koh's *FingerPrints* (2014).

First, we imagined the player simply as a jukebox - "put another nickel in, then you'll hear the music spin" - that would pick at random among the 30 quadraphonic files that the artists had composed. The selection mechanism developed into a small set of rules where the probability vector for the quadraphonic files is updated in cascade depending on which files have been played earlier. Also, the presently playing quad files limit which monophonic files can be selected, as well as when, and where (in the surround panning) they can appear (**Figure 4**).

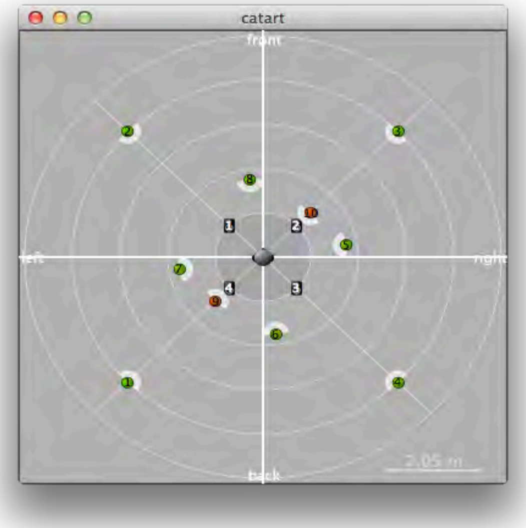


Figure 4. The surround panning space as it appears in the IRCAM spat user interface, in a 'bird's eye' perspective. The filled circles represent sound sources, and the squares outputs e.g. physical loudspeakers, or virtual sources for binaural reproduction.

4. Instruments

We defined four virtual instruments (or 'sub-players'), striving to integrate their functioning and sonic output, so as to yield the impression of an ensemble.

The first instrument, 'quadbase', consists of two quad sub-players taking turns, which may sometimes overlap by a certain amount. Their continuous output is analysed for loudness (using Ircamdescriptors) and main directivity, and a self-correcting mechanism tries to adjust the spatialisation to make the overall output as even as possible.

The second and third instruments, nicknamed 'recomposer' and 'cloud', are both based on *CataRT* (Diemo Schwarz 2015). They try, in different ways, to imitate what the quad players are doing. One selects among the 120 mono files originally submitted by the composers, and the other chops these files up into approximately 3000 fragments in order to produce a cloud-like sound (**Figure 5**). Both receive a mono mix and file IDs from the quadplayers in order to determine the output.

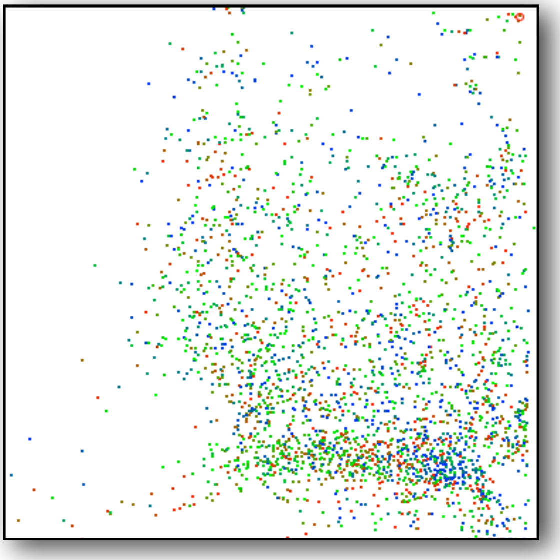


Figure 5. Approximately 3000 dots visualised in the *CataRT* LCD interface (Schwarz 2015). They represent sonic descriptors for short fragments of the mono files, projected onto a plane with spectral flatness and periodicity on x and y axes, respectively.

The fourth instrument, consisting of two sub-players so it might be called a duo, is called 'mirror solos'. It merges mono files two by two in cross-synthesis (using the supervp~ object from IRCAM). While the first two players (quad and CataRT-quad) respect the composers' original work, the last two (CataRT-cloud and the crossynths) produce new and quite unpredictable output.

5. Overall control

Finally (this was the last phase in the design work), an overall mechanism (at the middle level in the player structure; macro - meso - micro) decides which player plays when, by applying fade-ins and fade-outs according to a cyclical process (based on Lindborg's *Graviton Dance* pieces and our collaboration *On The String*, mentioned earlier), and 'freezes', that are applied stochastically. **Figure 6** illustrates the information flow in the player system.

The overall output was intended to have the character of 'chill-out electroacoustic music', yet the material and the players colluded to make it rather active at some moments. The superposition of deterministic and probabilistic mechanisms created a player structure where the interactive design and composed material collide.

Acknowledgement

The production of *When We Collide* was enabled by a grant from National Arts Council of Singapore. The first presentation took place at *Soundislands Festival*, 19-23 August 2015.

References

Hauswolff M von, Lindborg PM, Grönlund T., Harding M., Kirkegaard, J., LaBelle B., Nilsen B.J., Nisunen, P., Petursson, F., Pomassl, F., Tankred, K., Thirwell, J., Urstad, M. & Winderen, J. (2003, 2004, 2005, 2006, 2008, 2012, 2013, 2014). *Freq-out #*. Series of site-specific collaborative sound installations. Each instance is specifically created for the acoustics and context of the location. URL http://www.freq-out.org.

Koh JBT, Lindborg, PM, Stromberg D, Chian KH, & Lim WW. (2010) *On the String, theatre of* *music.* Multimedia performance (60 min.) Commissioned by National Arts Council for Singapore Arts Festival.

Lindborg PM & Koh, Joyce Beetuan (2011, August). "Multidimensional spatial sound design for 'On the String'". *Proceedings of the International Conference of Computer Music* (ICMC), University of Huddersfield, UK. July 2011.

Lindborg PM (2013). Graviton Dance - Semantic. Electroacoustic music composition with generative 3D audio diffusion; concert (8 min.) & interactive installation versions. *International Computer Music Conference 2013*. Perth, Australia. 11-17 August 2013.

Schwarz D, Beller G, Verbrugghe B & Britton S (2006). Real-time corpus-based concatenative synthesis with CataRT. *Proceedings of DAF-X*, p. 279-282, Montreal 2006.

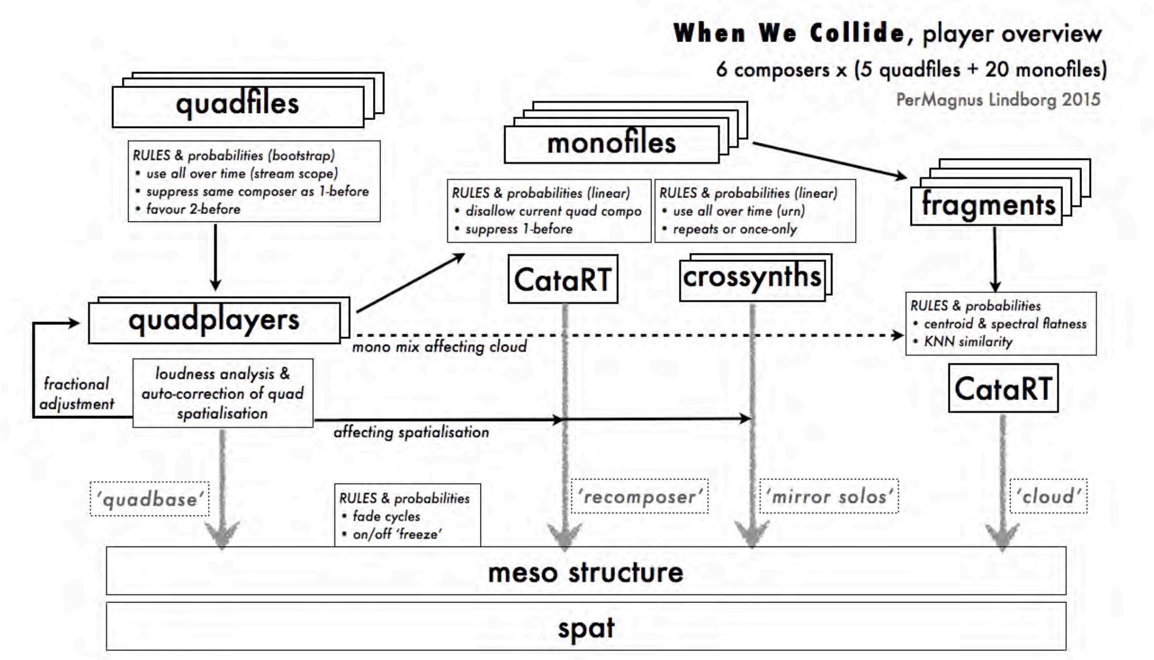


Figure 6. Overview of player design. Top: boxes represent sound files (provided by the artists), and descriptions of selection algorithms. Middle: boxes represent instruments, producing in all 18 channels of audio. The arrows indicate some of the ways that sub-systems influence each other. Bottom: boxes represent meso structure (i.e. control layer) and spatialisation.